



**HAL**  
open science

# ÉTUDE GÉNOMIQUE, MÉTAGÉNOMIQUE ET PHYSIOLOGIQUE DE LA DIVERSITÉ PIGMENTAIRE CHEZ LES CYANOBACTÉRIES DU GENRE SYNECHOCOCCUS

Florian Humily

► **To cite this version:**

Florian Humily. ÉTUDE GÉNOMIQUE, MÉTAGÉNOMIQUE ET PHYSIOLOGIQUE DE LA DIVERSITÉ PIGMENTAIRE CHEZ LES CYANOBACTÉRIES DU GENRE SYNECHOCOCCUS. Océanographie. Université Pierre et Marie Curie - Paris VI, 2013. Français. NNT: . tel-00851664

**HAL Id: tel-00851664**

**<https://theses.hal.science/tel-00851664v1>**

Submitted on 16 Aug 2013

**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.

**THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**  
Spécialité Océanographie Biologique  
Ecole Doctorale Sciences de l'Environnement d'Ile de France (ED 129)

Présenté par  
**Florian HUMILY**

Pour obtenir le grade de  
**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

**ÉTUDE GÉNOMIQUE, MÉTAGÉNOMIQUE ET PHYSIOLOGIQUE  
DE LA DIVERSITÉ PIGMENTAIRE CHEZ LES CYANOBACTÉRIES  
DU GENRE *SYNECHOCOCCUS***

Soutenue le 21 Juin 2013, devant le jury composé de:

<b>Pr Timothy VOGEL</b> , Professeur	Université Claude Bernard Lyon I	<b>Rapporteur</b>
<b>Pr Cécile BERNARD</b> , Professeur	Museum National Histoire Naturelle, Paris	<b>Rapporteur</b>
<b>Dr Ghada AJLANI</b> , Chargée de Recherches	CEA, Gif-Sur-Yvette	<b>Examineur</b>
<b>Dr Delphine LATOUR</b> , Maître de Conférences	Université Pascal Blaise, Clermont-Ferrand	<b>Examineur</b>
<b>Pr Robert BELLÉ</b> , Professeur	UPMC, Station Biologique de Roscoff	<b>Président du jury</b>
<b>Dr Laurence GARCZAREK</b> , Chargée de Recherches	Station Biologique de Roscoff	<b>co-Directrice de thèse</b>
<b>Dr Frédéric PARTENSKY</b> , Directeur de Recherches	Station Biologique de Roscoff	<b>co-Directeur de thèse</b>



*«Je sais pourquoi tant de gens aiment couper du bois.  
C'est une activité où l'on voit tout de suite le résultat.»*

**Albert Einstein.**



*«La théorie, c'est quand on sait tout et que rien ne fonctionne.  
La pratique, c'est quand tout fonctionne et que personne ne sait pourquoi.  
Ici, nous avons réuni théorie et pratique : Rien ne fonctionne... et personne ne sait pourquoi »*

**Albert Einstein.**



---

---

## REMERCIEMENTS

---

En premier lieu je tiens à remercier Mr Timothy Vogel, Mme Cécile Bernard, Mme Ghada Ajlani, Mme Delphine Latour et Mr Robert Bellé d'avoir accepté de juger ce travail. Merci également à Stéphane Hourdez et Alexis Dufresne pour vos conseils, remarques, et vos questions pertinentes qui m'auront fait prendre conscience d'un certain nombre de choses lors de mes deux comités de thèse.

16 Février 2013... Il a fait beau temps<sup>1</sup> ce jour (mention importante pour nos deux marseillais exilés qui reviendra tout au long de ces remerciements en guise de piqure de rappel). Il est tard. Ludovico Einaudi en fond (dédicasse à Baboun) et me voilà parti dans la rédaction d'une partie que beaucoup attendent. Moi en premier lieu.

Au départ, à la suite d'une discussion de 19h avec Justine Pittera en date du 30 Janvier 2013, celle-ci soulignait que "les remerciements se rédigent toujours à la fin". J'avais écrit cette phrase en lui promettant de l'inclure à cette partie. Chose faite, même si j'ai décidé de ne pas attendre la dernière minute pour écrire ces quelques lignes. J'ai en effet décidé de m'y mettre un peu avant, au cas où ... car il y a quelques jours j'ai été pris de panique car des météorites sont tombées sur l'Oural, alors je me suis dit que ça pouvait peut-être tomber sur Roscoff ("t'à qu'à croire" comme le dit souvent mon collègue Steven le Roy).

Arrivé il y a plus de 4 ans maintenant à la fameuse Station Biologique de Roscoff, après avoir fait un passage, finalement éclair, au sein de l'équipe des grandes algues, au sein de laquelle je remercie mes premiers encadrants scientifiques Philippe Potin et Ludovic Delage, c'est finalement sur des algues microscopiques que j'ai décidé de faire ma thèse. Là, entre ces vieux murs, dans l'île de Léon, pas loin d'Île de Batz. Sacré paysage. Et en plus on m'avait dit à l'époque « tu verras il y a plein de sacrés têtes là-bas ». Après plus de 4 ans passés entre ces murs : je confirme.

Vieux murs, vieux hommes (pas que !), jeunes esprits. Janvier 2009, le début d'un stage de Master 2 qui s'annonce passionnant, avec plein de techniques compliquées comme je les aime. Ce stage aura finalement été à la hauteur de mes attentes, et malgré beaucoup de travail et finalement peu de résultats scientifiques, m'aura réellement comblé. Quelques temps après, après quelques péripéties, Monsieur Frédéric Partensky, m'annonce au téléphone « Finalement Flo, c'est bon, on pourra te financer sur le projet ANR ». Sourire en coin, explosion de joie, je me souviens il faisait beau temps<sup>1</sup> ce jour-là. A la suite de cette nouvelle j'ai passé un super été mêlé de Vieilles Charrues, de Saint-Emilion, de concours de pétanque ... le plus sympa de toute ma vie.

C'était parti pour 3 ans. Finalement cela aura été plus long que prévu.

Maintenant par où commencer ? Il y a tant de choses à dire et à raconter. Je pense que dans la logique il faut commencer par ceux sans qui cela ne serait jamais arrivé. Merci à Fred Partensky, matinal comme jamais, prêt à vous envoyer des e-mails à 6h un dimanche matin, un homme de science, passionné, avec qui j'ai beaucoup appris. Même si on n'a pas réussi à trouver le rôle de MpeU, j'espère que ce problème aura attiré ton attention, et que le jour où tu trouveras la méchante

cystéine qui est sous l'influence de cette phycobiline-lyase, tu penseras à moi. Merci de m'avoir au final fait rentrer dans ce cercle si fermé de cette famille protéique, qui au départ m'effrayait au possible. Merci pour la dernière ligne droite, loin d'être la plus facile et la plus marrante, mais je garderais de celle-ci celle où j'aurais le plus interragit avec toi ... un peu trop tard surement. Merci biensur à Loh sans qui tout le travail divers et varié, fourni au cours de cette thèse n'aurait pas eu lieu. Si tout avait fonctionné pour le mieux au cours de mon stage de Master 2, j'avoue avoir connu davantage de difficultés une fois rentré dans le monde de la thèse. Il en est ainsi. Merci pour cette aventure, qui n'aura pas toujours été facile, mais qui m'aura beaucoup appris, à la fois en terme de rigueur et de perfectionnisme qu'à bien d'autres niveaux. Je ne fais nul doute que de ces longues étapes de purification, de clonage, de suivi cyto, sortiront de supers résultats et de publications ... Je le souhaite en tout cas. Merci à vous deux, pour toutes les aventures menées ensemble, entre les suivis à 8h du matin un dimanche, les réunions de 15 min qui duraient finalement 3 heures, les échanges scientifiques, les coups de gueule et tant d'autres choses. Puisse vos explosions d'idées se concrétiser comme vous le souhaiter. J'en profite également pour vous remercier de l'opportunité que vous m'avez offerte de voyager, et d'aller rencontrer les pros de la WGA autour d'un homard du golfe du Maine, discuter d'acclimatation chromatique dans la ville de Porto, ou encore de goûter aux paroles étonnantes d'un Monsieur Jeanthon à la fête concernant des louboutins. On n'a pas trop de quoi se plaindre au final.

Troisième homme de l'équipe, pro des caroténoïdes mineurs, de l'HPLC, et surtout des membres protéiques associés au PSII, un grand merci à toi Toif. Je n'oublierais pas tes conseils, ta disponibilité, tes cours sur le fonctionnement de l'appareil photosynthétique, ta diapo avec une croix barrée sur l'ordonancement linéaire des PBS, ainsi que l'aide sur le PAM (bien différente des explications de Mr Partensky ;-). Je n'oublierais pas nos discussions autour de raisin fermenté, et je serais heureux d'apprendre enfin le jour où tu trouveras que le Bourgogne, le bon hein ?, est bien meilleur que ses cousins de Bordeaux. Si un jour tu veux te mettre à la PCR ou la WGA, ou plein d'autres techniques mêlant acides nucléiques, n'hésite pas.

Le groupe Plancton est tellement grand que je ne sais pas par où commencer. Après tout je me dis que pour faire simple, il vaut mieux commencer par une journée type. Arrivée tôt, Marie-Noëlle est en général la première que je salue. Premier sourire matinal, et ce tous les jours pendant plus de 4 ans. Finalement les journées commençaient pas mal. Merci à toi pour ces labos propres et ta franche bonne humeur ! Enfin tous les jours est un peu mentir, car des fois Laure Guillou, qui venait de se réveiller de son lit sous la hotte à flux horizontal pour ses repiquages, lui gravissait la place. Je pense que tu auras été la plus matinale de nous deux au terme de ces 4 ans, et je t'en félicite.

Petite manip, petites bises aux filles, petit café, et puis le bonjour à Brigitte. Ah Brigitte ! Merci pour tout, pour tes envois carbo, et merci de n'avoir jamais diffusé ces photos prises aux Vieilles Charrues. 55 000 personnes et il a fallu que je tombe sur toi. Un grand moment d'anthologie. Au rez-de-chaussée se cache également une perle, celle qui vous aura appris le fonctionnement d'un autoclave humide, juste avant que sonne votre timer pour votre expérience imminente, qui biensur ne pouvait pas attendre. Si toi, lecteur, tu souris en lisant ces lignes, c'est que je ne suis pas le seul à qui c'est arrivé. Deux dames sans qui la SBR ne serait pas ce qu'elle est en terme de fonctionnement. J'en profite également pour remercier toutes les bonnes âmes de l'ombre qui contribuent de façon substantielle à la qualité de recherche menée à la Station : merci à tous les gens de l'administration, au service Mer, aux informaticiens/bioinformaticiens. Merci à tout le personnel technique, à Jack pour les planches de contreplaqués, Daniel pour ses conseils élec, ainsi qu'au personnel du Gulf ! Le meilleur moment de la journée, que même sur la planète Mars on nous envie. Cela va être un manque énorme de ne plus cottoyer Frank et sa bande ! Merci au personnel de la plateforme de séquençage, Morgan et Gwenn, pour les milliers d'échantillons que les séquenceurs auront analysés durant cette thèse. Enfin merci aux gardiens, et notamment Denis ! Merci pour ta franche bonne humeur et de m'avoir permis de m'évader quelque peu à certains moments de la rédaction de cette thèse.

Je reviens aux filles, noyau dur du groupe plancton. Merci à toutes mes « coupines » comme disait Anne-Laure Sauvadet. Entre Momo, Pris, Roselin(e) avec qui je n'aurais partagé que de bons moments, et qui auront toujours, mais alors toujours été là pour m'écouter, me conseiller, prendre un café, dire que « c'est pas normal », partager un tupperware, parler gosses, pontes et autres sujets indécents, et tout ça, tout ça... J'inclue bien évidemment la Gwenn locale à ce groupe de drôles de dames. Je pourrais continuer pendant des pages et des pages, mais on n'est pas dans un livre de J.K. Rowling.

Merci aux mamans Jaja et Flo, qui conditionnent l'âme de ce groupe Plancton, et sont au petit soin de tous les petits nouveaux, anciens, doctorants, étudiants, stagiaires, et tant d'autres. Merci à Sarah, la Pro de la raclette, du Picogreen et du 454 à travers le Génoscope. Longue vie aux jeunes mariés. Mathieu, que la force du repiquage soit avec toi.

Arrive enfin le moment de remercier mes deux pères de manip, avec qui j'aurais tellement appris. Commençons par Dominique Marie, celui qui m'auras tout appris du monde de la cyto, et fais rentrer dans le cercle ultra-fermé des utilisateurs de l'ARIA. Entre le SSC, le fluidic shutdown, le coupage des sondes, la raccourcissement des tubulures qui ne servent « à rien », les coups de tournevis, les grands fous rires, la tasse de café noire, les discussions bricolage, et surtout, n'oublions pas la charlotte et la blouse (un moment épique), que de bons moments ! Après la ferveur du PSG, passons à la ferveur du taboulé. Ahalalal Mr Jeanthon. Comment vous remercier de tout ce que vous m'avez appris ? Entre les recettes de la Saint-Jacques à l'orange, les conseils sur la purification d'ADN au CsCl, les discussions autour de mes manips, il y en aurait des choses à raconter ... Je ne garderais que de bons souvenirs de cette première année de thèse, marquée de covoiturage, et de commentaires miso par rapport à la voiture de devant qui n'avancait pas, et du vin algérien à Paris.

De ce groupe Plancton je me souviendrais également des group meeting, souvent marqués de présentation de Daniel Vaultot ou de Colomban de Vargas. Quel plaisir d'avoir tant appris à travers vos talks ! Daniel merci pour toutes ces discussions WGA et échanges autour de la SCG; je regrette néanmoins que tu ne te sois pas converti au GenomiPhi v2. Merci à Colomban, l'homme plancton comme il se définit lui-même, pour toute son œuvre. De l'homme suisse que tu es, je n'oublierais jamais les vidéos Thalassa, entre le filet à papillons qui entre dans la canopée de l'océan, le sperme de l'océan, et l'art de filtrer le plancton qui te rapellent ta campagne suisse et les vaches ... Exceptionnel.

Merci à tous les anciens de la SBR : Cécile, Anne-Cat, Sergio (l'homme capable de manger 2 homards du Maine sans aucun problème), Anne-Laure, François, Madhi, Catharina, Raffaele... Merci à tous les anciens stagiaires qui auront apporté gaieté et bonne humeur au sein de notre groupe. Mention spéciale à Maxine, la fille du Nord, qui m'aura appris à faire les frites en deux cuissons (un délice). Je n'oublie pas non plus la Marine, entre le stress parce que « pas le temps » ... et « non je ne prends pas de desserts » aura participé à cette bonne humeur au labo. Avoue que tu avais peur de te lever hein ? Cela me ramène à la Juju locale. Sacré brin de fille. Ces soirées aux Gulf avec la chemise rouge de Domi pour la paëlla, et Franky toujours impeccable auront été mémorables. L'ambiance de notre petit bureau aura bien changé depuis ton départ, ma Dani. Je n'ai plus de grands moments calmes, sans un bruit, s'arrêtant net à cause d'un fou rire attrapé à travers une vidéo FB. Bonne continuation à toi, et merci pour tout. Surtout pour les grains de café entourés de chocolat, une véritable tuerie que je ramènerais en container du Pérou si l'occasion se présente.

Mes remerciements les plus sincères vont également à mes deux stagiaires, travailleuses acharnées, pro de la WGA et des tris cellulaires. Merci à Jennyfer Langlois et Audrey Nicolas. Même si vos chemins se sont un peu écartés de la Recherche, je vous souhaite le meilleur ! Un grand merci à toi Audrey pour cette nuit dantesque de tri et de WGA...

Je n'oublie pas la Manceau ! Ah Céline, entre tes coups de gueule, tes « j'en ai marre », tes sourires, ta bonne humeur, et ton aide si précieuse à bien des égards ! Merci pour tout à la Maman de l'UMR7144.

Merci à mes collègues de galère, DB, Aliou et Frédo. Ah pas toujours simple la thèse hein ? Merci pour ces moments passés ensemble, à ces moments d'écoute, à ces repas (le must restant le sandwich au poulet) et bières partagées ! Bonne continuation à vous et merci pour tout. Courage pour la dernière ligne droite à ceux concernés ! Un grand merci à toi DB, collègue de la retouche d'image, et de tant d'autres choses. Merci de m'avoir entraîné dans cette galère de locations de vélos... :( Aliou merci pour ces séances de repiquage et m'avoir fait remarqué que prendre une après-midi n'était pas partir à 16h ...

Greg, j'espère que tu en tireras un max de papiers de ces 25 génômes ... Puisse la puissance des la, lb, CRD1a, WPC1 t'extasier longtemps. Merci pour ces assemblages et ton aide bioinfo.

A Justine Pitt : Alalalla que de choses à dire de cette virée Banyulencque ! Entre mer et soleil<sup>1</sup>, Denver et tunicien, la remise en place, un séjour fort en travail et en fou rire... Merci à toi pour ces moments ! Puisse la patience t'animer, et courage à toi avec ces souches qui ne poussent pas !

Merci à tous mes potes, qui auront été là pour me faire penser à autre chose. T'à qu'a croire on va en refaire un max de soirées à déguster de la Tripel Karm ou du Saint-é, à profiter de la plaine de Kerampuilh sous le soleil<sup>1</sup> ... Merci pour votre soutien dans certains de mes choix. Merci pour les foires à la Penfeld, mon RDV de l'année ! On verra ce que cela donne ces nouvelles aventures ... Merci à la famille, qui m'aura également permis de m'évader, à travers des séances de tronçonnage, ou des nombreuses expériences de bricolage, qui m'auront permis sans nul doute de tenir jusqu'à la fin ! Merci à toi Papa, de m'avoir donné l'opportunité de faire ces longues études !

Enfin merci à celle qui aura le plus souffert durant cette thèse. Agathe à travers tes sacrifices, tes coups de gueule, tu m'auras permis d'aller jusqu'au bout de cette chose. Tu auras été le soutien de tous les jours. Désolé pour cette période pas facile, pour ces choses qu'on n'a pas pu faire. Comme disait Jean-Louis Aubert, voilà c'est fini ... il était temps. Cela va changer.

A tous ceux m'ayant permis de remporter 2 fois de suite la soirée Wizz : je vais rajouter ça sur mon CV c'est sur !

Je remercie en dernier lieu mon nouveau PC qui aura évité que je pète un plomb, Europe 1 et la bande à Ruquier, Renaud, AC/DC, Ludovico Einaudi, et le club des chanteurs dont je tairais le nom (Ju, à quand le Zenith?), etc ...

A tous ceux que j'ai oublié ... mais que je n'oublierais pas.

Sachez enfin que je reste disponible pour tout conseil de Weber ... Parce que la pizz' ou l'poulet au Weber, et ben c'est complètement différent !

17 Février 2013. Jacques Brel a remplacé Ludovico Einaudi. Grand soleil<sup>1</sup>. Les PBS de Syn sont au max de leur puissance, dégagement d'oxygène, bouffée d'air frais.

L'art de me dire, qu'après finalement tant de temps : Ca, c'est fait.



Figure numéro la plus belle : T'as vu mon pays ? ! Faut pas avoir des pikouz dans les yeux pour être genaoueg devant la beauté du paysage. Si t'apprécie pas, soit t'es une brell, soit t'es une pikez, soit t'es un drôle. Hopala, quel glaouch va doue ! Ca va faire du reuz la beauté du paysage quand même t'a qu'a croire ! Bon va falloir arreter d'glabousser et d'chaoquer, j'sens que j'traîne. Echu tout !

PS : si en lisant cela tu te dis Pebezh klapez celui-là : garde mon 06, on s'fra une bouffe et après tu seras complètement poazh à en prendre les ribins ... Mat pell 'zo.

MERSIBRAS \*

CHOUK'RAN\*\*

<sup>1</sup> mention importante pour nos deux marseillais.

\* Merci en Breizhoneg

\*\*Merci en Arabe (dédicasse à Aliou, qui m'a dit que cela n'existait pas).



---

---

# SOMMAIRE

---

LISTE DES FIGURES .....	V
LISTE DES TABLEAUX.....	VIII
LISTE DES ABBRÉVIATIONS .....	IX
VALORISATION SCIENTIFIQUE DU PROJET DE THÈSE .....	X
CHAPITRE I INTRODUCTION GÉNÉRALE.....	1
<i>PRÉAMBULE</i> .....	1
<i>PARTIE I L'IMPORTANCE DES PICOCYANOBACTÉRIES DANS LE PHYTOPLANCTON MARIN</i> .....	3
I.1. Généralités sur le phytoplancton marin.....	4
I.1.1 Définition et découverte du picophytoplancton.....	4
I.1.2 Rôle du phytoplancton dans la pompe biologique.....	9
I.2. Les picocyanobactéries marines.....	10
I.2.1 Origine et évolution des cyanobactéries marines.....	10
I.2.2 Morphologie et organisation cellulaire des picocyanobactéries.....	14
I.2.3 Distribution et dynamique des picocyanobactéries marines.....	18
I.2.4 Diversité génétique chez les picocyanobactéries marines et phylogéographie.....	21
I.2.4.1. <i>Prochlorococcus</i> .....	21
I.2.4.2. <i>Synechococcus</i> .....	24
<i>PARTIE II ROLE DES FACTEURS ENVIRONNEMENTAUX DANS LA DISTRIBUTION DES PICOCYANOBACTÉRIES MARINES</i> .....	31
II.1. Facteurs environnementaux.....	32
II.1.1 Facteurs biotiques.....	33
II.1.1.1. Le broutage.....	33
II.1.1.2. La lyse virale.....	34
II.1.2 Facteurs abiotiques.....	34
II.1.2.1. Intensité lumineuse.....	34
II.1.2.2. Qualité de lumière.....	36
II.1.2.3. La disponibilité en nutriments.....	38
II.1.2.4. La température.....	39
II.2. La compétition pour la lumière dans l'océan.....	39
II.2.1 La photosynthèse oxygénique.....	39
II.2.2 L'organisation générale de l'appareil photosynthétique oxygénique.....	41
II.2.3 Les photosystèmes.....	43
II.2.3.1. Le photosystème II.....	43

II.2.3.2.	Le photosystème I .....	44
II.2.3.3.	Production d'énergie au sein des membranes thylacoïdales.....	45
II.2.4	Les antennes collectrices de <i>Prochlorococcus</i> .....	45
II.2.5	Les phycobilisomes de <i>Synechococcus</i> .....	49
II.2.5.1.	Structure des phycobilisomes.....	49
II.2.5.2.	Phycobiliprotéines et chromophorylation .....	54
II.2.5.3.	Biosynthèse et attachement des phycobilines .....	59
II.2.5.4.	Transfert orienté d'énergie .....	62
II.2.6	Gènes codant pour les polypeptides des phycobilisomes.....	64
II.2.6.1.	Nature et organisation génique.....	64
II.2.6.2.	Processus évolutifs des phycobiliprotéines .....	67
II.3.	L'acclimatation chromatique chez les cyanobactéries.....	71
II.3.1	Diversité des souches capables d'acclimatation chromatique chez les cyanobactéries.....	72
II.3.2	Bases moléculaires de l'acclimatation chromatique de type III .....	73
II.3.3	Bases moléculaires de l'acclimatation chromatique de type II .....	75
II.3.4	Originalité de l'acclimatation chromatique de type IV présente chez les <i>Synechococcus</i> marins.....	77
<b>PARTIE III LES APPORTS DE LA GÉNOMIQUE ET DE LA METAGÉNOMIQUE EN MICROBIOLOGIE MARINE.....</b>		<b>79</b>
III.1.	Les apports de la génomique marine .....	80
III.1.1	Génomique des espèces marines cultivées.....	80
III.1.2	Réduction génomique et succès écologique.....	83
III.1.3	Dynamique des génomes picocyanobactériens.....	84
III.2.	L'étude des communautés non cultivées : l'ère de la métagénomique .....	87
III.2.1	De la classification des microorganismes à la phylogénomique.....	87
III.2.2	La naissance de la métagénomique .....	88
III.2.2.1.	Métagénomique et clonage : choix du vecteur .....	89
III.2.2.2.	Approches "shotgun" et nouvelles technologies de séquençage à haut-débit.....	92
III.2.2.3.	Principaux problèmes méthodologiques d'une approche métagénomique .....	94
III.2.3	Intérêts de l'approche métagénomique .....	95
III.2.3.1.	Classification, diversité phylogénétique et évolution .....	95
III.2.3.2.	Découverte de nouvelles fonctions métaboliques .....	97
III.3.	L'entrée dans l'ère de la métagénomique ciblée .....	99
III.3.1	La métagénomique ciblée .....	99
III.3.1.1.	Définition et principe.....	99
III.3.1.2.	Intérêts de la métagénomique ciblée .....	99
III.3.2	Un champs en pleine mutation .....	100
III.3.2.1.	Isoler les microorganismes d'intérêt .....	100
III.3.2.2.	Techniques de lyse cellulaire .....	102
III.3.2.3.	Amplification de génomes complets : techniques et contraintes .....	102
III.3.3	Apports de la métagénomique ciblée .....	104
III.3.3.1.	Obtenir de nouveaux génomes de référence .....	104
III.3.3.2.	Interactions microbiennes in situ .....	105
III.3.3.3.	Diversité et fonction .....	106
<b>OBJECTIFS.....</b>		<b>109</b>

## CHAPITRE II ROLE DE LA DIVERSITÉ PIGMENTAIRE DANS L'ADAPTATION A LA NICHE CHEZ *SYNECHOCOCCUS*..... 111

### **ARTICLE I : DIVERSITÉ, ABONDANCE ET DISTRIBUTION DES DIFFÉRENTS TYPES PIGMENTAIRES DANS L'ENVIRONNEMENT PAR UNE APPROCHE CIBLÉE DE METAGÉNOMIQUE..... 113**

I. CONTEXTE DU PROJET .....	113
II. MATÉRIELS ET MÉTHODES .....	115
II.1. Stratégie d'échantillonnage et tri cellulaire .....	115
II.1.1 Collecte et concentration des échantillons environnementaux.....	115
II.1.2 Détermination des abondances picophytoplanctoniques par cytométrie en flux	116
II.1.3 Tri de cellules du genre <i>Synechococcus</i> à partir d'échantillons naturels.....	117
II.1.3.1. Nettoyage du cytomètre et préparation d'expérience .....	117
II.1.3.2. Tri cellulaire.....	118
II.2. Amplification de génomes complets à partir d'échantillons triés.....	118
II.3. Fabrication de bibliothèques de fosmidomes.....	119
II.3.1 Traitement enzymatique à la Nuclease S1 .....	119
II.3.2 Sélection des fragments d'intérêt par PFGE.....	120
II.3.3 Traitement enzymatique.....	120
II.3.4 Ligation, empaquetage et titrage.....	121
II.3.5 Détermination de la taille des inserts .....	121
III. ARTICLE.....	121
IV. DONNÉES SUPPLÉMENTAIRES A CETTE ÉTUDE.....	155

## CHAPITRE III DIVERSITÉ PIGMENTAIRE DES SOUCHES EN CULTURE DE *SYNECHOCOCCUS* 161

### **APPORTS DE LA GÉNOMIQUE COMPARATIVE A L'ÉTUDE DE LA PIGMENTATION CHEZ *SYNECHOCOCCUS*..... 163**

I. CONTEXTE DU PROJET .....	163
II. MATÉRIELS ET MÉTHODES .....	165
II.1. Préparation de l'ADN génomique.....	165
II.1.1 Culture cellulaire .....	165
II.1.2 Purification des souches de <i>Synechococcus</i> .....	165
II.1.3 Extraction d'ADN génomique de souches du genre <i>Synechococcus</i> .....	167
II.2. Analyses phylogénétiques des opérons codant pour les sous-unités des phycobiliprotéines.....	168
III. RÉSULTATS ET DISCUSSION.....	169
III.1. Caractérisation des gènes impliqués dans la capture de lumière .....	169
III.2. Processus évolutifs des phycobiliprotéines.....	173

### **ARTICLE II : DIVERSITÉ ET ÉVOLUTION DU PHÉNOMÈNE D'AC4 CHEZ *SYNECHOCOCCUS* 175**

I. CONTEXTE DE L'ÉTUDE .....	175
------------------------------	-----

---

II. ARTICLE.....	176
III. DONNÉES COMPLÉMENTAIRES A CETTE ÉTUDE .....	215
<b>CHAPITRE IV DISCUSSION GÉNÉRALE ET PERSPECTIVES .....</b>	<b>227</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>	<b>239</b>
<b>ANNEXES .....</b>	<b>277</b>
<i>ANNEXE A : ARTICLE III IMPACT DU RÉCHAUFFEMENT CLIMATIQUE SUR LA DISTRIBUTION DES ÉCOTYPES DE PROCHLOROCOCCUS ET SYNECHOCOCCUS EN MER MEDITERRANÉE.....</i>	<b>279</b>
<i>ANNEXE B : ARTICLE IV CYANOLYASE : UNE BASE DE DONNÉES DÉDIÉE AUX PHYCOBILINES LYASES.....</i>	<b>302</b>
<i>ANNEXE C : POSTER I A TARGETED METAGENOMIC APPROACH TO STUDY THE PIGMENT DIVERSITY OF MARINE SYNECHOCOCCUS AND THE DISTRIBUTION OF PIGMENT TYPES IN SITU.....</i>	<b>310</b>
<i>ANNEXE D : POSTER II DIVERSITY AND FUNCTIONAL CHARACTERIZATION OF CHROMATIC ACCLIMATION IN MARINE SYNECHOCOCCUS CYANOBACTERIA.....</i>	<b>312</b>
<i>ANNEXE E : NOUVEAUX GÉNOMES DE SYNECHOCOCCUS SÉQUENCÉS.....</i>	<b>314</b>

---

---

## LISTE DES FIGURES

---

<b>FIGURE 1</b> : Distribution des organismes phytoplanctoniques au sein de l'arbre du vivant.....	4
<b>FIGURE 2</b> : Efflorescences de phytoplancton.....	6
<b>FIGURE 3</b> : Magnitude et distribution de la production primaire, à la fois océanique et terrestre.....	7
<b>FIGURE 4</b> : Classifications du plancton selon leur taille, leur mode de nutrition et leur règne.....	8
<b>FIGURE 5</b> : Rôle de la boucle microbienne dans le recyclage de la matière organique au niveau océanique .....	10
<b>FIGURE 6</b> : Colonnes stromatolithiques de la baie Shark sur le littoral ouest de l'Australie .....	11
<b>FIGURE 7</b> : Chronogramme de reconstruction de l'état ancestral par horloge moléculaire en utilisant un arbre <i>rpoC</i> - <i>ssu</i> .....	12
<b>FIGURE 8</b> : Arbre en maximum de vraisemblance basé sur l'analyse de séquences codant pour l'ARNr 16S des espèces cyanobactériennes dont le génome a été entièrement séquencé.....	13
<b>FIGURE 9</b> : Photographies en microscopie électronique de cellules de <i>Prochlorococcus</i> et <i>Synechococcus</i>	15
<b>FIGURE 10</b> : Représentation schématisée des différents compartiments cellulaires chez les picocyanobactéries .....	17
<b>FIGURE 11</b> : Distributions verticales typiques des cyanobactéries <i>Prochlorococcus</i> et <i>Synechococcus</i> .....	20
<b>FIGURE 12</b> : Arbre phylogénétique basé sur l'espace intergénique 16S-23S montrant les relations entre les différents clades ou écotypes de <i>Prochlorococcus</i> .....	22
<b>FIGURE 13</b> : Arbre phylogénétique en maximum de vraisemblance basé sur l'espace intergénique 16S-23S montrant les relations entre les différents clades de <i>Synechococcus</i> du subcluster 5.1 .....	27
<b>FIGURE 14</b> : Arbre phylogénétique en maximum de vraisemblance basé sur l'espace intergénique 16S-23S montrant les relations entre les différents clades de <i>Synechococcus</i> du subcluster 5.3 .....	28
<b>FIGURE 15</b> : Analyse factorielle menée à partir d'échantillons environnementaux provenant de différentes campagnes océanographiques .....	32
<b>FIGURE 16</b> : Intensités lumineuses de surface obtenues sur la période 1983 - 2005, à partir des moyennes annuelles sur la période, du mois de Décembre et du mois de Juin .....	35
<b>FIGURE 17</b> : Pénétration des radiations lumineuses dans trois types d'environnements aquatiques : océan ouvert (station ALOHA de l'océan Pacifique), eaux côtières (mer Baltique) et eaux turbides de lacs (lac Groote Moost) .....	37
<b>FIGURE 18</b> : Apparition de la photosynthèse, augmentation de la teneur en oxygène atmosphérique et évolution du vivant. ....	40

<b>FIGURE 19</b> : Organisation de l'appareil photosynthétique d'une plante vasculaire .....	41
<b>FIGURE 20</b> : Schéma en "Z" représentant le transfert d'électrons lors de la photosynthèse oxygénique conduisant à la photolyse de l'eau conduisant à la formation de dioxygène, d'ATP et de NADPH.....	42
<b>FIGURE 21</b> : Spectres d'absorption de la lumière visible des pigments majoritaires retrouvés au sein des cellules phytoplanctoniques et structures chimiques associées .....	43
<b>FIGURE 22</b> : Antennes collectrices de lumière retrouvées au sein des picocyanobactéries marines au niveau du photosystème II.....	46
<b>FIGURE 23</b> : Pigments majoritaires de la picocyanobactérie marine <i>Prochlorococcus</i> et implications écologiques.....	48
<b>FIGURE 24</b> : Photos en microscopie électronique à transmission montrant la présence des phycobilisomes au niveau des thylacoïdes.....	49
<b>FIGURE 25</b> : Structure du phycobilisome .....	50
<b>FIGURE 26</b> : Rôle des protéines de liaison dans l'assemblage et l'architecture du phycobilisome .....	52
<b>FIGURE 27</b> : Représentation schématique des localisations probables des protéines de liaison au sein d'un phycobilisome de <i>Synechococcus</i> sp. WH8102, comportant de la phycoérythrine II .....	53
<b>FIGURE 28</b> : Chromophorylation des différentes phycobiliprotéines.....	55
<b>FIGURE 29</b> : Liaison des différentes phycobilines aux résidus cystéines des phycobiliprotéines par des liaisons thioéthers.....	55
<b>FIGURE 30</b> : Types pigmentaires rencontrés au sein des <i>Synechococcus</i> marins et spectres d'absorption des cellules en culture .....	58
<b>FIGURE 31</b> : Biosynthèse des phycobilines à partir de biliverdine IX $\alpha$ .....	59
<b>FIGURE 32</b> : Transfert orienté d'énergie au sein du phycobilisome .....	63
<b>FIGURE 33</b> : Comparaisons des régions génomiques impliquées dans la synthèse et la régulation des bras des phycobilisomes au sein des différentes souches marines de <i>Synechococcus</i> .....	65
<b>FIGURE 34</b> : Modèle d'évolution des gènes des phycobiliprotéines et de développement du phycobilisome .....	69
<b>FIGURE 35</b> : Arbre en maximum de vraisemblance produit à partir des séquences d'acides aminés concatenés de 51 protéines ribosomales .....	70
<b>FIGURE 36</b> : L'acclimatation chromatique de type III chez <i>Fremyella diplosiphon</i> .....	72
<b>FIGURE 37</b> : Bases moléculaires de la régulation de l'acclimatation chromatique de type III.....	74
<b>FIGURE 38</b> : Bases moléculaires de l'acclimatation chromatique de type II chez <i>Nostoc punctiforme</i> .....	76
<b>FIGURE 39</b> : Changement de chromophorylation des phycoérythrines de <i>Synechococcus</i> sp. RS9916 lors de l'acclimatation chromatique de type IV.....	78
<b>FIGURE 40</b> : Microorganismes marins entièrement séquencés.....	81
<b>FIGURE 41</b> : Taille et GC% des génomes des picocyanobactéries marines, comparés à une sélection représentative d'autres cyanobactéries .....	82
<b>FIGURE 42</b> : Ilots génomiques chez <i>Prochlorococcus marinus</i> sp. MED4 et <i>Synechococcus</i> sp. WH8102 ....	85
<b>FIGURE 43</b> : Intersections des disciplines traditionnelles de la microbiologie et de la métagénomique .....	89

<b>FIGURE 44</b> : Principe de la fabrication d'une banque métagénomique utilisant des fragments d'ADN environnemental de petite ou grande taille et possibilités d'analyses .....	90
<b>FIGURE 45</b> : Nouvelles technologies de séquençage et domaines d'utilisation .....	93
<b>FIGURE 46</b> : Techniques utilisées pour l'échantillonnage et le fractionnement des différentes classes de taille du plancton .....	94
<b>FIGURE 47</b> : Diversité microbienne observée au sein de différents habitats océaniques, côtiers et estuariens.....	96
<b>FIGURE 48</b> : Principe de la réaction d'amplification par déplacement de brins utilisant la <i>phi29</i> .....	103
<b>FIGURE 49</b> : Biais d'amplification lors de la réaction d'amplification de génomes complets .....	114
<b>FIGURE 50</b> : Localisation géographique des différentes stations échantillonnées .....	155
<b>FIGURE 51</b> : Abondance relative des différents clades de <i>Synechococcus</i> , en se basant sur l'analyse du gene <i>petB</i> à une résolution taxonomique de 94% d'identité .....	157
<b>FIGURE 52</b> : Représentation schématique de la distribution écologique des différents clades de <i>Synechococcus</i> marins en relation avec les principaux gradients environnementaux. ....	159
<b>FIGURE 53</b> : Arbre phylogénétique en Neighbor-Joining basé sur l'analyse des séquences <i>petB</i> obtenues à partir d'échantillons environnementaux et d'isolats en culture .....	164
<b>FIGURE 54</b> : Analyses phylogénétiques des opérons codant pour les sous-unités $\alpha$ et $\beta$ de la phycocyanine, de la phycoérythrine I et de la phycoérythrine II des différents génomes de <i>Synechococcus</i> séquencés .....	174
<b>FIGURE 55</b> : Role of protein synthesis in CA4 process .....	217
<b>FIGURE 56</b> : Variations of the ratios of the two major carotenoids (zeaxanthin and $\beta$ -carotene) to Chl <i>a</i> for <i>Synechococcus</i> strains acclimated to 20 photons $m^{-2} s^{-1}$ of GL or BL.....	219
<b>FIGURE 57</b> : World map illustrating position of oceanographic stations with at least one positive sequence affiliated to <i>mpeZ</i> or <i>mpeW</i> and <i>Synechococcus</i> strains used in this study .....	221
<b>FIGURE 58</b> : Phylogenetic tree showing the relationships between <i>mpeZ</i> and <i>mpeW</i> sequences from <i>Synechococcus</i> isolates and metagenomic reads reads retrieved in CAMERA portal .....	223
<b>FIGURE 59</b> : Partitionnement du spectre solaire dans l'environnement océanique .....	228
<b>FIGURE 60</b> : Mécanisme d'action de Cpcs1.....	232
<b>FIGURE 61</b> : Schéma hypothétique de la régulation de l'acclimatation chromatique de type IV chez les <i>Synechococcus</i> marins en se basant sur le mutant naturel BIOS-E4-1 et mutants de RS9916 étudiés à ce jour.....	235
<b>FIGURE 62</b> : Taille des génomes des <i>Synechococcus</i> et <i>Cyanobium</i> marins, contenu en GC et nombre de gènes.....	314

---

---

## LISTE DES TABLEAUX

---

<b>TABLEAU 1</b> : Diversité écologique des <i>Synechococcus</i> marins et biogéographie des différents clades .....	29
<b>TABLEAU 2</b> : Classification des différents types de phycocyanine selon leur chromophorylation.....	56
<b>TABLEAU 3</b> : Avantages et inconvénients des bibliothèques métagénomiques fabriquées à partir de petits ou grands inserts.....	91
<b>TABLEAU 4</b> : Méthodes d'isolement utilisées pour des approches génomiques à partir d'une cellule .....	101
<b>TABLEAU 5</b> : Paramètres physico-chimiques et abondances des cellules phytoplanctoniques dans les différents échantillons environnementaux analysés.....	156
<b>TABLEAU 6</b> : Souches de <i>Synechococcus</i> séquencées à ce jour et caractéristiques principales.....	166
<b>TABLEAU 7</b> : Gènes codant pour les polypeptides de liaison présents dans les différents génomes .....	170
<b>TABLEAU 8</b> : Gènes codant pour des phycobilines lyases putatives présents dans les différents génomes .....	172
<b>TABLE 9</b> : Presence of <i>mpeZ</i> or <i>mpeW</i> genes in <i>Synechococcus</i> strains .....	220
<b>TABLE 10</b> : List of positive reads affiliated to <i>mpeZ</i> or <i>mpeW</i> sequences and oceanographic stations .....	222

---

---

## LISTE DES ABBRÉVIATIONS

---

<b>AC</b>	Acclimatation chromatique
<b>AC4</b>	Acclimatation chromatique de type IV
<b>ACC</b>	Acclimatation chromatique complémentaire
<b>ADN</b>	Acide désoxyribonucléique
<b>APC</b>	Allophycocyanine
<b>BL</b>	Lumière bleue
<b>Chl</b>	Chlorophylle
<b>Cys</b>	Résidu cystéine
<b>DOC</b>	Dissolved Organic Carbon
<b>DOM</b>	Dissolved Organic Matter
<b>Ga</b>	Milliards d'années
<b>GL</b>	Lumière verte
<b>HL</b>	Ecotype de forte lumière (pour High Light)
<b>ITS</b>	Espace intergénique (pour Internal Transcribed Spacer)
<b>LL</b>	Ecotype de faible lumière (pour Low Light)
<b>LMP</b>	Low melting point
<b>MDA</b>	Amplification par déplacement de brins (pour Multiple Displacement Amplification)
<b>MEB</b>	Microscope électronique à balayage
<b>MET</b>	Microscope électronique à transmission
<b>ML</b>	Maximum de vraisemblance (pour Maximum likelihood)
<b>MO</b>	Microscope optique
<b>NGS</b>	Nouvelles technologies de séquençage (NGS)
<b>NJ</b>	Neighbor-Joining
<b>OEC</b>	Oxygen Evolving Complex
<b>pb</b>	Paire de bases
<b>PBP</b>	Phycobiliprotéine
<b>PBS</b>	Phycobilisome
<b>PC</b>	Phycocyanine
<b>PCB</b>	Phycocyanobilin
<b>PE</b>	Phycoérythrine
<b>PEB</b>	Phycoerythrobiline
<b>PEI</b>	Phycoérythrine de type I
<b>PEII</b>	Phycoérythrine de type II
<b>PS</b>	Photosystème
<b>PUB</b>	Phycourobiline
<b>PVB</b>	Phycoviolobilin
<b>RL</b>	Lumière rouge
<b>SCG</b>	Génomique à partir d'une cellule (pour Single Cell Genomics)
<b>TFF</b>	Filtration tangentielle (pour Tangential Flow Filtration)
<b>UV</b>	Rayons ultraviolets
<b>WGA</b>	Amplification de génomes complets (pour Whole Genome Amplification)

---

---

## VALORISATION SCIENTIFIQUE DU PROJET DE THÈSE

---

### PUBLICATIONS SCIENTIFIQUES

- Boeuf, D., Humily, F., Jeanthon, C. (2013) Mini-review : Diversity and functions of Arctic pelagic bacteria. **Biogeosciences** *in preparation*
- Pittera, J., Humily, F., Garczarek, L., Ratin, M., Thorel, M., Six, C. (2013) Adaptation and response to temperature changes among *Synechococcus*. **The ISME Journal** *in preparation*
- Humily, F., Marie, D., Farrant, G., Perennou, M., Partensky, F., Mazard, S., Perennou, M., Labadie, K., Aury, J-M., Wincker, P., Nicolas A., Scanlan, D.J. and Garczarek, L. Use of targeted metagenomic approach to study *in situ* diversity of pigment types in marine *Synechococcus*. **FEMS Microbiology Ecology** *in preparation*
- Humily, F., Partensky, F., Six, C., Farrant, G., Ratin, M., Marie, D., Kehoe, D.M., Garczarek, L. (2013) A gene island with dual evolutionary origin is involved in chromatic acclimation in marine *Synechococcus*. **Environmental Microbiology** *Submitted*
- Bretaudeau, A., Coste F., Humily, F., Garczarek L., Six, C., Ratin M., Collin O., Schluchter W., Partensky F. (2013) CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. **Nucleic Acid Research** 41, doi:10.1093/nar/gks1091
- Mella-Flores D., Mazard S., Humily S., Partensky F., Mahé F., Bariat L., Courties C., Marie D., Ras J., Mauriac R., Jeanthon C., Bendif E.M., Ostrowski M., Scanlan D.J., Garczarek L. (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? **Biogeosciences** 8, 2785–2804

### PRESENTATIONS ORALES

- Garczarek L., Humily F., Habib, C., Le Corguillé G., Corre E., Brillet L., Hoebeke M., Caron C., Dufresne A. and Frédéric Partensky. Ecological genomics of the marine cyanobacterium *Synechococcus*. **Colloque Génomique Environnementale**, Lyon, France, November 28 to 30 Nov, 2011. (LG)
- Mella-Flores D., Mazard S., Humily F., Partensky F., Mahé F., and Garczarek L. Use of picocyanobacteria as biomarkers of the effects of global warming: a case study in the Mediterranean Sea. Colloque **Microorganisms facing their environment**, Museum National d'Histoire Naturel, Paris, France, October 27 to 28, 2011.
- Garczarek L., Humily F., Habib C., Le Corguillé G., Corre E., Brillet L., Hoebeke M., Six C., Dufresne A., Scanlan D.J., Caron C., and Partensky, F. Niche adaptation in the marine cyanobacterium *Synechococcus*: insights from comparative genomics and metagenomics. **Conférence Jacques-Monod: Génomique écologique intégrative Intégrative ecological genomics**, Roscoff, France, October 15 to 19, 2011. (LG)
- Humily F., Marie D., Habib C., Six C., Ratin M., Le Corguillé, G., Caron, C., Partensky F. and Garczarek L. Use of comparative genomics and metagenomics approaches to study pigment

diversity in natural populations of marine *Synechococcus* cyanobacteria **Workshop Aquagenet**, La Rochelle, France, October 5<sup>th</sup> 2011.

- Humily F., Marie D., Mazard S., Scanlan D.J., Mary I., Partensky F. and Garczarek L. A targeted metagenomic approach to study the pigment diversity of marine *Synechococcus* and the distribution of pigment types *in situ*. **Journée des doctorants**, Paris, France, September 12 to 15, 2011.
- Marie D., Humily F., Six C. Approches et contraintes du tri de bactéries. **Journées Thématiques de l'Association Française de Cytométrie**, Saint-Malo, France, November 25 to 26 Novembre, 2009. (DM)

## POSTERS

- Humily F., Shukla, A., Garczarek L., Six C., Marie D., Schlutchter W.M., Karty J.A., Hammad, L.A. Kehoe D.M. and Partensky F. Diversity and functional characterization of chromatic acclimation in marine *Synechococcus* cyanobacteria. **14th International Symposium on Phototrophic Prokaryotes**, Porto, Portugal, August 5 to 9, 2012
- Humily F., Marie D., Mazard S., Scanlan D.J., Partensky F. and Garczarek L. A metagenomic approach to study the diversity and distribution of cyanobacterial pigment types *in situ*. **Journées Jeunes Chercheurs**, Roscoff, France, November 24th, 2011.
- Humily F., Marie D., Mazard S., Scanlan D.J., Partensky F. and Garczarek L. A metagenomic approach to study the diversity and distribution of cyanobacterial pigment types *in situ*. **Colloque Génomique Environnementale**, Lyon, France, November 28 to 30, 2011. (FP).
- Humily F., Marie D., Mazard S., Scanlan D.J., Mary I., Partensky F. and Garczarek L. A targeted metagenomic approach to study the pigment diversity of marine *Synechococcus* and the distribution of pigment types *in situ*. **Jacques-Monod conference: Génomique écologique intégrative. Integrative ecological genomics**. Roscoff, France, October 15 to 19, 2011.



---

# CHAPITRE I

## INTRODUCTION GÉNÉRALE

---



---

---

## PRÉAMBULE

---

L'océan mondial représente environ 71% de la surface de la Terre. De par leur profondeur, les océans représentent un volume habitable ( $\sim 1,37 \cdot 10^9 \text{ km}^3$ ) près de 300 fois plus important que les habitats terrestres. Les microorganismes marins (ou fraction microbienne) sont les organismes invisibles à l'œil nu vivant dans ces masses d'eau. Cette définition simple cache en réalité une très grande diversité d'êtres vivants, décrits ou restants à découvrir, appartenant à des groupes évolutifs distincts. Les trois domaines du vivant y sont ainsi représentés: bactéries, archées et eucaryotes unicellulaires (ou protistes; Woese et Fox, 1977). Ces microorganismes ont colonisé tous les environnements marins propices à la vie. Ils prospèrent ainsi de la surface jusqu'au plancher océanique, des eaux froides profondes ou de l'Arctique jusqu'aux températures élevées des sources hydrothermales, des eaux désalées de la mer Baltique aux eaux hypersalées de la mer Morte, des zones côtières riches en nutriments aux eaux pauvres du large (Macdonald et al., 1980; Blochl et al., 1997; Rothschild et Mancinelli, 2001). De par cette formidable ubiquité et versatilité, les microorganismes marins constituent les organismes les plus abondants sur la planète d'un point de vue numérique. Bien que ce chiffre varie largement suivant les environnements, il est généralement considéré que chaque  $\text{mm}^3$  d'eau de mer ( $1 \mu\text{L}$ ) contient environ 10 000 virus, 1000 bactéries hétérotrophes, 100 bactéries photosynthétiques, et 10 protistes photosynthétiques (Azam et Malfatti, 2007). Le nombre de cellules procaryotes est ainsi estimée à  $10^{29}$ , un chiffre largement supérieur aux potentielles  $10^{21}$  étoiles de l'univers (Whitman et al., 1998).

En terme de biomasse, plus d'un tiers se trouve dans les 200 premiers mètres en dessous de la surface, là où l'énergie lumineuse est capable de pénétrer. Au sein de cette couche euphotique, le monde microbien est essentiellement constitué de procaryotes, de protistes bactériovores et de eucaryotes autotrophes (Pomeroy et al., 2007). Outre leur abondance, les procaryotes présentent une très grande diversité métabolique. En conséquence, les microorganismes marins sont virtuellement impliqués dans tous les grands processus biogéochimiques de l'océan et même de la planète (Kirchman, 2008).



---

---

## PARTIE I

# L'IMPORTANCE DES PICOCYANOBACTÉRIES DANS LE PHYTOPLANCTON MARIN

---

I.1. Généralités sur le phytoplancton marin.....	4
I.1.1 Définition et découverte du picophytoplancton.....	4
I.1.2 Rôle du phytoplancton dans la pompe biologique.....	9
I.2. Les picocyanobactéries marines.....	10
I.2.1 Origine et évolution des cyanobactéries marines.....	10
I.2.2 Morphologie et organisation cellulaire des picocyanobactéries.....	14
I.2.3 Distribution et dynamique des picocyanobactéries marines.....	18
I.2.4 Diversité génétique chez les picocyanobactéries marines et phylogéographie.....	21
I.2.4.1. <i>Prochlorococcus</i> .....	21
I.2.4.2. <i>Synechococcus</i> .....	24

# I. L'IMPORTANCE DES PICOCYANOBACTÉRIES DANS LE PHYTOPLANCTON MARIN

## I.1. Généralités sur le phytoplancton marin

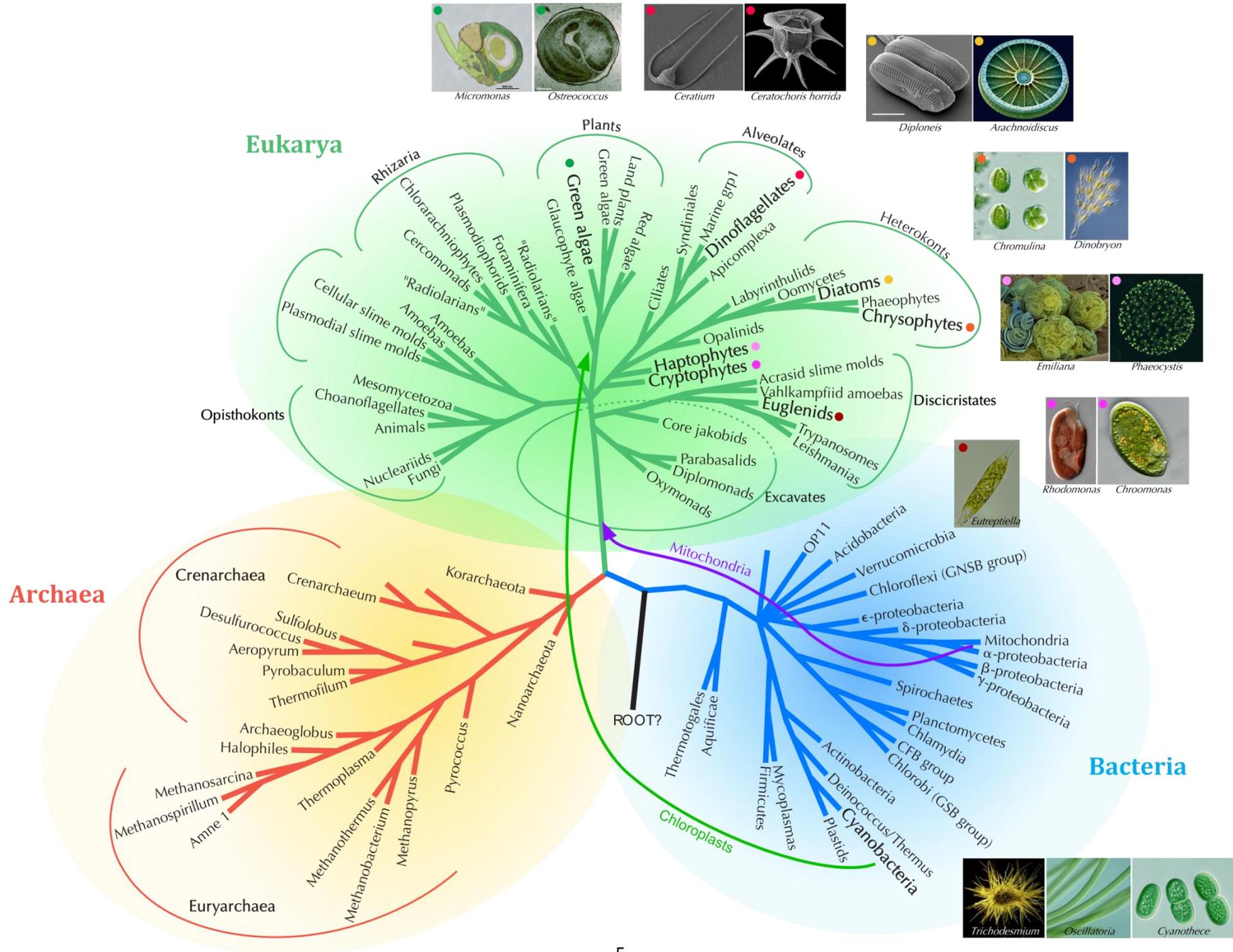
### I.1.1 Définition et découverte du picophytoplancton

Étymologiquement, le mot phytoplancton désigne les organismes photosynthétiques, souvent unicellulaires, dérivant avec les courants (du grec *Phyton* signifiant « végétal » et *Planktós* signifiant « dériver ») par opposition au necton, rassemblant les organismes capables de se mouvoir. Bien qu'il ne représente que 1% de la biomasse de la composante photosynthétique de la planète, le phytoplancton est responsable de 40 à 50% de la production primaire globale (Falkowski, 1994). Si l'on considère que le phytoplancton comporte moins d'espèces décrites que les plantes supérieures, il est en revanche beaucoup plus diversifié d'un point de vue phylogénétique puisqu'il compte huit embranchements taxonomiques et trois lignées évolutives distinctes (Fig. 1).

La composante procaryotique du phytoplancton est constituée par les cyanobactéries, le plus ancien groupe d'organismes capable de photosynthèse oxygénique. Toutes les autres espèces phytoplanctoniques sont des eucaryotes, dont l'histoire évolutive a été marquée par plusieurs évènements endosymbiotiques, suivis de transferts massifs de gènes des endosymbiontes vers le noyau de leurs hôtes. On distingue ainsi la lignée verte de la lignée rouge. Les membres de la lignée verte, aussi appelés chlorophytes, sont issus d'un évènement endosymbiotique primaire entre un organisme proche des cyanobactéries et un organisme eucaryote hétérotrophe possédant des mitochondries, qui aurait eu lieu il y a environ 1.6 à 2.1 milliards d'années (Ga).

---

**Figure 1 (page suivante): Distribution des organismes phytoplanctoniques au sein de l'arbre du vivant.** Des photographies de microorganismes appartenant à chaque lignée sont montrées à titre d'exemple. Les chlorophytes sont représentés par *Micromonas* [MET, image colorisée, Monterey Bay Research Institute] et *Ostreococcus* [MET, Martin F. Hohmann-Marriott]; les dinoflagellés par *Ceratium* [MEB, Philippe Crassous, Ifremer] et *Ceratochoris* [MEB, Maria Faust, Smithsonian Institution]; les diatomées par *Diploneis* [MEB, Mary Ann Tiffany] et *Arachnoidiscus* [MEB, image colorisée, Science Photo Library]; les chrysophytes par *Chromulina* [MO, <http://vietsciences.free.fr/khaocuu/nguyenlandung/vitao02.htm>] et *Dinobryon* [MO, <http://micropolitan.org>]; les haptophytes par *Emiliana* [MEB, image colorisée, Steve Gschmeissner] et *Phaeocystis* [colonie observée au MO, Dr. Peter Countway]; les cryptophytes par *Rhodomonas* [MO, William Bourland] et *Chroomonas* [MO, Gevelsberg Stefansbachtal]; les euglenides par *Eutreptiella* [MO, Fredrik Norén]. Le phytoplankton est également représenté par des organismes procaryotes, au sein du domaine Bacteria, tous inclus au sein de l'embranchement des cyanobactéries. *Trichodesmium* est retrouvé principalement dans les eaux oligotrophes tropicales et sub-tropicales [Abby Heithoff, Woods Hole Oceanographic Institution]. Autres exemples: *Oscillatoria* [MO, <http://silicasecchidisk.conncoll.edu>] et *Cyanothece* [MO, Y. Tsukii]. Les évènements endosymbiotiques sont montrés par des flèches. Arbre modifié d'après Cracraft et Donoghue (2004).



Au sein de la lignée verte (dont font partie les plantes supérieures; Knoll, 1992), la grande majorité des taxons décrits à ce jour dans le milieu marin appartiennent à la classe des Mamiellophyceae, un des groupes écologiquement les plus abondants, comprenant notamment les genres *Micromonas*, *Bathycoccus* et *Ostreococcus* (Vaulot et al., 2008; Marin et Melkonian, 2010). Les chlorophytes forment un groupe fortement soutenu à travers les différentes études phylogénétiques et caractérisé par différentes propriétés biochimiques et ultrastructurales (Leliaert et al., 2011). Tous utilisent la chlorophylle *b* (Chl *b*) comme pigment accessoire principal.

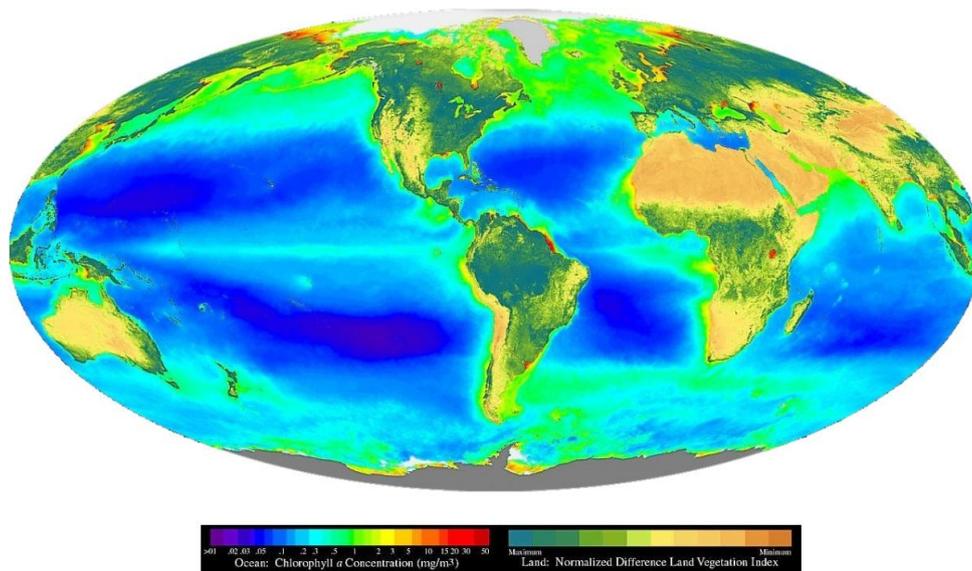
Trois groupes majeurs du phytoplancton sont issus d'une endosymbiose secondaire avec une algue unicellulaire de la lignée rouge, les diatomées, les dinoflagellés et les haptophytes (Fig. 1). Les diatomées constituent un des groupes les plus diversifiés au sein des eucaryotes (Mann et Droop, 1996) et sont caractérisées par une coque siliceuse ou frustule. Les dinoflagellés ont la particularité de posséder deux flagelles orientés perpendiculairement l'un par rapport à l'autre et peuvent être photoautotrophes, hétérotrophes ou mixotrophes. Certains dinoflagellés peuvent avoir un mode de vie parasitaire ou symbiotique, tandis que d'autres provoquent des efflorescences importantes, appelées 'marées rouges', au cours desquelles certaines espèces produisent des toxines pouvant affecter les activités économiques, telles que la pêche (Fig. 2A; Taylor et al., 2008; Hallegraeff, 2010). Les haptophytes présentent eux aussi des modes de vie divers, et arborent différents types structuraux. Ils sont néanmoins caractérisés par la présence d'un appendice particulier, appelé l'haptonème. On retrouve au sein de cet ordre le genre *Phaeocystis*, dont les représentants forment des colonies qui peuvent être à l'origine d'efflorescences planctoniques de grande ampleur (Fig. 2B), ainsi que des espèces calcifiantes ou coccolithophores (Fig. 2C).



**Figure 2 : Efflorescences de phytoplancton.** (A) Marée rouge sur la plage australienne Covelly, proche de Sydney, provoquée par le dinoflagellé *Noctiluca scintillans* [Craig Greenhill, The Telegraph]. (B) Efflorescence de la microalgue haptophyte *Phaeocystis pouchetii* en Péninsule de Gower, se traduisant par l'apparition de colonies flottantes formant un épais mucilage dans l'eau [David John]. (C) Efflorescence de grande ampleur du coccolithophore *Emiliana huxleyi* au large de la Cornouaille [Steve Groom, Plymouth Marine Laboratories].

Cette richesse du phytoplancton, tant du point de vue de l'abondance que de la diversité est restée insoupçonnée pendant de nombreuses décennies, notamment dans les zones centrales des océans, pauvres en Chl *a* et en éléments nutritifs (azote, phosphore et métaux traces), et longtemps

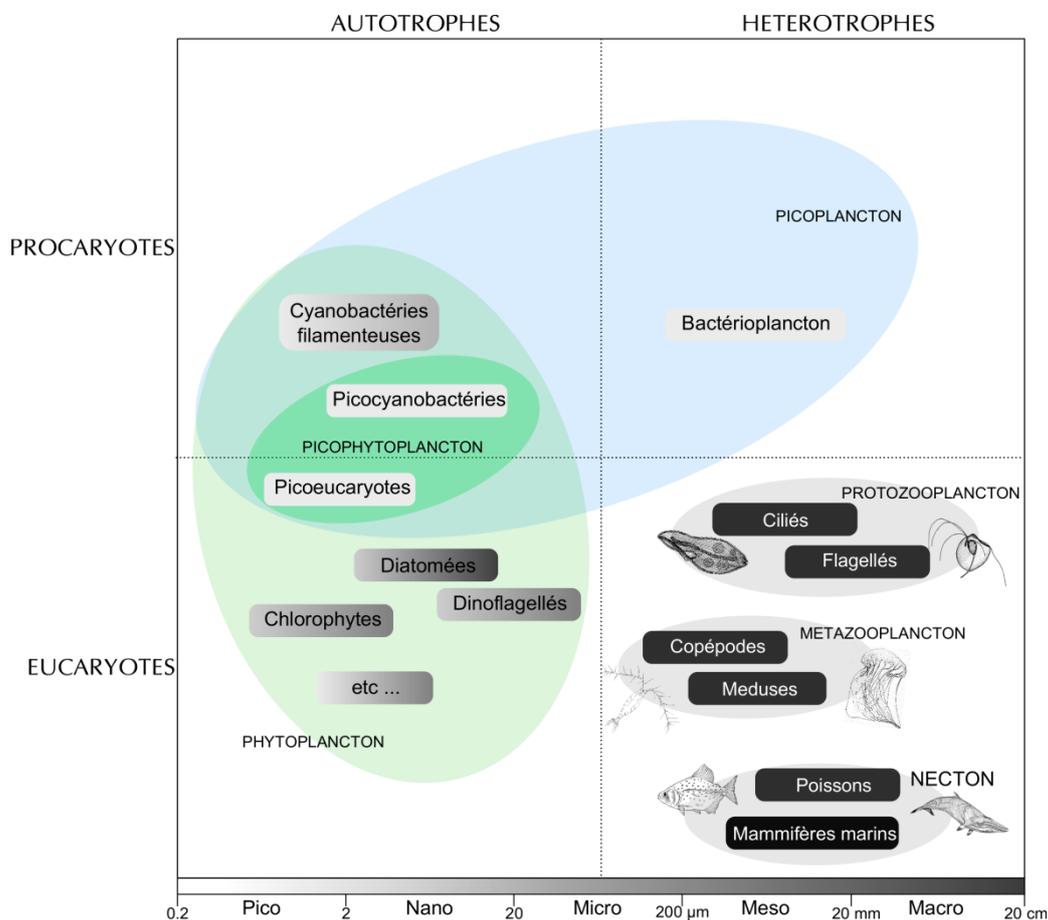
considérées comme de véritables déserts biologiques, alors qu'elles sont en fait peuplées d'une multitude de cellules photosynthétiques de petite taille, constituant le 'picophytoplancton' (Fig. 3). La présence de petites cellules dans le plancton était connue bien avant que le terme 'picoplancton' soit proposé. La première description d'une algue microscopique, *Stichococcus bacillaris*, remonte à 1849 (Nägeli, 1849). Au cours du XXème siècle, les scientifiques prirent conscience que des organismes bien plus petits que ce qui pouvait être capturé à l'aide de filets existaient dans les océans (Lohmann, 1908, 1911), menant à la description de la première « pico » cyanobactérie, *Synechocystis salina*, en 1924 (Wislough, 1924). La microalgue chlorophyte *Micromonas pusilla* fut quant à elle décrite en 1952, sous le nom initial de *Chromulina pusilla* (Butcher, 1952). Néanmoins, il faut attendre la fin des années 1960 pour que l'importance écologique des microorganismes soit enfin suggérée par quelques scientifiques pionniers (Wauthy et al., 1967; Jeffrey, 1976). Par la suite, Sieburth et collaborateurs proposent une classification afin de caractériser les différentes classes de taille que l'on peut retrouver au sein de la fraction planctonique (Fig. 4; Sieburth et al., 1978).



**Figure 3** : Magnitude et distribution de la production primaire, à la fois océanique (exprimée en  $\text{mg}\cdot\text{m}^{-3}$  de chlorophylle  $a$ ) et terrestre (index de végétation normalisé par rapport à la surface). Données provenant du SeaWiFS Project, NASA/Goddard Space Flight Center & ORBIMAGE.

Certains développements technologiques de la fin des années 1970, notamment l'avènement de la microscopie à épifluorescence (Hobbie et al., 1977), permettront de mettre en évidence l'abondance des microorganismes phytoplanctoniques de très petite taille, inférieure à  $2\ \mu\text{m}$ , appelé picoplancton (Johnson et Sieburth, 1979; Waterbury et al., 1979). En particulier, cette technique a révélé l'abondance de cellules possédant de la phycoérythrine (PE) comme principal pigment accessoire, morphologiquement proches d'organismes déjà décrits dans les eaux douces et assignées au genre *Synechococcus*. En 1986, John Waterbury et collaborateurs firent la première description détaillée de la diversité, biologie et écologie de ces cellules, grâce aux premières cultures de souches de

*Synechococcus* (Waterbury et al., 1986). A la fin des années 1980, une autre révolution technologique vint bouleverser l'étude du picophytoplancton. Initialement développée pour des applications cliniques dans le domaine médical, l'application de la cytométrie en flux au domaine de l'océanographie marqua en effet un réel tournant. Cette technologie permet l'analyse en temps réel des signaux de fluorescence et de diffusion optique émis par des particules individuelles. Elle a notamment permis de détecter une autre cyanobactérie marine, dépourvue de phycobiliprotéines (PBPs), et qui en microscopie à épifluorescence ne présentait aucune fluorescence orange et une fluorescence rouge si faible et si labile qu'elle était difficilement distinguable des bactéries hétérotrophes. C'est ainsi qu'en 1988, l'importance écologique de *Prochlorococcus* fut démontrée par Chisholm et collaborateurs (Chisholm et al., 1988). Ce microorganisme photosynthétique est actuellement considéré comme étant le plus abondant sur Terre (Partensky et al., 1999b).



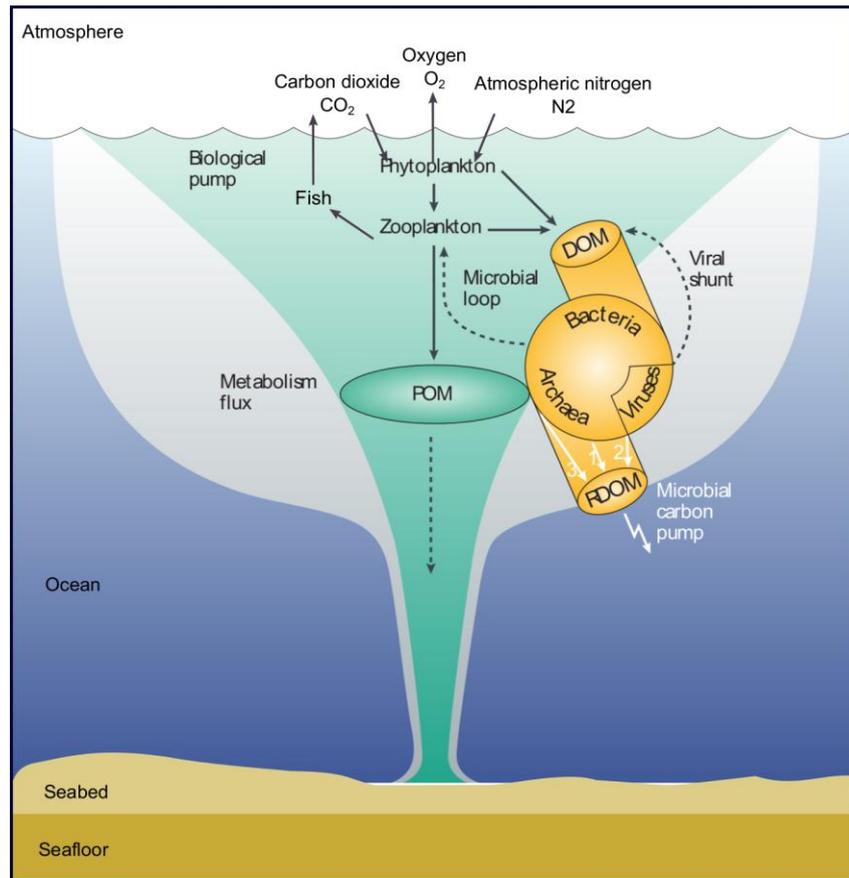
**Figure 4 :** Classifications du plancton selon leur taille, leur mode de nutrition et leur règne. Le plancton par opposition au necton désigne les organismes incapables de se mouvoir, dérivant donc avec les courants. Le plancton peut être autotrophe, et se nourrir par réduction de la matière inorganique, ou hétérotrophe et assimiler directement la matière organique. Le plancton, à la fois procaryote et eucaryote, est divisé selon différentes classes de taille, proposées par Sieburth et al. (1978).

### I.1.2 Rôle du phytoplancton dans la pompe biologique

De par leur abondance, les organismes phytoplanctoniques jouent un rôle essentiel dans les grands cycles biogéochimiques de la planète, notamment celui du carbone puisqu'ils sont capables de le fixer sous forme de CO<sub>2</sub> et de le transformer en composés organiques carbonés en utilisant la lumière. Cette réaction chimique nécessite de casser une molécule d'eau pour fixer une molécule de CO<sub>2</sub> et ainsi libérer une molécule d'oxygène. C'est la photosynthèse dite oxygénique, qui diffère de la photosynthèse anoxygénique réalisée par certaines bactéries unicellulaires et filamenteuses mais ne donnant pas lieu à une émission d'oxygène. Contrairement au milieu terrestre où elle est principalement effectuée par les plantes supérieures, la photosynthèse en milieu océanique est réalisée en grande partie par des microorganismes. Près de la moitié de la production primaire globale de la planète a lieu au sein des masses d'eau océaniques (49.4%), notamment dans les eaux mésotrophes (34.6% de la production totale; Field et al., 1998; Carr et al., 2006). La production primaire au niveau des milieux oligotrophes (9.2%) serait supérieure à celles des milieux eutrophes (5.6%). Par adsorption à de larges particules, cette matière organique, alors appelée neige marine, peut couler et être séquestrée pour une partie d'entre elle (2%) au niveau du plancher océanique. Ce phénomène contribue ainsi à la baisse du taux de dioxyde de carbone dans l'atmosphère, et est à l'origine d'une partie de l'énergie fossile que nous utilisons aujourd'hui (Hedges et al., 2001).

L'essentiel de cette matière organique produite à la suite de la photosynthèse oxygénique se retrouve sous forme dissoute, appelée DOM (pour Dissolved Organic Matter). La DOM contient à la fois du carbone (DOC pour Dissolved Organic Carbon) mais aussi les principaux éléments que sont l'azote, le phosphore et le soufre (Biddanda et Benner, 1997). La DOM n'est presque exclusivement accessible qu'aux archées et aux bactéries hétérotrophes. Cette observation a conduit les scientifiques, au début des années 80, à distinguer la chaîne alimentaire classique (allant des grosses cellules de phytoplancton et zooplancton aux poissons) de celle représentée par les microorganismes autotrophes, appelée boucle microbienne (Azam et al., 1983). Les microbes ayant assimilé ces éléments dissous sont ensuite lysés par la composante virale de l'océan (Suttle, 2007) ou broutés par les protistes. Ces derniers servent ainsi de passerelle vers les niveaux supérieurs de la chaîne alimentaire, une fois consommés par les métazoaires (Azam et al., 1983).

Le rôle du phytoplancton dans cette boucle microbienne est particulièrement important dans les environnements oligotrophes. En effet, il est désormais bien établi que ces régimes océaniques pauvres en nutriments sont dominés par le picophytoplancton, composé à la fois par de petits eucaryotes (Moon-van der Staay et al., 2001; Vaulot et al., 2008) et des deux genres cyanobactériens *Prochlorococcus* et *Synechococcus* qui dominent, en nombre de cellules, ces larges étendues marines (voir §1.2.3).



**Figure 5 : Rôle de la boucle microbienne dans le recyclage de la matière organique au niveau océanique.** Le dioxyde de carbone, fixé dans les couches supérieures de l'océan par le phytoplancton, est transporté vers l'océan profond, par l'intermédiaire des particules biogéniques (POM pour Particulate Organic Matter) ou est transformé en matière organique dissoute (DOM). Cette dernière est en partie (5-10%) assimilable par les bactéries et archées hétérotrophes, et sera donc transférée aux niveaux trophiques supérieurs. C'est l'ensemble de ces processus que l'on appelle la boucle microbienne. La majeure partie de la matière organique dissoute n'est cependant pas reminéralisée (RDOM pour Recalcitrant Dissolved Organic Matter). Trois processus majeurs ont ainsi été identifiés au niveau de la pompe à carbone : la production d'exudats lors de la production et prolifération microbienne (voie 1), la lyse virale entraînant le relargage de parois cellulaires ou macromolécules de surface (voie 2) et la dégradation de la POM (voie 3). Le dégradé gris indique le flux total de carbone au sein de la colonne d'eau. Modifié d'après Jiao et al. (2010).

## I.2. Les picocyanobactéries marines

### I.2.1 Origine et évolution des cyanobactéries marines

Au sein des eubactéries, les cyanobactéries représentent un groupe évolutif monophylétique (Stanier et al., 1971; Castenholz et Waterbury, 1989; Wilmotte, 2004). Les nombreuses phylogénies réalisées à l'aide de différents marqueurs montrent que ces organismes auraient été un des premiers groupes à diverger au sein de la lignée bactérienne ancestrale (Blank, 2004). Ce groupe a en effet une origine très ancienne et serait apparu juste après les bactéries mésophiles réductrices de sulfates (Blank, 2004). L'origine évolutive des cyanobactéries reste cependant sujette à controverse. Diverses études utilisant des traces fossiles, ont proposé que les cyanobactéries seraient apparues il y a 3.5 Ga (Schopf et Packer, 1987; Schopf, 1993). Une analyse plus récente combinant des données

paléobiologiques et des comparaisons phylogénétiques entre de nombreuses souches cyanobactériennes propose que ce groupe daterait plutôt de la fin de l'archéen, il y a 2.7 Ga (Ga, Blank et Sánchez-Baracaldo, 2010). Par ailleurs, une origine marine de ce groupe a longtemps été envisagée du fait que les plus anciennes traces fossiles, notamment les stromatolithes, ont toutes été identifiées dans des écosystèmes marins ou saumâtres (Fig. 6; Honda et al., 1999; Bolhar et Van Kranendonk, 2007).

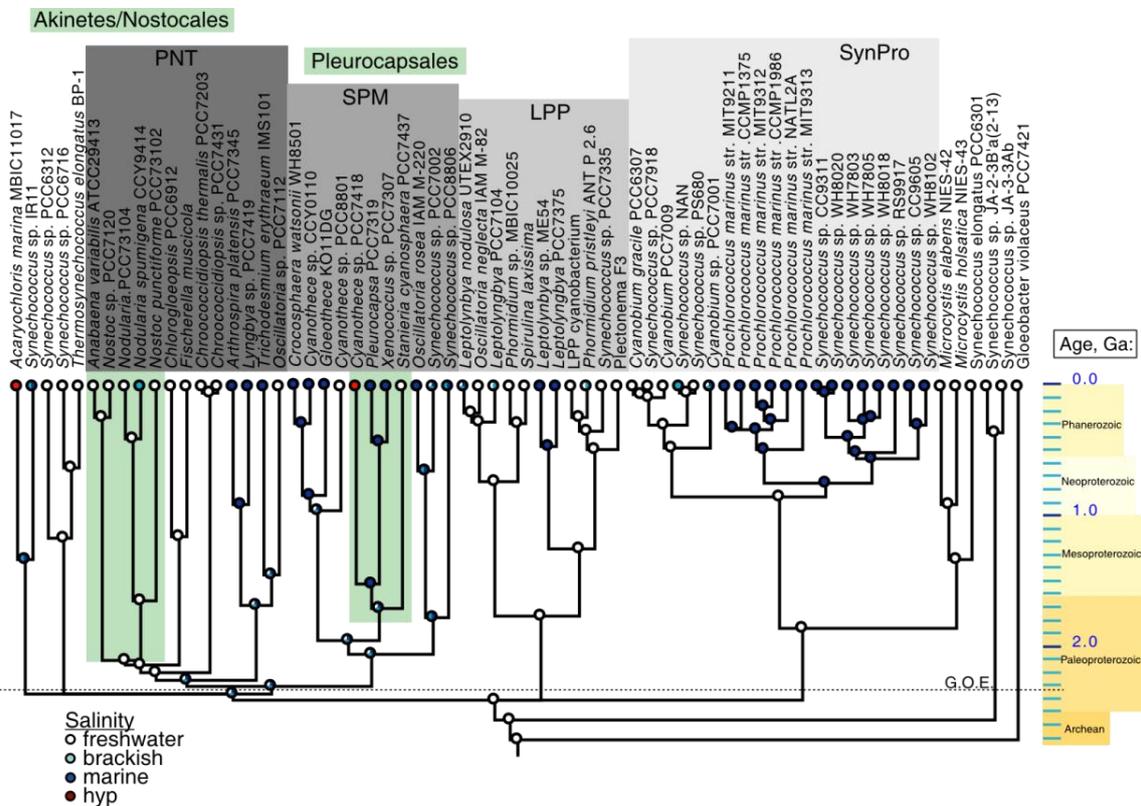


---

**Figure 6** : Colonnes stromatolithiques de la baie Shark sur le littoral ouest de l'Australie. Ces structures calcaires sont dites biogéniques car elles sont le résultat de la construction par des communautés bactériennes, principalement des cyanobactéries.

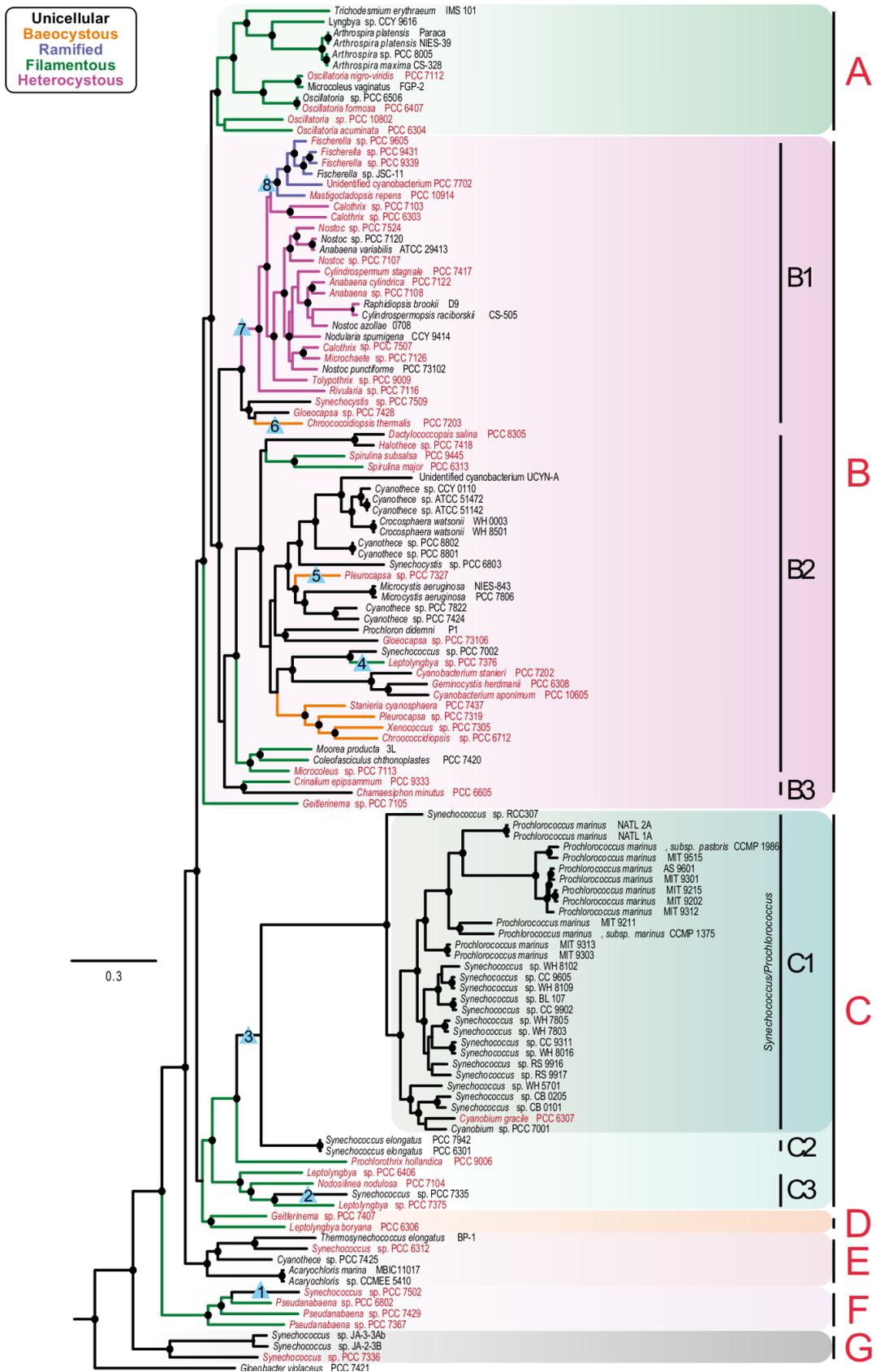
Ces structures fossiles datent effectivement de la première partie de l'archéen mais, d'après Blank et Sánchez-Baracaldo (2010), elles auraient en fait été formées par des procaryotes phototrophes anoxygéniques, chimioautotrophes ou par des procédés abiotiques. Ces auteurs proposent donc, sur des bases phylogénétiques, que l'ancêtre commun des cyanobactéries serait plutôt apparu dans un environnement terrestre dulçaquicole et/ou endolithique. De plus, cet ancêtre était probablement unicellulaire, avec une taille cellulaire inférieure à 2.5  $\mu\text{m}$  (Battistuzzi et al., 2004; Sánchez-Baracaldo et al., 2005). Les premières diversifications au sein de la lignée cyanobactérienne seraient ensuite apparues durant le Protérozoïque, avant l'augmentation de la concentration en oxygène atmosphérique (Fig. 7). Ces diversifications auraient conduit à la colonisation des écosystèmes de plus forte salinité, estuariens puis océaniques, conduisant à accroître de façon importante la taille de la niche écologique disponible pour la photosynthèse oxygénique. La colonisation de ces espaces côtiers, riches en nutriments, aurait conduit à l'augmentation des flux de carbone et d'oxygène. Extrêmement faible avant 2.45 Ga, l'oxygénation de l'atmosphère primitive aurait débuté il y a 2.32 Ga, pour atteindre des niveaux importants il y a 2.22 Ga (Bekker et al., 2004). En modifiant les grands cycles biogéochimiques, ce phénomène a conduit à l'émergence de nouveaux environnements, terrestres et aquatiques, rapidement colonisés par des organismes utilisant la respiration oxygénique (Farquhar et al., 2000; Kirschvink et al., 2000; Melezhik et al., 2005).

Différents évènements de colonisation du milieu marin se seraient produits, de manière indépendante, au cours de l'histoire évolutive des cyanobactéries (Sánchez-Baracaldo et al., 2005). Plusieurs lignées actuelles attestent du développement de cette tolérance aux fortes salinités comme les Nostocales, le clade formé par *Phormidium/Pseudanabaena*, le clade formé par *Gleothoece/Cyanothece/Crocospaera/Synechocystis/Microcystis* ainsi que le groupe des cyanobactéries planctoniques *Synechococcus/Prochlorococcus/Cyanobium* (Fig. 7). Phylogénétiquement très proches, ces trois derniers genres auraient divergé à partir d'un ancêtre commun avec les autres cyanobactéries, il y a environ 600 millions d'années. Cette branche serait donc plus ancienne que l'avaient suggéré Rappé et Giovanonni (Fig. 7; Rappé et Giovanonni, 2003; Sánchez-Baracaldo et al., 2005; Blank et Sánchez-Baracaldo, 2010).



**Figure 7 : Chronogramme de reconstruction de l'état ancestral par horloge moléculaire en utilisant un arbre RpoC-SSU.** Les branches indiquées sont proportionnelles à l'âge, l'échelle étant indiquée en milliards d'années (Ga). L'oxygénation atmosphérique est indiquée par la ligne en pointillés. Les caractères ancestraux liés à la tolérance à la salinité sont indiqués par des cercles de couleur. D'après Blank and Sánchez-Baracaldo (2010).

Du point de vue taxonomique, les genres *Synechococcus/Prochlorococcus/Cyanobium* font partie de la classe des *Cyanophyceae* décrite par Schaffner en 1909, et font partie de l'ordre des *Chroococcales*, dont tous les représentants se divisent par fission binaire (Herdman et al., 2001). À noter cependant que Guiry (2012), dans le cadre du 'World Register of Marine Species', a proposé une nouvelle classification dans laquelle ce groupe a été placé dans le nouvel ordre des '*Synechococcales*'.



**Figure 8** : Arbre en Maximum de vraisemblance (ML) basé sur l'analyse de séquences codant pour l'ARN 16S des espèces cyanobactériennes dont le génome a été entièrement séquencé. Les couleurs des branches se réfèrent aux caractères morphologiques. Les taxons indiqués en rouge ont été séquencés dans le cadre du projet CyanoGEBa. Les valeurs de bootstraps supérieures à 70% sont indiquées par un cercle noir. Les cyanobactéries sont divisées en sept grands clades (A à G), dont certains ont été divisées en sous-clades. Modifié d'après Shih et al. (2013).

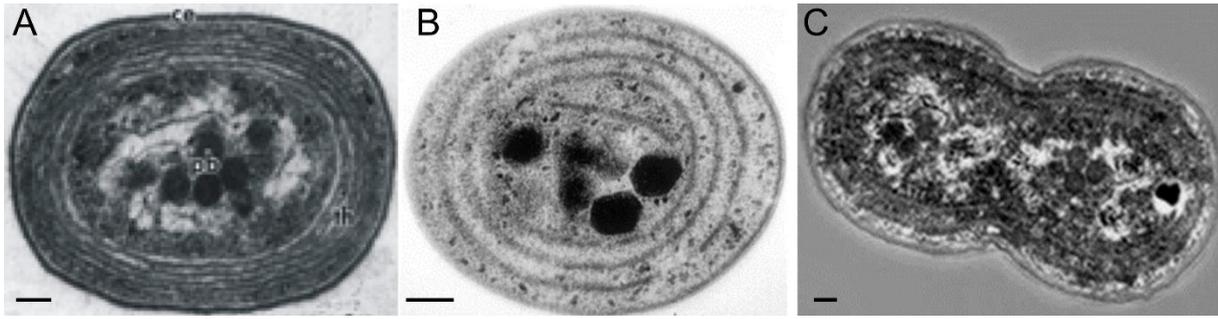
Le genre *Synechococcus* qui comprend à la fois des souches marines et d'eau douce, est présent au sein de différentes radiations et apparaît clairement polyphylétique (Fig. 8; Honda et al., 1999; Robertson et al., 2001; Fuller et al., 2003; Gupta et Mathews, 2011). Néanmoins, des analyses phylogénétiques, basées sur le gène de l'ARNr 16S et des séquences concaténées de 44 protéines conservées chez les cyanobactéries, ont permis de montrer que les *Synechococcus* marins forment (avec *Prochlorococcus*) une radiation bien différenciée des *Synechococcus* d'eau douce (Fuller et al., 2003; Gupta et Mathews, 2011). Initialement défini à partir de critères morphologiques, ce nom de genre n'est donc clairement plus adapté et mériterait une révision taxonomique.

### I.2.2 Morphologie et organisation cellulaire des picocyanobactéries

Les picocyanobactéries marines sont de forme ovoïde et se divisent toujours par fission binaire, perpendiculairement au grand axe de la cellule. La taille cellulaire de *Synechococcus* est comprise entre 0,6  $\mu\text{m}$  et 2  $\mu\text{m}$  (Waterbury et al., 1986). Les souches de surface de *Prochlorococcus* possèdent une forme coccoïde, avec un diamètre de 0,7  $\mu\text{m}$ , mais certaines, notamment celles du clade LLIV qui vivent plus bas dans la colonne d'eau (voir ci-dessous), sont plus de plus grande taille (Morel et al., 1993; Lichtlé et al., 1995; Rippka et al., 2000a; Ting et al., 2007). Ceci est en adéquation avec les observations réalisées *in situ*, démontrant une augmentation de la taille des cellules de *Prochlorococcus* avec la profondeur (Sieracki et al., 1995), le plus faible rapport surface/volume des souches de surface leur permettant de maximiser la capture de nutriments.

Comme toutes les cyanobactéries, à l'exception de *Gloeobacter violaceus*, *Synechococcus* et *Prochlorococcus* possèdent des membranes thylacoïdales, où se produisent les réactions de la photosynthèse (Schneider et Jürgens, 1991). Ces membranes forment des cercles concentriques, parallèles à la membrane cytoplasmique. L'aspect des thylacoïdes diffère néanmoins entre les deux genres, puisqu'ils sont accolés chez *Prochlorococcus* alors qu'un large espace est visible entre ces structures chez *Synechococcus* (Fig. 9; Lichtlé et al., 1995). Le nombre de ces membranes diffère également entre les souches. Chez *Prochlorococcus*, peu de membranes thylacoïdales sont observées chez la souche de surface MED4, comparé à la souche de profondeur MIT9313 (Ting et al., 2007). Des canaux ont été mis en évidence entre ces membranes thylacoïdales et pourraient participer au maintien d'un gradient électrochimique uniforme (Ting et al., 2007).

D'autres structures cellulaires également indispensables à la photosynthèse et spécifiques des cyanobactéries se trouvent dans le cytoplasme: les carboxysomes (Kerfeld et al., 2005; Kerfeld et al., 2010). De forme polyédrique et d'environ 100 nm de diamètre (Figs. 9 et 10; Stanier et Cohen-Bazire, 1977; Iancu et al., 2007), ces derniers servent de lieu de stockage de la Rubisco (ribulose 1,5-bisphosphate carboxylase / oxygenase), enzyme essentielle pour la fixation du  $\text{CO}_2$  à travers la voie réductrice du cycle des pentoses phosphates.



**Figure 9:** Photographies en microscopie électronique de cellules de (A) *Prochlorococcus* et (B, C) *Synechococcus*. La cyanobactérie marine *Synechococcus* a été décrite pour la première fois en 1979 (Waterbury et al., 1979). Le genre *Prochlorococcus* a été découvert en 1988 par cytométrie en flux (Chisholm et al., 1992). Néanmoins des photographies des communautés picoplanctoniques *in situ* avaient déjà rapporté son existence dès 1979 (Johnson et Sieburth, 1979). (C) Les picocyanobactéries marines se divisent toujours par fission binaire perpendiculairement au grand axe de la cellule. Images de (A) Johnson et Sieburth (1979), (B) du Dr T. Beveridge et (C) de The Lawrence Berkeley National Laboratory. Echelle : 0.1  $\mu\text{m}$ .

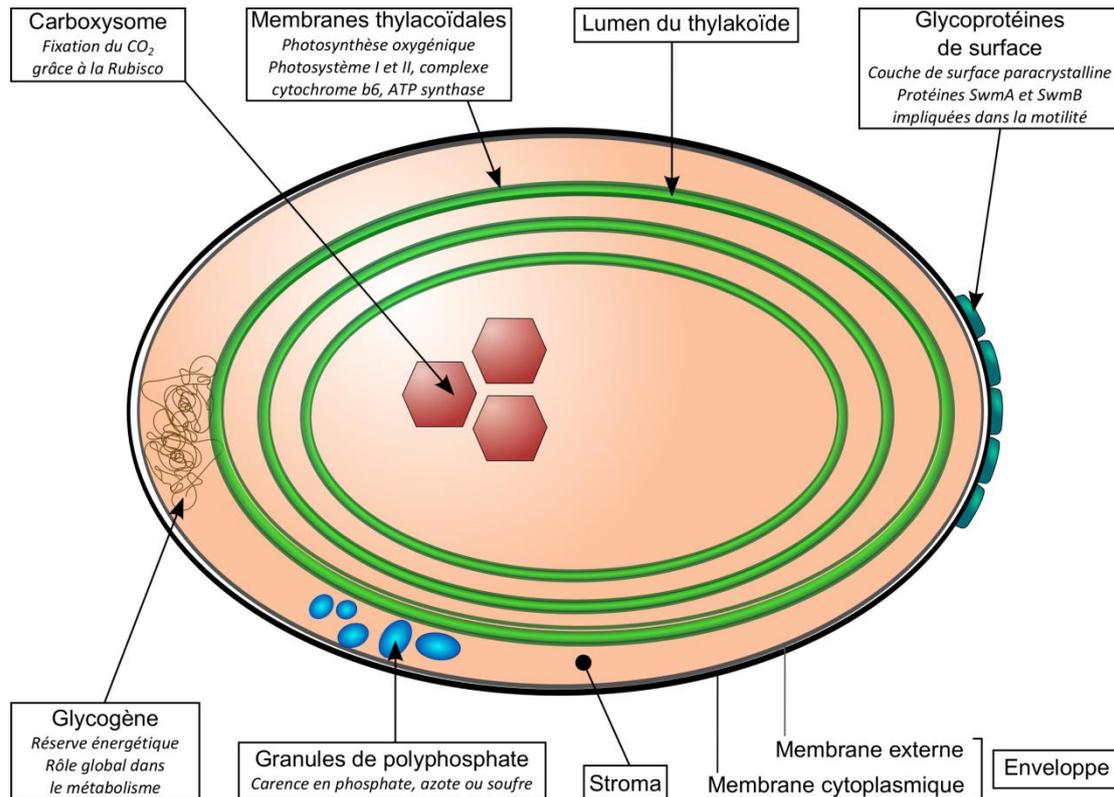
Comme toutes les bactéries à Gram négatif, *Synechococcus* et *Prochlorococcus* possèdent une membrane cytoplasmique et une membrane externe, qui forment l'enveloppe cellulaire (Fig. 10). Les Gram positif, en revanche, ne possèdent qu'une membrane plasmique et une couche épaisse de peptidoglycane. Cependant, malgré cette similarité structurale avec les Gram négatif, l'enveloppe cellulaire des cyanobactéries partagent certaines caractéristiques avec les Gram positif, notamment l'épaisseur de la couche de peptidoglycane (Hoiczky et Hansel, 2000). Généralement fine ( $\sim 2$  nm chez la plupart des Gram négatif), elle peut être très épaisse chez certaines cyanobactéries. Chez *Synechococcus elongatus*, elle mesure 10 nm (Golecki, 1977), et peut atteindre 15-16 nm chez *Synechococcus* sp. WH8102/WH8113 (Samuel et al., 2001; Ting et al., 2007). Une autre étude rapporte toutefois que celle-ci n'atteindrait que 6 nm chez *Synechococcus* sp. WH8102 (McCarren et al., 2005). Certaines caractéristiques chimiques, telles que le degré d'entrecroisement entre les chaînes de peptidoglycane et la muréine (Jürgens et al., 1983), la structure chimique des pentapeptides impliqués (Kodani et al., 1999), ainsi que l'arrangement des lipopolysaccharides avec le peptidoglycane sont plus proches de ce que l'on rencontre chez les Gram positif (Jürgens et Weckesser, 1986; Jürgens et Burger-Wiersma, 1989; Ting et al., 2007).

Chez certaines souches de *Synechococcus*, des appendices de surface (de type spinae) de longueur variable ont été rapportés à partir d'isolats d'eau douce et de milieu côtier (Perkins et al., 1981; Schultze-Lam et Beveridge, 1994; McCarren et al., 2005). Des structures similaires ont été observées chez *Cyanobium*, qui pourraient permettre à ces cellules de s'aggréger afin de se protéger de la prédation par les brouteurs (Jezberová et Komárková, 2007). Des composants de type polysaccharidique sont aussi présents chez les *Synechococcus* marins, à la fois sur des souches isolées en milieu côtier ou dans l'océan ouvert. Des variations dans la composition en ces composés entraînent également différentes susceptibilité à la prédation selon les souches (Snyder et al., 2009; Zwirgmaier et al., 2009; Apple et al., 2011).

Enfin chez les souches motiles, comme *Synechococcus* sp. WH8102 ou WH8113, une membrane additionnelle de surface est systématiquement présente (Samuel et al., 2001; McCarren et al., 2005). Cette dernière possède les propriétés d'une couche S en formant une couche paracrystalline tout autour de la cellule (S-layer; Šmarda et al., 2002). Chez *Synechococcus* sp. WH8113, de fines fibrilles ou des spicules émanant de cette membrane ont été observés (Samuel et al., 2001). Initialement il avait été proposé que ces structures fibrillaires auraient un rôle dans la transmission du mouvement, mais n'ayant jamais été observées chez *Synechococcus* sp. WH8102, celles-ci ne semblent donc pas indispensables à la motilité.

Chez les *Synechococcus* marins, seules les cellules du clade III (voir Fig. 13) sont capables de se mouvoir (Toledo et al., 1999). En dépit d'une apparente ressemblance à la motilité par nage (Adams, 2001), aucun flagelle n'a pu être mis en évidence (Waterbury et al., 1985; Brahmsha, 1996). En revanche, cette capacité est perdue chez les mutants du gène *swmA* qui sont dépourvus de couche S, suggérant que cette membrane est directement ou indirectement (site d'attachement d'autres composants membranaires) impliquée dans la motilité (McCarren et al., 2005). La couche S est ainsi formée par deux glycoprotéines de surface, très abondantes dans les cellules motiles, SwmA et SwmB, d'un poids moléculaire respectif de 130 kDa et 1.12 MDa (Brahmsha, 1996; McCarren et Brahmsha, 2007, 2009). Si la motilité chez *Synechococcus* reste encore un mystère sur de nombreux points, ce phénomène serait sous l'influence d'une force proton-motrice, et le mouvement se déplacerait au sein de l'enveloppe cellulaire le long d'une « piste » hélicoïdale continue en boucle (Ehlers et Oster, 2012).

Chez certaines cyanobactéries filamenteuses, de l'ordre des *Nostocales* exposées aux changements de leur milieu environnant, notamment à des carences nutritives, on observe des réponses au niveau développemental, incluant des différenciations cellulaires telles que des hétérocystes, des akinètes ou des hormogonies. Ce type de cellules différenciées n'a cependant jamais été observé chez *Synechococcus* et *Prochlorococcus*.



**Figure 10** : Représentation schématique des différents compartiments cellulaires chez les picocyanobactéries. L'enveloppe cellulaire des picocyanobactéries marines présente certaines caractéristiques spécifiques assurant sa singularité. Les thylakoïdes sont composées de deux membranes cellulaires, l'espace situé entre ces membranes correspondant à la lumière thylacoïdale, où sont générés les protons issus de la photosynthèse.

Quatre types d'inclusion sont fréquemment trouvées chez les cyanobactéries soumises à un stress. Des granules de polyphosphate peuvent se former en cas de carence en phosphate, en azote ou en soufre (Lawry et Jensen, 1979, 1986). Cette capacité d'accumulation est présente au sein de la radiation marine des *Synechococcus*, mais n'est pas partagée par toutes les souches en culture. En effet, les granules de phosphate sont plus fréquemment observés et importants au niveau des isolats des régions côtières (clades I et IV notamment), soumis à des variations plus importantes des concentrations en nutriments (Mazard et al., 2012b). Comme les autres bactéries, *Synechococcus* et *Prochlorococcus* accumulent des polysaccharides de stockage, comme du glycogène, durant la journée, de manière à fournir une source de carbone importante et une énergie suffisante pour supporter l'activité respiratoire nocturne (Stanier et Cohen-Bazire, 1977; Lichtlé et al., 1995). Cependant les réserves de glycogène ne serviraient pas uniquement de réserve énergétique, mais auraient également un rôle plus global dans le métabolisme, notamment en réponse au stress oxydatif ou osmotique (Suzuki et al., 2010). L'expression du gène impliqué dans la synthèse de glycogène diffère d'un point de vue temporel entre *Synechococcus* et *Prochlorococcus*, et reflèterait de stratégies distinctes d'assimilation du carbone et de l'azote au cours du cycle journalier (Zinser et al., 2009; Wyman et Thom, 2012).

### I.2.3 Distribution et dynamique des picocyanobactéries marines

Les picocyanobactéries marines jouent un rôle majeur dans les grands cycles biogéochimiques océaniques, en dépit de leur faible contenu en carbone, estimé entre 200 à 250 fg/cellule pour *Synechococcus* et 50 à 130 fg/cellule pour *Prochlorococcus* (Li, 1994; Scanlan et West, 2002; Richardson et Jackson, 2007; Buitenhuis et al., 2012). *Prochlorococcus* représenterait à lui seul jusqu'à 73% du carbone photosynthétique dans la couche de surface mélangée au large d'Hawaï (Campbell et al., 1994; Partensky et al., 1999b). La biomasse globale de *Synechococcus* à l'échelle mondiale a été estimée à 111 Mt C et celle de *Prochlorococcus* à 126 Mt C (Buitenhuis et al., 2012). Des mesures directes ont par ailleurs établi que ces deux genres contribuaient à plus de 25% de la production primaire dans de nombreux environnements océaniques (Li, 1994), mais aucune estimation de leur contribution à la production primaire globale n'a été publiée à ce jour.

Depuis plus de trente ans, de nombreuses études ont permis d'identifier les principaux paramètres physico-chimiques affectant la distribution des populations naturelles de *Prochlorococcus* et *Synechococcus*. En effet, les diverses campagnes océanographiques ont permis de couvrir une large diversité de régimes océaniques (Zubkov et al., 2000; Heywood et al., 2006). D'autre part, la dynamique temporelle de ces populations a également été étudiée à des stations de suivis à long terme telle que la station BATS dans l'Atlantique Nord (Bermuda Atlantic Time-series Study; DuRand et al., 2001; Steinberg et al., 2001) ou HOT dans le Pacifique N (Hawaii Ocean Time-series; (Campbell et Vaulot, 1993; Karl et Lukas, 1996; Campbell et al., 1997; Karl et al., 2001).

*Prochlorococcus* est confiné entre les latitudes 48°N et 40°S (Buck et al., 1996; Partensky et al., 1999a; Partensky et al., 1999b; Johnson et al., 2006; Zwirgmaier et al., 2007). En effet, ce genre est particulièrement bien adapté aux eaux chaudes oligotrophes (Zubkov et al., 2000), et aux masses d'eau fortement stratifiées, notamment durant l'été en zone tempérée (Olson et al., 1990a; Campbell et Vaulot, 1993; Lindell et Post, 1995; DuRand et al., 2001). Cette cyanobactérie présente des concentrations moyennes comprises entre  $10^4$  et  $10^5$  cellules.mL<sup>-1</sup> (Partensky et al., 1999b), mais dépasse fréquemment  $10^5$  cellules.mL<sup>-1</sup> (Olson et al., 1990a; Campbell et Vaulot, 1993; Zubkov et al., 2000; Johnson et al., 2006; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008), atteignant même  $7.10^6$  cellules.mL<sup>-1</sup> au niveau de la mer d'Arabie (Campbell et al., 1998).

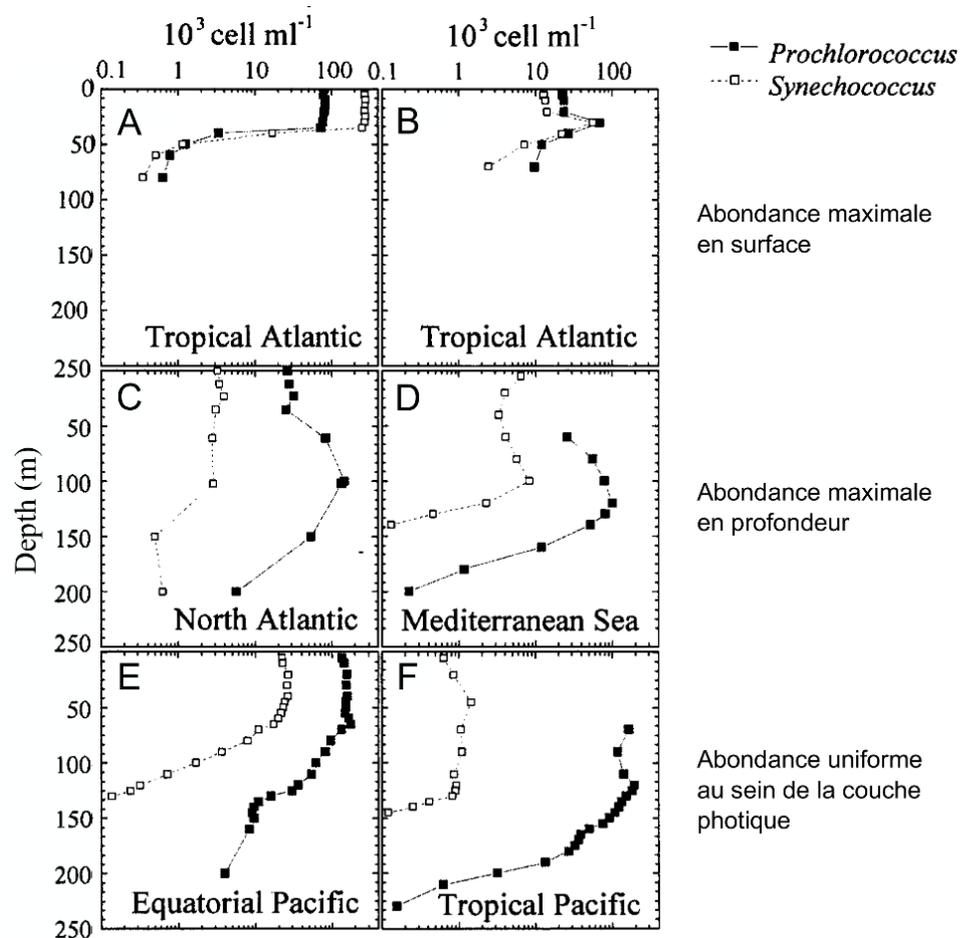
*Synechococcus* est quant à lui beaucoup plus ubiquiste et présent dans tous les environnements océaniques, des régimes estuariens et côtiers à l'océan ouvert, et de l'équateur jusqu'aux eaux polaires arctiques (Partensky et al., 1999a). Comme les picoeucaryotes, *Synechococcus* semble mieux adapté aux zones plus riches en sels nutritifs (régimes eutrophes et

mésotrophes) et supportent une amplitude thermique beaucoup plus importante (Zwirgmaier et al., 2008). Les concentrations moyennes de cette picocyanobactérie, bien que plus faibles que *Prochlorococcus*, sont généralement comprises entre  $10^3$  et  $10^4$  cellules.mL<sup>-1</sup> (Partensky et al., 1996; Partensky et al., 1999a; Zhao et al., 2007b; Zwirgmaier et al., 2007; Tai et Palenik, 2009; Mella-Flores et al., 2011; Post et al., 2011). Si *Synechococcus* est capable de former des efflorescences lorsque les eaux sont mélangées, son abondance n'excède cependant que très rarement  $10^5$  cellules.mL<sup>-1</sup> notamment en mer d'Arabie (Fuller et al., 2006), au niveau des remontées d'eau froide ('*upwelling*') à l'ouest de l'Afrique (Partensky et al., 1996; Zubkov et al., 2000; Zwirgmaier et al., 2007), en Polynésie Française (Charpy et Blanchot, 1998) ou le long des côtes américaines (Waterbury et al., 1979). Les plus fortes concentrations cellulaires ont été rapportées au niveau des remontées d'eau froide du Costa Rica, avec des valeurs généralement comprises entre  $1.5$  et  $3.7 \times 10^6$  cellules.mL<sup>-1</sup> (Li et al., 1983; Saito et al., 2005).

*Synechococcus* est retrouvé dans des environnements couvrant une large gamme de température, depuis 2°C à plus de 30°C (Shapiro et Haugen, 1988; Fuller et al., 2006; Zwirgmaier et al., 2008). Néanmoins, sa concentration est fortement influencée par la température et diminue de plusieurs ordres de magnitude en dessous de 14°C (Li, 1998; Fouilland et al., 1999; Wilmotte et al., 2002) jusqu'à atteindre environ  $10^2$  cellules au niveau de l'Arctique (Waleron et al., 2007; Cai et al., 2010). Au delà du cercle polaire, les abondances de *Synechococcus* sont généralement négligeables (Gradinger et Lenz, 1989, 1995), et seraient principalement d'origine allochtone, provenant plutôt d'écosystèmes d'eau douce (Waleron et al., 2007). Cependant une étude récente menée en mer de Chukchi souligne que certaines populations seraient autochtones et adaptées aux températures polaires (Huang et al., 2012).

Les distributions verticales de *Synechococcus* et *Prochlorococcus* sont également contrastées. Elles peuvent être classées en trois grandes catégories en fonction du régime trophique de l'environnement océanique, ainsi que du degré de mélange de la colonne d'eau. On distingue ainsi les eaux où les abondances maximales sont mesurées en surface et diminuent rapidement en dessous la thermocline (Fig. 11A-B). Ce type de distribution prédomine en milieu mésotrophe, où *Synechococcus* peut être plus abondant que *Prochlorococcus*. En milieu océanique tempéré, les abondances picocyanobactériennes sont généralement maximales en profondeur (Fig. 11C-D). Au niveau de l'équateur, où les eaux sont riches en azote et phosphore mais carencées en fer, l'abondance des deux groupes est maximale dans la couche mélangée, puis diminue en profondeur mais plus progressivement qu'en zone mésotrophe typique (Fig. 11E). Enfin, dans les environnements océaniques tropicaux et subtropicaux oligotrophes, le cas le plus fréquent, les

abondances des deux groupes sont assez uniformes sur une couche de près de 100 m avant de diminuer fortement vers le bas de la couche euphotique (Fig. 11C). Au-delà des abondances moyennes rencontrées en milieu océanique, généralement 1 à 2 ordres de magnitude plus élevées pour *Prochlorococcus* que *Synechococcus*, il est surtout intéressant de noter que le premier est retrouvé en dessous 150 m, profondeur en dessous de laquelle *Synechococcus* est en général indétectable, même en milieu oligotrophe. *Prochlorococcus* a ainsi développé des mécanismes lui permettant de croître à de très faibles niveau d'éclairement, de l'ordre de 0.1% du flux de photons incident en surface (voir §II.2.4, Partensky et al., 1996).



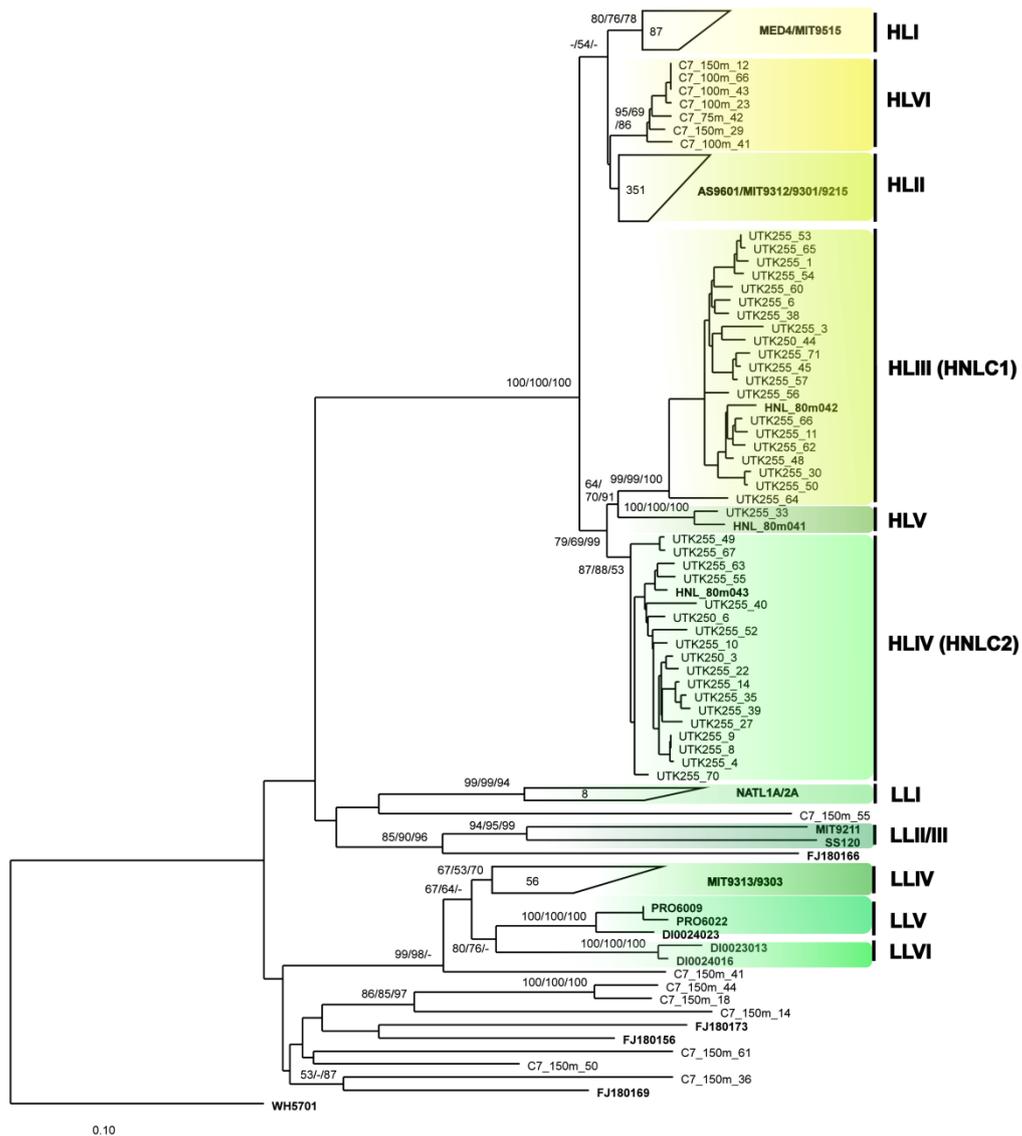
**Figure 11** : Distributions verticales typiques des cyanobactéries *Prochlorococcus* et *Synechococcus*. Les abondances maximales peuvent être restreintes à la couche mélangée superficielle dans les eaux mésotrophes (A-B) avec une dominance de *Synechococcus* ou co-dominance des deux genres, ou étalées sur une plus grande profondeur dans les eaux plus pauvres (C-F), avec une dominance de *Prochlorococcus* sur *Synechococcus* d'un facteur 10 dans les eaux tempérées de l'Atlantique N ou de Méditerranée ou les eaux équatoriales (C-E) ou d'un facteur 100 dans les eaux tropicales et subtropicales (F). Les profils présentés ont été obtenus (A & B) au large de la côte mauritanienne (à 20°N, 18°W et 18°N, 21°W, respectivement), (C) dans l'Atlantique N (30°N, 23°W), (D) en Mer Méditerranée (34°N, 18°E), (E) dans le Pacifique équatorial (150°W, 5°S) et (F) dans le Pacifique tropical (150°W, 16°S). La fluorescence de *Prochlorococcus* est souvent trop faible pour que ces cellules soient détectées dans la couche superficielle. D'après Partensky et al. (1999b).

## I.2.4 Diversité génétique chez les picocyanobactéries marines et phylogéographie

### I.2.4.1. *Prochlorococcus*

L'étude de la photophysologie de souches en culture de *Prochlorococcus* a souligné qu'il existait une grande disparité entre souches dans leur capacité d'acclimatation à la lumière (Moore et al., 1995; Moore et al., 1998). Ainsi des souches isolées dans la partie supérieure de la couche photique, comme MED4, sont capables de croître jusqu'à des intensités lumineuses très élevées ( $450 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$ ) alors que pour des souches comme MIT9313 et SS120, isolées en profondeur, des intensités supérieures à  $150 \mu\text{mol photons.m}^{-2}.\text{s}^{-1}$  s'avèrent létales (Moore et al., 1995; Moore et al., 1998). Ces observations ont en partie été expliquées par des différences dans le contenu pigmentaire de ces isolats. Les souches de profondeur présentent des rapports chl  $b_2$ /chl  $a_2$  supérieurs ou égaux à 1, alors que ce rapport est jusqu'à 10 fois plus faible chez les souches capables de croître à de plus fortes intensités lumineuses (Partensky et al., 1996). Cette différenciation entre populations adaptées aux fortes lumières (HL pour « high-light ») et aux faibles lumières (LL pour « low-light ») avait préalablement été décrite dans le milieu naturel par (Campbell et Vaulot, 1993), qui ont observé par cytométrie en flux une distribution bimodale de la fluorescence rouge de cellules de *Prochlorococcus* en profondeur.

De plus, des analyses phylogénétiques ont montré que ces populations présentant des caractères physiologiques distincts appartenaient à des clades différents (Ferris et Palenik, 1998; Moore et al., 1998; Urbach et al., 1998; Scanlan et West, 2002). Ces clades sont soutenus en utilisant différents marqueurs génétiques, à savoir le gène codant pour l'ARNr 16S (Moore et al., 1998; West et al., 2001) ou la région intergénique située entre l'ARNr 16S et l'ARNr 23S (ITS; Fig. 12; Rocap et al., 2002; Johnson et al., 2006; Huang et al., 2012). D'autres gènes ont également été utilisés comme le gène codant pour la sous-unité  $\beta$  de la PE (*cpeB*; Steglich et al., 2003b), l'ARN polymérase (*rpoC1*; Ferris et Palenik, 1998; Jameson et al., 2010), la grande sous-unité de la ribulose-1,5-bisphosphate carboxylase oxygenase ou RUBISCO (*rbcl*; Wawrick et al., 2003). Ces clades ont ainsi été nommés. A noter que le terme d'écotype est souvent utilisé pour qualifier les différents clades de *Prochlorococcus* (Coleman et Chisholm, 2007). Dans les modèles théoriques de l'évolution bactérienne, le terme d'écotype est plutôt utilisé pour désigner des populations de cellules partageant une niche écologique commune (voir revue de Cohan, 2006). L'équipe de S.W. Chisholm utilise généralement pour nommer les écotypes de *Prochlorococcus* le nom d'une souche représentative, précédée de la lettre « e » (par exemple: eMED4; Kettler et al., 2007), mais d'autres équipes préfèrent utiliser une nomenclature en HL/LL suivi de chiffres romains.



**Figure 12** : Arbre phylogénétique basé sur l'espace intergénique 16S-23S montrant les relations entre les différents clades ou écotypes de *Prochlorococcus*, par la méthode du Neighbor-Joining. Les valeurs de bootstraps montrées aux différents nœuds correspondent respectivement à NJ-PAUP, NJ-PHYLIP et ML-PAUP. Modifié d'après Huang et al. (2012).

Les écotypes HL et LL sont tous les deux divisés en six groupes (Fig. 12). Les nombreuses études de diversité menées au sein de différents environnements océaniques ont permis d'identifier les niches écologiques de ces différents écotypes (Bouman et al., 2006; Zinser et al., 2006; Huang et al., 2012). Les écotypes HLII (ou eMIT9312) dominent les eaux superficielles tropicales à subtropicales, entre les latitudes 30°N et 30°S. Cet écotype est progressivement remplacé à haute latitude par l'écotype HLI (ou eMED4) qui devient dominant dans les eaux plus froides situées entre 35-48°N et 35-40°S (Ahlgren et Rocop, 2006; Johnson et al., 2006; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008). Cette distribution latitudinale s'explique par des différences au niveau des optima de température observés chez les différentes souches en culture de ces deux écotypes, l'écotype HLII étant incapable de croître à des températures inférieures à 15°C, contrairement à l'écotype HLI qui présente une plus grande tolérance aux températures froides (Johnson et al., 2006;

Zinser et al., 2007). Récemment de nouveaux clades HL, nommés HNLC1 et HNLC2, ont été décrits en utilisant la métagénomique. Ces clades semblent restreints aux régions HNLC ("High Nutrients, Low Chlorophyll") de l'océan Pacifique équatorial et de l'Océan Indien (Rusch et al., 2010; West et al., 2010) où la production primaire est limitée par les faibles concentrations en fer (Martin et al., 1994; Mann et Chisholm, 2000). Cependant grâce à l'apport de la génomique à partir d'une cellule (SCG pour Single Cell Genomics), il a récemment été démontré que dans ces deux études utilisant des marqueurs génétiques différents (gènes *core* pour Rusch et collaborateurs, 16S rRNA pour West et collaborateurs), ces deux clades ont été dénommés avec une nomenclature inverse (Malmstrom et al., 2013). Pour davantage de clarté et afin de suivre la nomenclature existante, ces clades sont à présent nommés HLIII et HLIV (Huang et al., 2012). Le nouveau clade HLV est uniquement retrouvé dans la zone équatoriale (Huang et al., 2012). Contrairement aux écotypes HLI et HLII qui présentent une large distribution géographique, ces trois derniers clades semblent donc être confinés à des régions aux conditions hydrologiques très particulières. Enfin, le clade HLVI a été découvert dans la partie inférieure de la couche photique de la station C7 au sud de la Mer de Chine. Si seulement 7 séquences définissent ce clade, les auteurs suggèrent qu'il serait le premier clade "HL" adapté à des conditions de faible intensité lumineuse, remettant donc en question cette nomenclature (Huang et al., 2012).

La situation est plus complexe au niveau des écotypes LL. L'écotype LLIV (eMIT9313) est retrouvé à de fortes concentrations à la base de la couche photique, de l'équateur jusqu'aux latitudes 35°N et 40°S (Johnson et al., 2006). Le clade LLII (eSS120) présente une distribution similaire et coexiste souvent avec l'écotype LLIV (Malmstrom et al., 2010). L'absence de ces écotypes en surface est en accord avec les faibles optima lumineux de croissance des souches en culture (Moore et al., 1998). L'écotype LLI (eNATL), semble présenter une distribution latitudinale plus marquée. Aux basses latitudes, on le retrouve en dessous la thermocline, dans une position intermédiaire entre les écotypes HL et les écotypes strictement de basse lumière (LLII et LLIV). A haute latitude, cet écotype peut être présent jusqu'en surface, où il coexiste avec l'écotype HLI. Cet écotype correspond en réalité à un écotype intermédiaire d'un point de vue physiologique entre les écotypes HL et les écotypes LL stricts, ce qui est en accord avec sa position phylogénétique (Kettler et al., 2007; Partensky et Garczarek, 2010). Peu de données ont été obtenues concernant la distribution de l'écotype LLIII (eMIT9211), qui semble être peu abondant *in situ*. Des clades additionnels ont été révélés grâce aux banques de clones, mais sans représentants en culture, avec le clade NC1 (pour "Non-Cultured 1") retrouvé abondant au niveau de plusieurs stations des océans Atlantique et Pacifique (Martiny et al., 2009b), et les clades LLV et LLVI qui semblent spécifiques des zones pauvres en oxygène (Lavin et al., 2010). Tout comme les écotypes HL récemment mis en évidence grâce à la métagénomique, ces clades semblent être restreints à des conditions hydrologiques très spécifiques.

I.2.4.2. *Synechococcus*

Les *Synechococcus* marins peuvent être divisés en trois groupes phylogénétiques majeurs ou « sub-clusters »: 5.1, 5.2 et 5.3 (Herdman et al., 2001; Dufresne et al., 2008). Initialement défini comme le « marine cluster A », ou MC-A, le sub-cluster 5.1 domine la zone photique des eaux côtières et océaniques et est celui qui présente la plus grande diversité génétique (Fuller et al., 2003). Les deux autres sub-clusters apparaissent beaucoup moins représentés dans l'environnement océanique, notamment le sub-cluster 5.2, qui comprend principalement des souches halotolérantes, confinées aux régions côtières et plus particulièrement aux zones estuariennes. Si la taxonomie du genre *Synechococcus* mériterait d'être revue d'un point de vue global, celle du sub-cluster 5.2 apparaît comme la plus problématique en fonction des différents marqueurs génétiques utilisés. En effet, la délimitation entre ce cluster et le genre *Cyanobium* est très mal définie (Fig. 8 et voir Fig. 1 de Chen et al., 2006). En se basant sur le gène marqueur *rbcL* ou l'ITS, les clades CB4 et CB5 découverts dans la baie de Chesapeake partageraient une origine évolutive commune avec les souches du sub-cluster 5.1, alors que la souche WH5701, utilisée initialement pour décrire le sub-cluster 5.2 (Herdman et al., 2001), tomberait en dehors de ce groupe, avec des *Cyanobium* d'eau douce (Chen et al., 2006). D'autre part, la répartition géographique de ce sub-cluster reste largement méconnue. Néanmoins, il est intéressant de noter que le clade CB5 dominerait à haute latitude en mer de Bering et en mer des Tchouktches (Huang et al., 2012). Initialement le subcluster 5.2 ne comprenait que des souches à phycocyanine (PC), mais l'isolement de souches contenant de la PE, au sein des clades CB4 et CB5, est venue contredire la validité de ce caractère taxonomique (Chen et al., 2004).

La diversité génétique du sub-cluster 5.1, le plus abondant dans la nature et le mieux représenté en culture, a fait l'objet de nombreuses études. Basée sur l'analyse phylogénétique du gène codant pour l'ARNr 16S, 9 clades ont ainsi pu être mis en évidence au sein de ce sub-cluster, mais ce marqueur est assez peu résolutif. De nombreux clades supplémentaires ont en effet été trouvés en utilisant d'autres marqueurs génétiques tels que l'ITS (Figs. 13 et 14; Ahlgren et al., 2006; Choi et Noh, 2009; Huang et al., 2012), *rpoC1* (Palenik, 1994; Toledo et al., 1999; Mühling et al., 2005), *cpeB* (Haverkamp et al., 2008), le régulateur de l'azote (*ntcA*; Penno et al., 2006), *rbcL* (Paerl et al., 2011b), la nitrate reductase (*narB*; Jenkins et al., 2006; Paerl et al., 2008; Paerl et al., 2011a) ou le cytochrome *b<sub>6</sub>* (*petB*; Mazard et al., 2012a), même s'il reste délicat de comparer les phylogénies obtenues avec ces différents marqueurs. En se basant sur l'ITS (Fig. 13), une vingtaine de clades sont actuellement répertoriés au sein du sub-cluster 5.1, un marqueur très résolutif, deux au niveau du sub-cluster 5.2 et huit au sein du sub-cluster 5.3 (Fig. 14; Huang et al., 2012). Ce nombre devrait encore augmenter dans les années à venir avec l'essor des nouvelles technologies de séquençage (NGS) et l'utilisation croissante des approches métagénomiques. Néanmoins de nombreux groupes

phylogénétiques ainsi décrits ne comportent aucun représentant cultivé, soulignant l'importance de l'approche culturelle, au niveau de l'isolement et de la caractérisation de nouvelles souches (Huang et al., 2012; Mazard et al., 2012a).

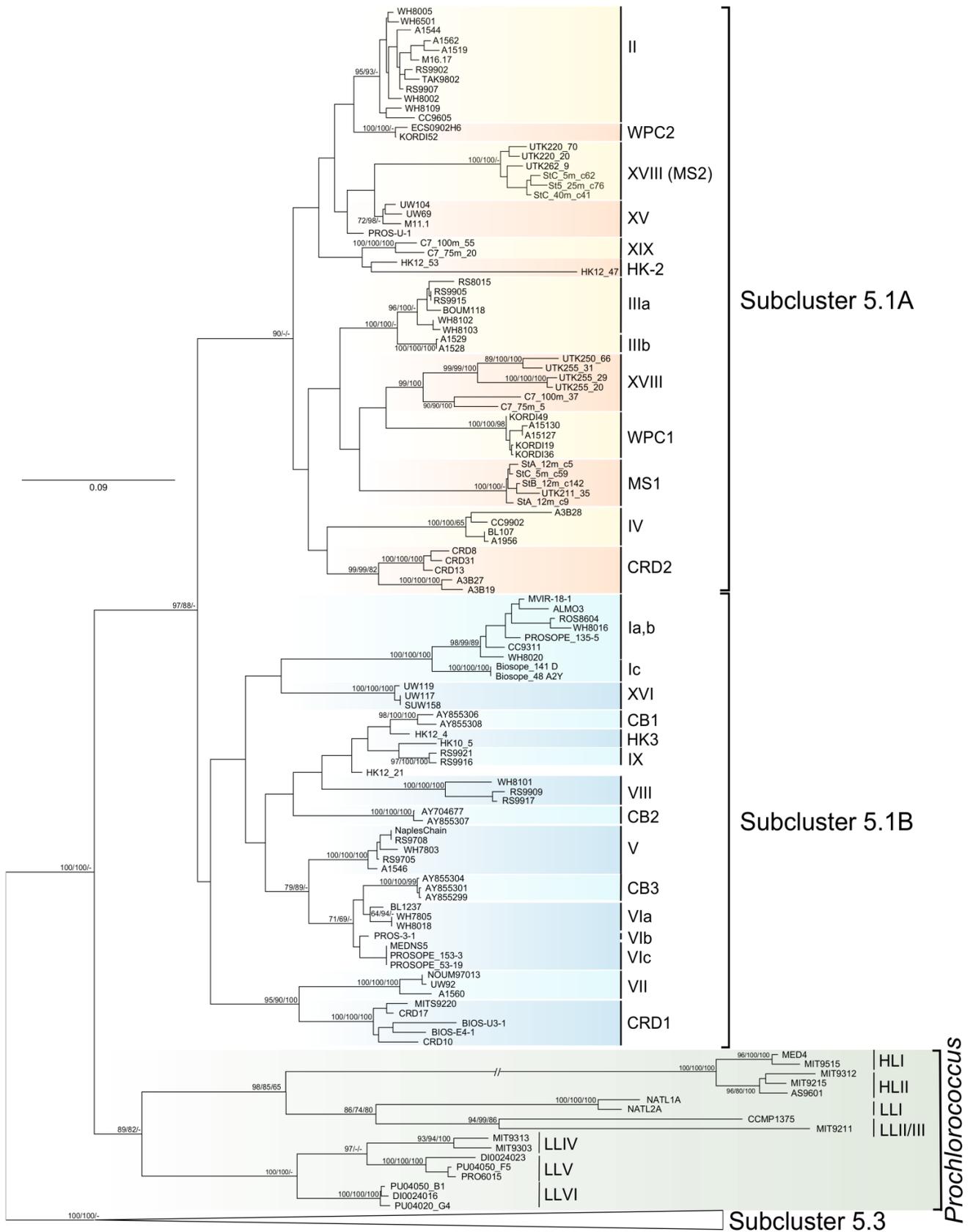
On connaît encore peu de caractères phénotypiques qui soient véritablement spécifiques de chacun des différents clades (Tableau 1). La motilité est un caractère uniquement retrouvé au sein du clade III (Toledo et al., 1999). De même, les représentants du clade VIII sont halotolérants et partagent une pigmentation de type 1, et ne possèdent ni phycoérythrobiline (PEB) ni phycourobiline (PUB) et leur complexe collecteur de lumière n'est composé que de PC (voir §1.2.5; Fuller et al., 2003). Les membres des clades CB4 et CB5 ne possèdent quant à eux pas de PUB (Chen et al., 2004).

Contrairement à *Prochlorococcus*, il est donc difficile chez *Synechococcus* de définir la niche occupée par les différents clades ou écotypes. Cependant, les outils moléculaires ont permis de tirer quelques généralités concernant la distribution géographique et écologique de certains clades (Tableau 1). Tout d'abord, l'ensemble des études de diversité convergent pour dire que les clades I à IV dominent la structure des communautés de *Synechococcus* dans la grande majorité des environnements océaniques. Parmi ces 4 clades, la coexistence des deux clades I et IV a souvent été rapportée. Ces deux clades sont prédominants à haute latitude (au dessus de 30°N et en dessous de 30°S) dans les environnements eutrophes à mésotrophes. Ces groupes phylogénétiques sont particulièrement abondants dans les environnements côtiers dont les eaux riches en nutriments présentent une température faible, une salinité et une concentration en chl *a* plus élevées (Brown et Fuhrman, 2005; Brown et al., 2005; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008; Jing et al., 2009a; Tai et Palenik, 2009; Cai et al., 2010; Paerl et al., 2011a). Ces clades sont ainsi présents dans divers environnements océaniques comme le Golfe du Lion en mer Méditerranée (Mella-Flores et al., 2011; Annexe A), les côtes de l'océan Pacifique Nord (Brown et al., 2005; Tai et Palenik, 2009; Tai et al., 2011), les eaux eutrophes à mésotrophes de l'océan Atlantique Nord (Zwirgmaier et al., 2007; Huang et al., 2012) et l'océan Arctique (Zwirgmaier et al., 2008). Des cycles saisonniers d'abondance ont également pu être observés (Brown et al., 2005; Tai et Palenik, 2009). Généralement les représentants du clade IV sont deux fois plus abondants que ceux appartenant au clade I (Tai et Palenik, 2009; Mella-Flores et al., 2011). Cependant le clade I devient dominant chaque année juste avant le bloom printanier de *Synechococcus* sur la côte californienne et constitue également une proportion plus significative de la communauté comparé au clade IV au fur et à mesure que l'on se rapproche du pôle Arctique (Tai et Palenik, 2009; Huang et al., 2012; Mazard et al., 2012a). La souche CC9311, représentative du clade I, possède un génome de taille plus importante que les souches du clade IV BL107 et CC9902 et possède un plus grand répertoire protéique impliqué dans la réponse à différents signaux environnementaux (histidine kinases et régulateurs de réponse) ainsi que dans l'assimilation et le transport de métaux et de composés organiques (Palenik et al., 2006; Scanlan et

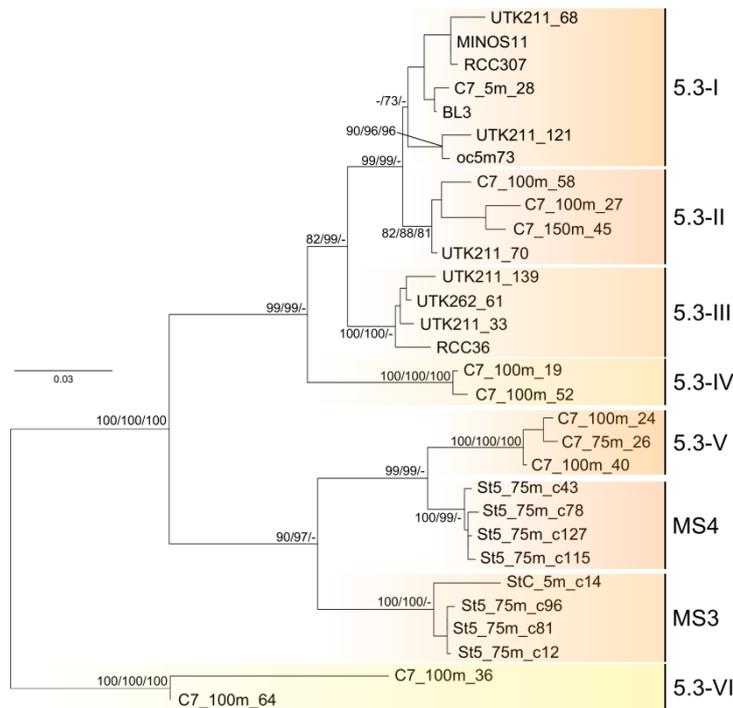
al., 2009). Ceci pourrait donc expliquer l'abondance du clade I dans les environnements côtiers où la disponibilité en nutriments est particulièrement variable, mais n'explique pas la dominance du clade IV dans la plupart de ces régions. Il a ainsi été proposé que le clade I serait plus « opportuniste » alors que le clade IV serait « spécialiste » de ces environnements (Dufresne et al., 2008; Zwirgmaier et al., 2008). Des comparaisons faites sur un nombre plus important de génomes de chacun des deux clades s'avèrent nécessaires afin de mieux comprendre la coexistence de ces clades.

Contrairement aux clades I et IV, confinés aux eaux froides (7°C à 18°C), le clade II est particulièrement abondant dans les eaux chaudes avec une température comprise entre 22°C et 28°C (Zwirgmaier et al., 2008). Ce clade domine ainsi la mer Rouge (Fuller et al., 2003), la région Nord de la mer d'Arabie (Fuller et al., 2006), le courant californien (Toledo et Palenik, 2003), la région de remontée d'eaux froides du Maroc (Zwirgmaier et al., 2007), et la bande équatoriale de l'océan Atlantique (Zwirgmaier et al., 2008; Mazard et al., 2012a). Il est aussi bien présent au large qu'en zone côtière et en surface qu'en profondeur. Ce clade présente la plus grande diversité génétique au sein du sub-cluster 5.1, avec 7 sous-clades décrits en utilisant le gène marqueur *petB* (incluant le clade IIh qui correspond au clade XV tel que défini en se basant sur l'ITS; Mazard et al., 2012). Par opposition à ces trois clades qui démontrent une préférence latitudinale, le clade III préfère les environnements oligotrophes, où il domine la partie supérieure de la couche photique (Toledo et al., 1999; Toledo et Palenik, 2003; Zwirgmaier et al., 2008). Les représentants de ce clade semblent adaptés à une fenêtre assez étroite de concentration en nitrate et phosphate (Zwirgmaier et al., 2008). Le clade III domine ainsi dans la partie Est de la mer Méditerranée (Mella-Flores et al., 2011; Annexe A), ainsi qu'à la station BATS en Mer des Sargasses (Huang et al., 2012; Mazard et al., 2012a).

Les autres clades du genre *Synechococcus* semblent être moins abondants dans l'environnement. Quelques exceptions sont à noter, comme l'exemple de l'environnement ultra-oligotrophe du gyre du Pacifique Sud et celui du Pacifique équatorial où les clades CRD1 et VII seraient respectivement les clades dominants (Ostrowski M. et Scanlan D., comm. pers.; Huang et al., 2012). Le clade VII domine aussi dans les eaux de remontées d'eau froide du Costa Rica (Saito et al., 2005). Néanmoins la position phylogénétique de ces deux clades, en utilisant les marqueurs ITS et *petB*, montre qu'ils sont très proches (Mella-Flores et al., 2011; Huang et al., 2012; Mazard et al., 2012a). Il pourrait s'agir en réalité de deux sous-groupes d'un même clade dont la distribution écologique et les caractères physiologiques seraient similaires. La distribution des clades V et VI reste peu documentée, mais semblent abondants, voire dominants dans la mer d'Arabie et plusieurs autres régimes océaniques (Fuller et al., 2006). Cependant la majorité de ces profils de distribution ont été obtenus par Dot-Blot, en utilisant des sondes ciblant collectivement les clades V/VI/VII qui pourrait se révéler en partie aspécifique (Zwirgmaier et al., 2008; Mella-Flores et al., 2011).



**Figure 13 :** Arbre phylogénétique en maximum de vraisemblance (ML) basé sur l'espace intergénique 16S-23S (longueur totale, incluant les ARNt), montrant les relations entre les différents clades de *Synechococcus* du subcluster 5.1. Les nombres présents à chaque nœud représentent les valeurs de bootstraps obtenues pour les analyses en ML, par la méthode du Neighbor-Joining (NJ) et en maximum de parcimonie (PARS), respectivement. Les bootstraps représentés en pourcentage, ont été obtenus à partir de 100 répétitions pour les analyses ML, 1000 pour les analyses NJ et PARS, et uniquement les valeurs supérieures à 70% sont montrées. (F. Humily, non publié).



**Figure 14** : Arbre phylogénétique en maximum de vraisemblance (ML) basé sur l'espace intergénique 16S-23S (longueur totale, incluant les ARNt), montrant les relations entre les différents clades de *Synechococcus* du subcluster 5.3. Les nombres présents à chaque nœud représentent les valeurs de bootstraps obtenues pour les analyses en ML, en Neighbor-Joining (NJ) et en maximum de parcimonie (PARS), respectivement. Les bootstraps représentés en pourcentage, ont été obtenus à partir de 100 répétitions pour les analyses ML, 1000 pour les analyses NJ et PARS, et uniquement les valeurs supérieures à 70% sont montrées. (F. Humily, non publié).

L'environnement estuarien de la baie de Cheasepeake apparaît principalement peuplé par des souches halotolérantes des clades CB6 et CB7 du sub-cluster 5.2 (Cai et al., 2010). Le clade CB5, du même sub-cluster, domine quant à lui dans le Nord de la mer de Béring et la mer des Tchoukhes (Huang et al., 2012). Si, dans le premier cas, les populations majoritaires de *Synechococcus* proviendraient principalement des apports d'eau douce se déversant dans la baie, comme observé dans la baie d'Hong-Kong avec les clades HK-4 et HK-5 (Jing et al., 2009b), la prévalence du clade CB5 sur de si vastes étendues suggèrent que ces populations seraient autochtones de l'océan Arctique (Huang et al., 2012). Ces génotypes seraient ainsi adaptés aux régions polaires et subpolaires dont la température est proche de 0°C.

**Tableau 1 : Diversité écologique des *Synechococcus* marins et biogéographie des différents clades.** Pour chacun des clades, l'existence d'isolats en culture, les types pigmentaires retrouvés (1, 2, 3 et acclimateurs chromatiques; voir §II.2.5) ainsi que les différents gènes marqueurs utilisés pour la définition des différents clades phylogénétiques sont mentionnés. Les autres gènes marqueurs sont a : *rbcL*, b : *narB*, c : *rpoC1*.

Sub-cluster	C lade	Subclade	Physiology	Isol-ates	Pigment types				Genetic markers				Biogeography	References			
					1	2	3	CA	16S	ITS	<i>petB</i>	<i>ntcA</i>			others		
5.1A	II			●	●	●	●	●	●	●	●	●	abc	Abundant in warm waters, both in coastal and open-ocean regions	Ahlgren and Rocap (2006); Zwiarglmaier et al. (2007, 2008); Chung et al. (2011); Jing et Liu (2011); Mella-Flores et al. (2011); Post et al. (2011); Ahlgren et Rocap (2012); Choi (2012); Huang et al. (2012); Mazard et al. (2012b); Guo et al. (2013)		
		IIa		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)	
		IIb, IIc, IIe, IIg, IIh		●	●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
		IIc		●	●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
		XV	IIh	Slow growth on nitrate	●	●	●	●	●	●	●	●	●	bc	Seen in open ocean, East China Sea, Sargasso Sea	Ahlgren and Rocap (2006, 2012); Choi (2012)	
		III		Motile	●	●	●	●	●	●	●	●	●	bc	Restricted to warm, oligotrophic waters	Jing and Liu (2011); Mella-Flores et al. (2011); Post et al. (2011); Ahlgren and Rocap (2012); Mazard et al. (2012b)	
			IIIa		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
			IIIb		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
			IIIc		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
		IV			●	●	●	●	●	●	●	●	●	abc	Dominant in cold coastal waters, co-occurs with clade I. Present in polar waters	Zwiarglmaier et al. (2007, 2008); Jing et al. (2009a); Tai and Palenik (2009); Mella-Flores et al. (2011); Paerl et al. (2011a); Ahlgren and Rocap (2012); Choi (2012); Mazard et al. (2012)	
			IVa		●	●	●	●	●	●	●	●	●	b		Paerl et al. (2011a); Ahlgren and Rocap (2012); Mazard et al. (2012b)	
			IVb		●	●	●	●	●	●	●	●	●	b		Paerl et al. (2011a); Ahlgren and Rocap (2012); Mazard et al. (2012b)	
			HK-2							●						Only observed in subtropical coastal water east of Hong Kong	Jing et al. (2009b)
			MS1							●						Only observed in surface waters of Mediterranean Sea	Mella-Flores et al. (2011)
			MS2/XVIII							●						Present in oligotrophic waters of Atlantic, Pacific and Mediterranean Sea	Mella-Flores et al. (2011); Huang et al. (2012)
	XIX							●						Only observed in South China Sea at depth (C7 station)	Mella-Flores et al. (2011);		
	WPC1			●	●	●	●	●	●	●	●	●	●	Present in East China Sea, East Sea, Mediterranean Sea, Pacific and Atlantic	Choi and Noh (2009); Mella-Flores et al. (2011); Huang et al. (2012)		
	WPC2			●	●	●	●	●	●	●	●	●	●	Present in East China Sea	Choi and Noh (2009); Mella-Flores et al. (2011)		
	WPC3							●						Only observed in North Equatorial Current (Northwestern Pacific Ocean) and East China Sea	Choi et al. (2011); Choi (2012)		
5.1B	CB1			●	●	●	●	●	●	●	●	●	●	Related to HK-3. Described on environmental sequences from the Chesapeake Bay, but present in col/arctic environments.	Chen et al. (2004); Chen et al. (2006); Choi et Noh (2009); Cai et al. (2010); Huang et al. (2012)		
														Only seen in Chesapeake bay	Chen et al. (2006); Cai et al. (2010)		
														Only seen in Chesapeake bay	Chen et al. (2006); Cai et al. (2010)		
														Related to I. Present in winter in coastal waters of Chesapeake bay	Cai et al; (2010)		
		CRD1	CRD1a, CRD1b		●	●	●	●	●	●	●	●	bc	Related to VII. Dominant in Costa Rica upwelling and southern Pacific. Present in Atlantic Ocean	Saito et al. (2005); Ahlgren and Rocap (2012); Huang et al. (2012); Mazard et al. (2012b); Scanlan et Ostrowski (comm. pers.)		
		CRD2			●	●	●	●	●	●	●	●	●	●	Dominant in Costa Rica upwelling. Present in oligotrophic environments of Atlantic and Pacific.	Saito et al. (2012); Huang et al. (2012)	
		I			●	●	●	●	●	●	●	●	●	ac	Dominant in cold coastal waters, co-occurs with clade IV. Present in polar waters	Ahlgren and Rocap (2006); Mella-Flores et al. (2011); Mazard et al. (2012b); Huang et al. (2012)	
			Ia		●	●	●	●	●	●	●	●	●	●	b		Paerl et al. (2011a); Ahlgren and Rocap (2012); Mazard et al. (2012b)
			Ib		●	●	●	●	●	●	●	●	●	●	b		Paerl et al. (2011a); Ahlgren and Rocap (2012); Mazard et al. (2012b)
			Ic		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
		V			●	●	●	●	●	●	●	●	●	●	b	Distribution unclear due to aspecificity of 16S probe used in diversity studies. Found in East China Sea and South Sea.	Fuller et al. (2003); Choi and Noh (2009); Choi et al. (2011); Ahlgren and Rocap (2012)
		VI			●	●	●	●	●	●	●	●	●	●	b	Distribution unclear due to aspecificity of 16S probe used in diversity studies. Found in East China Sea, South Sea, Mediterranean Sea	Fuller et al. (2003); Choi and Noh (2009); Choi et al. (2011); Ahlgren and Rocap (2012); Mella-Flores (2011)
			VIa		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)
	VIb, VIc		●	●	●	●	●	●	●	●	●	●	●		Mazard et al. (2012b)		
VII/XIII/NC3				●	●	●	●	●	●	●	●	ab	Related to CRD1. Found in tropical and sub-tropical oceans.	(Lindell et al. (2005); Penno et al. (2006); Ahlgren and Rocap (2012); Huang et al. (2012)			

Sub-cluster	Clade	Subclade	Physiology	Isol-	Pigment types				Genetic markers				Biogeography	References		
				ates	1	2	3	CA	16S	ITS	petB	ntcA			others	
<b>5.1B</b>	VIII		NoPE, some strains are unable to grow on nitrate	●	●					●	●	●	●	bc	Found in hypersaline waters, Red Sea, East China Sea.	Fuller et al. (2003); Dufresne et al. (2008); Choi and Noh (2009); Huang et al. (2012)
	IX			●		●	●			●	●	●		b	Rarely observed but seen in East China Sea, Red Sea and upwelling off of Africa	Fullet et al. (2003); Zwirgmaier et al. (2008); Choi and Noh (2009); Choi et al. (2011); Ahlgren and Rocap (2012); Choi (2012)
	HK-2									●					Only observed in subtropical coastal water east of Hong Kong	Jing et al. (2009b)
	HK-3									●					Related to CB1. Only observed in subtropical coastal water east of Hong Kong	Jing et al. (2009b)
	XVI				●				●	●	●	●	●	bc	Present in North Atlantic, Mediterranean Sea, Pacific Ocean and East China Sea.	Ahlgren and Rocap (2006, 2012); Jing and Liu (2011); Huang et al. (2012); Mazard et al. (2012)
<b>5.1</b>	XI/NC1											●	a	Only observed in Gulf of Aqaba. Predominant during spring bloom. Possibly adapted to waters of high trophic level	Lindell et al. (2005); Penno et al. (2006); Post et al. (2011)	
	XII/NC2											●	a	Only observed in Gulf of Aqaba. Predominant during winter mixing and summer stratification	Lindell et al. (2005); Penno et al. (2006); Post et al. (2011)	
	XIV											●	a	Only observed in Gulf of Aqaba	Penno et al. (2006); Post et al. (2011)	
	EnvA											●		Found in Atlantic Ocean	Mazard et al. (2012b)	
	EnvB											●		Found in Atlantic Ocean	Mazard et al. (2012b)	
	EnvC											●		Found in Atlantic Ocean	Mazard et al. (2012b)	
<b>5.2</b>	CB4		Lack PUB and some strains are motile	●	●							●			Found in temperate coastal waters	Chen et al. (2004); Chen et al. (2006); Cai et al. (2010); Huang et al. (2012)
	CB5		Lack PUB. One motile strain	●	●	●				●	●				Present in temperate estuarine/coastal waters and prevalent in polar/subpolar waters	Waterbury et al. (1986); Chen et al. (2004); Chen et al. (2006); Cai et al. (2010); Choi et al. (2011); Huang et al. (2012)
	HK-4											●			Only observed in subtropical coastal water east of Hong Kong. Transient presence of HK-4 along a seasonal study.	Jing et al. (2009b)
	HK-5											●			Only observed in subtropical coastal water east of Hong Kong	Jing et al. (2009b)
<b>5.3</b>				●		●	●			●	●	●	b	Rarely seen <i>in situ</i> , but predominant in Sargasso Sea and highly abundant in China Sea	Ahlgren and Rocap (2006); Mella-Flores et al. (2011); Post et al. (2011); Ahlgren and Rocap (2012); Choi (2012); Guo et al. (2013)	
	I			●		●	●			●	●	●	b	Found in Mediterranean Sea, East China Sea and North Atlantic Ocean	Fuller et al. (2003); Choi and Noh (2009); Ahlgren and Rocap (2012); Huang et al. (2012)	
	II											●		Found in Atlantic Ocean and East China Sea	Huang et al. (2012)	
	III			●		●						●		Found in Pacific and Atlantic Ocean	Huang et al. (2012)	
	IV, V, VI											●		Only observed in South China Sea at depth	Huang et al. (2012)	
	MS3											●		Only observed in Mediterranean Sea (St5) at depth	Mella-Flores et al. (2011)	
	MS4											●		Only observed in Mediterranean Sea (St5) at depth	Mella-Flores et al. (2011)	

---

---

## PARTIE II

# ROLE DES FACTEURS ENVIRONNEMENTAUX DANS LA DISTRIBUTION DES PICOCYANOBACTÉRIES MARINES

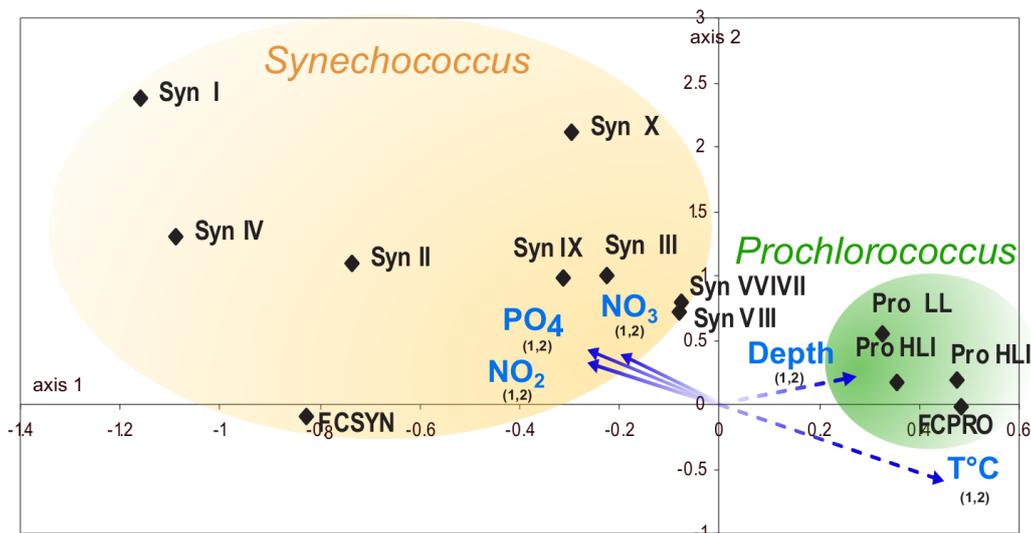
---

II.1. Facteurs environnementaux.....	32
II.1.1 Facteurs biotiques.....	33
II.1.1.1. Le broutage.....	33
II.1.1.2. La lyse virale.....	34
II.1.2 Facteurs abiotiques.....	34
II.1.2.1. Intensité lumineuse.....	34
II.1.2.2. Qualité de lumière.....	36
II.1.2.3. La disponibilité en nutriments.....	38
II.1.2.4. La température.....	39
II.2. La compétition pour la lumière dans l'océan.....	39
II.2.1 La photosynthèse oxygénique.....	39
II.2.2 L'organisation générale de l'appareil photosynthétique oxygénique.....	41
II.2.3 Les photosystèmes.....	43
II.2.3.1. Le photosystème II.....	43
II.2.3.2. Le photosystème I.....	44
II.2.3.3. Production d'énergie au sein des membranes thylacoïdales.....	45
II.2.4 Les antennes collectrices de <i>Prochlorococcus</i> .....	45
II.2.5 Les phycobilisomes de <i>Synechococcus</i> .....	49
II.2.5.1. Structure des phycobilisomes.....	49
II.2.5.2. Phycobiliprotéines et chromophorylation.....	54
II.2.5.3. Biosynthèse et attachement des phycobilines.....	59
II.2.5.4. Transfert orienté d'énergie.....	62
II.2.6 Gènes codant pour les polypeptides des phycobilisomes.....	64
II.2.6.1. Nature et organisation génique.....	64
II.2.6.2. Processus évolutifs des phycobiliprotéines.....	67
II.3. L'acclimatation chromatique chez les cyanobactéries.....	71
II.3.1 Diversité des souches capables d'acclimatation chromatique chez les cyanobactéries.....	72
II.3.2 Bases moléculaires de l'acclimatation chromatique de type III.....	73
II.3.3 Bases moléculaires de l'acclimatation chromatique de type II.....	75
II.3.4 Originalité de l'acclimatation chromatique de type IV présente chez les <i>Synechococcus</i> marins.....	77

## II. RÔLE DES FACTEURS ENVIRONNEMENTAUX DANS LA DISTRIBUTION DES PICOCYANOBACTÉRIES MARINES

### II.1. Facteurs environnementaux

La dynamique et la distribution des populations naturelles de cyanobactéries marines sont influencées aussi bien par des facteurs abiotiques (intensité lumineuse, température, stabilité de la colonne d'eau, nutriments, etc.) que biotiques (broutage, toxines, infection virale; Scanlan, 2003). Pour *Prochlorococcus* il est maintenant admis que les paramètres physiques, intensité lumineuse et température, sont les principaux facteurs influençant la structure des communautés (Fig. 15). Pour *Synechococcus*, la mise en évidence de ces facteurs est rendue plus compliquée en raison de la plus grande diversité génétique existante au sein de ce genre. Contrairement à *Prochlorococcus*, il semble que *Synechococcus* soit principalement affecté par des variables chimiques comme la disponibilité en azote et phosphore et également par la température (Fig. 15), même si d'autres études ont suggéré l'importance des métaux traces et de la qualité de lumière (voir §II.1.2.2; Mann et al., 2002).



**Figure 15 :** Analyse factorielle menée à partir d'échantillons environnementaux provenant de différentes campagnes océanographiques (AMT, BIOSOPE, PROSOPE, AMBITION, BEAGLE et Océan Arctique, n=314). Les flèches représentent les coefficients de corrélation entre les paramètres environnementaux. Les cercles définissent les différents clades de *Synechococcus* et écotypes de *Prochlorococcus*. Les flèches pleines et pointillées représentent respectivement les paramètres chimiques et physiques. Modifié d'après Zwirgmaier et al. (2008).

## II.1.1 Facteurs biotiques

### II.1.1.1. *Le broutage*

Parmi les facteurs biotiques, la lyse virale et le broutage par les prédateurs jouent un rôle majeur sur la mortalité cellulaire et la régulation de la biomasse nouvelle des différentes populations picophytoplanctoniques (Scanlan, 2003). Le picoplancton est brouté par un grand nombre de microorganismes, principalement des protistes hétérotrophes (Quevedo et Anadón, 2001; Calbet et al., 2008). Cependant les résultats des différentes études menées sur le broutage des cellules picoplanctoniques diffèrent quant à l'importance relative des différents consommateurs qui pourraient être constitués de petits nanoflagellés hétérotrophes (HNFs pour heterotrophic nanoflagellates; Christaki et al., 2001; Sato et al., 2007; Hirose et al., 2008a; Apple et al., 2011), de ciliés de tailles variées (Christaki et al., 1999; Quevedo et Anadón, 2001; Christaki et al., 2002), ou encore d'appendiculaires (Gorsky et al., 1999). Le broutage est cependant considéré par tous comme la cause majeure de mortalité chez les picocyanobactéries par rapport à la lyse virale (Sherr et Sherr, 2002; Suttle, 2007), et peut avoir une très grande influence sur la structure des communautés *in situ* (Jürgens et al., 1999; Jürgens et Matz, 2002).

Le broutage est néanmoins un phénomène sélectif et les taux d'ingestion peuvent varier entre les différents taxons au sein d'une communauté complexe. Christaki et collaborateurs ont ainsi mis en évidence qu'au sein de la fraction picophytoplanctonique certains picoeucaryotes, dont *Ostreococcus tauri*, étaient beaucoup moins susceptibles au broutage que les procaryotes *Synechococcus* et *Prochlorococcus* (Christaki et al., 2005). Les taux d'ingestion par des nanoflagellés hétérotrophes sont par ailleurs 3 à 10 fois plus élevés pour *Synechococcus* que pour *Prochlorococcus* (Christaki et al., 1999; Christaki et al., 2002). En revanche *Picophagus flagellatus*, un petit flagellé eucaryote, semble préférer *Prochlorococcus* (Guillou et al., 2001). La révélation d'un brouteur si efficace et de si petite taille a d'ailleurs remis en question le paradigme anciennement accepté selon lequel les bactéries (de taille comprise entre 0,4  $\mu\text{m}$  et 1  $\mu\text{m}$ ) ne pouvaient être broutées par des nanoflagellés de taille inférieure à 5  $\mu\text{m}$ . De plus, les différentes souches de *Synechococcus* ne sont pas ingérées et digérées de la même manière (Zwirgmaier et al., 2009). Ces différences intrinsèques pourraient s'expliquer par la présence de gènes impliqués dans la synthèse de la membrane cellulaire, que l'on retrouve couramment au sein d'ilôts génomiques (voir §III.1.3; Dufresne et al. 2008). Dans certains cas une certaine forme de résistance aux brouteurs peut être mise en place, telle qu'une modification de la morphologie ultrastructurale chez *Synechococcus*, avec une haute glycosylation de la couche S paracrystalline (Scanlan, 2003). Ainsi, la taille et la forme des cellules (Gonzalez et al., 1990), l'activité métabolique (del Giorgio et al., 1996), la carence nutritionnelle des

brouteurs (Gonzalez et al., 1993; Boenigk et al., 2001), la qualité nutritionnelle des proies (Martel, 2009) sont autant de facteurs qui permettraient d'expliquer cette sélectivité de broutage.

#### II.1.1.2. *La lyse virale*

Les virus, qui font partie du femtoplankton (0,02 – 0,2 µm), sont typiquement un ordre de magnitude plus abondants que les bactéries hétérotrophes et de ce fait considérés comme les entités biologiques marines les plus abondantes (Fuhrman, 1999). De par leur impact sur les communautés bactérienne, les bactériophages jouent un rôle majeur sur les grands cycles biogéochimiques (voir revues de Fuhrman et Schwabach, 2003; Suttle, 2007). Les virus ont aussi un rôle sur la diversité génétique des communautés (Waterbury et Valois, 1993; Fuhrman et Schwabach, 2003), notamment en tant que médiateurs d'échanges génétiques (Lindell et al., 2004; Sullivan et al., 2005; Coleman et al., 2006; Avrani et al., 2011).

L'impact des virus sur la mortalité bactérienne diffère selon les études (Fuhrman et Schwabach, 2003; Suttle, 2005). Chez *Synechococcus*, la lyse virale serait responsable de 5 à 14% de la mortalité de cellules dans l'océan ouvert (Suttle et Chan, 1994). En provoquant la lyse de bactéries hétérotrophes, la lyse virale pourrait également stimuler la croissance cyanobactérienne (Weinbauer et al., 2011). En théorie toutes les cellules bactériennes peuvent être susceptibles à l'infection, souvent par plusieurs types de virus. Différents virus de *Synechococcus* ont ainsi été isolés à partir des estuaires, des eaux côtières ainsi que de l'océan ouvert (Suttle et al., 1990; Suttle et Chan, 1993; Waterbury et Valois, 1993; Lu et al., 2001; Marston et Sallee, 2003; Wang et Chen, 2008). De nombreux cyanophages de *Prochlorococcus* ont également été isolés (Sullivan et al., 2003; Sullivan et al., 2005). La majorité appartient à trois familles de phages à ADN double brin : *Myoviridae*, *Podoviridae* et *Siphoviridae*. Les deux premières sont les familles dominantes au sein des phages isolés à partir d'environnements marins (Labrie et al., 2013). Certains myoviridae sont capables d'infecter à la fois *Synechococcus* et *Prochlorococcus* (Weigele et al., 2007).

### II.1.2 Facteurs abiotiques

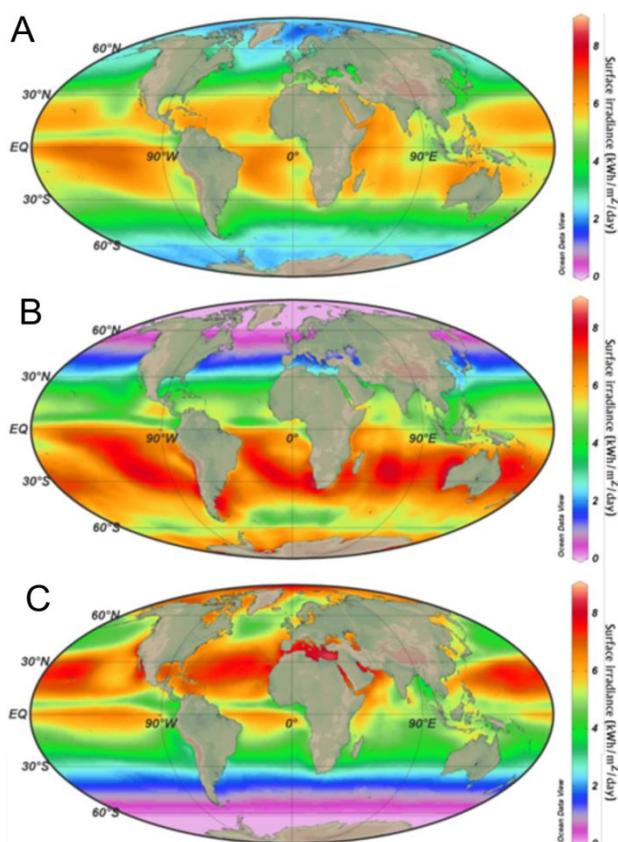
Différents facteurs physiques, tels que la température, la qualité et l'intensité de la lumière incidente, les mouvements des masses d'eau, ou chimiques, comme la disponibilité en nutriments et métaux traces sont parmi les principaux facteurs abiotiques influençant ces microorganismes.

#### II.1.2.1. *Intensité lumineuse*

Parmi l'ensemble des facteurs, biotiques ou abiotiques, dont le picoplancton dépend, la lumière est l'un des plus importants car elle est indispensable à tous les processus liés à la photosynthèse, et donc à la croissance cellulaire. Cette ressource varie aussi bien d'un point de vue

qualitatif que quantitatif, ainsi que spatio-temporellement. Ainsi le rythme des saisons, variant avec la latitude (Fig. 16), et l'alternance jour/nuit sont les deux principales variables influençant l'intensité lumineuse à laquelle les cellules picophytoplanctoniques sont soumises (Garg et Prakash, 2000). L'épaisseur de la couche nuageuse, le pouvoir réfléchissant de la surface océanique (ou albédo) et la diffusion directe ou indirecte des radiations solaires interviennent également.

Les effets de l'intensité lumineuse sur l'activité photosynthétique sont connus depuis longtemps (Ryther, 1956; Lorenzen, 1963; Eppley et Sloan, 1966). Chez les cyanobactéries, de nombreuses fonctions cellulaires, comme le cycle cellulaire (Vaultot et al., 1995; Shalapyonok et al., 1998; Jacquet et al., 2001), la synthèse pigmentaire (Claustre et al., 2002), la fixation de carbone (Bryant et al., 2005) et la capture d'acides aminés (Mary et al., 2008) sont synchronisées par le rythme jour-nuit. Chez *Crocospaera watsonii* sp. WH8501, c'est aussi le cas pour la fixation d'azote (Dron et al., 2012). D'une manière générale, ces variations lumineuses se traduisent par une modulation circadienne de l'expression génique (Wyman, 1999; Garczarek et al., 2001; Holtzendorff et al., 2001; Zinser et al., 2009; Wyman et Thom, 2012).



**Figure 16** : Intensités lumineuses de surface obtenues sur la période 1983 – 2005, à partir des moyennes annuelles sur la période (A), du mois de décembre (B) et du mois de juin (C). Les données ont été obtenues à partir de la NASA Surface meteorology and Solar Energy (<http://eosweb.larc.nasa.gov/sse/>) et sont exprimées en kWh/m<sup>2</sup>/jour.

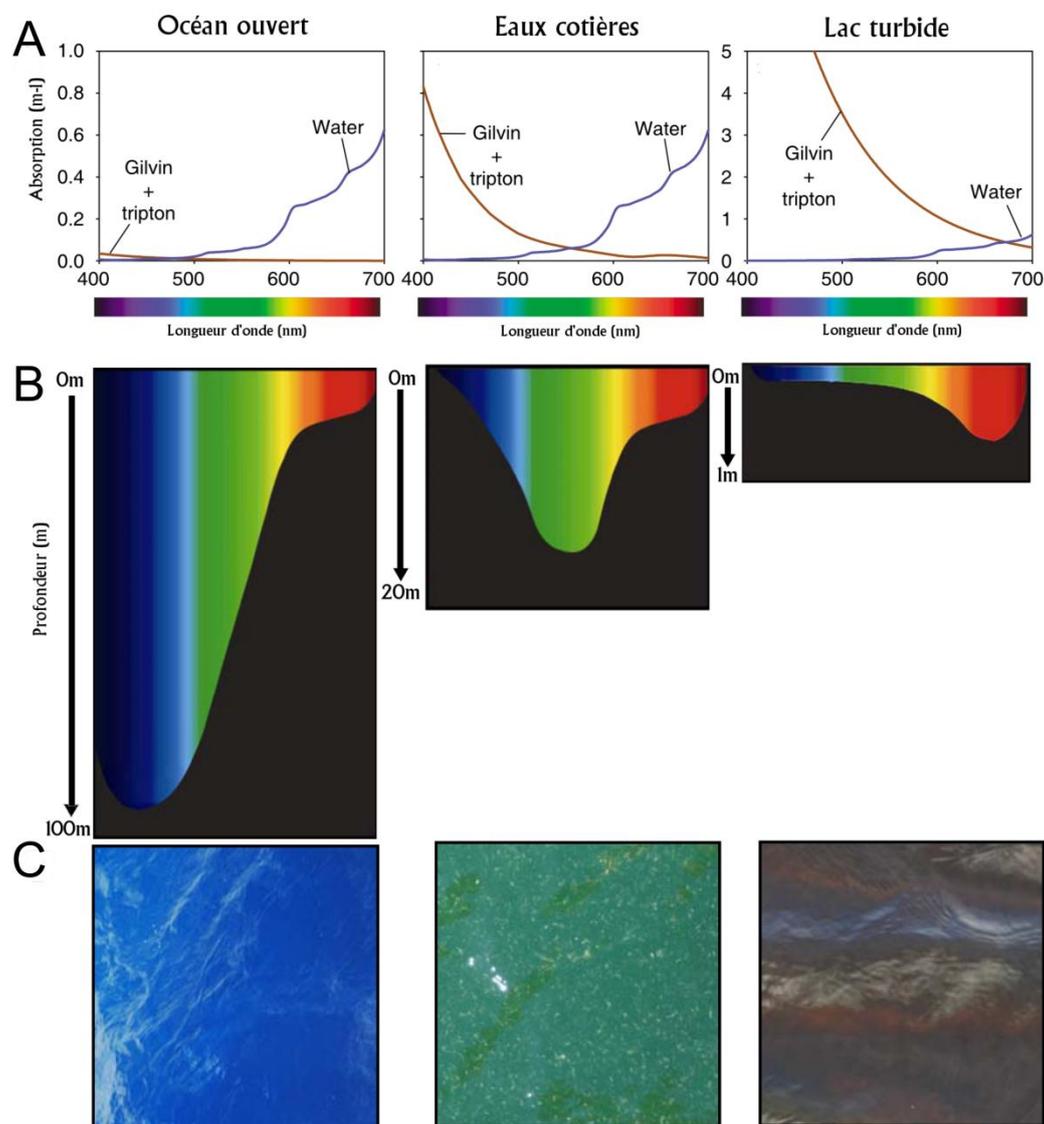
### II.1.2.2. *Qualité de lumière*

La pénétration des photons au sein des environnements océaniques n'est pas uniforme. Celle-ci dépend à la fois de la longueur d'onde considérée et des propriétés optiques de l'environnement (Morel et Maritorena, 2001). Les molécules d'eau elles-mêmes absorbent les radiations lumineuses (Stomp et al., 2007). Si cette absorption est faible dans la partie bleue du spectre solaire (400-500 nm), elle devient très forte dans la partie rouge, expliquant ainsi l'atténuation rapide avec la profondeur des radiations rouges dans l'océan ouvert (Fig. 17). D'autre part, la dégradation de la matière organique par les microorganismes conduit à la production de substances humiques, des composés de haut poids moléculaire, de couleur jaune à marron (Schnitzer, 1978; Ghosh et Schnitzer, 1980). Principalement trouvées dans les zones estuariennes, ces molécules absorbent principalement les longueurs d'onde bleues, expliquant la couleur brunâtre de certaines eaux de drainage ou de certains lacs tropicaux très chargées en ces composés (Fig. 17). La fraction vivante, principalement le phytoplancton, joue aussi un rôle important dans l'absorption de la lumière au sein de la couche photique. Cependant ces organismes ne modifient significativement les propriétés optiques d'un environnement que si les concentrations en nutriments sont suffisantes pour supporter une forte biomasse (Morel, 1988).

Ces différentes fractions, dissoutes ou insolubles, organiques ou minérales, vivantes ou non, modifient de façon considérable les propriétés optiques des différents environnements aquatiques. Les longueurs d'onde capables de pénétrer le plus en profondeur dans la couche photique de l'océan oligotrophe sont centrées autour de 475 nm, conférant leur couleur bleue à ces zones océaniques (Kirk, 1994; Morel et al., 2007; Mass et al., 2010). Les eaux des environnements côtiers, tout comme des lacs « clairs », sont en revanche considérées comme vertes. Enfin la couleur rouge à brunâtre domine dans certains lacs tropicaux eutrophes (Fig. 17; Kirk, 1994; Stomp et al. 2007). Globalement, plus la turbidité d'un environnement aquatique augmente, plus le spectre lumineux est décalé vers le rouge.

La meilleure connaissance des propriétés optiques différentielles des environnements aquatiques a permis d'établir un certain nombre de catégories en se basant sur la transmittance de chacune des longueurs d'onde au sein de la couche photique. En 1976, Jerlov proposa ainsi 3 types pour les régimes océaniques ouverts (I, II et III), et 9 pour les eaux côtières (1 à 9; Jerlov, 1976). Sur la base de mesures des variations spectrales, Smith et Baker ont par la suite établi une classification se basant uniquement sur l'influence des apports terrestres (Smith et Baker, 1981). Ainsi dans les régions où cette influence est minimale, l'atténuation lumineuse est principalement due au phytoplancton et à ses produits dérivés (pigments). Une nouvelle classification, proposée par Morel et Prieur en 1977 émana de ces travaux (Morel et Prieur, 1977; Gorbon et Morel, 1983). Les eaux de

« cas 1 » sont ainsi celles pour lesquelles le phytoplancton et produits dérivés dus à la lyse des compartiments cellulaires (pigments, débris organiques, composés jaunes, etc.) influencent majoritairement les propriétés optiques de l'océan (Morel, 1988). A l'inverse, les propriétés optiques des eaux de « cas 2 » varient principalement avec les apports terrestres (apports des rivières, influence anthropique, etc.) et océaniques (sédiments remis en suspension, etc.).



**Figure 17** : Pénétration des radiations lumineuses dans trois types d'environnements aquatiques : océan ouvert (station ALOHA de l'océan pacifique), eaux côtières (mer Baltique) et eaux turbides de lacs (lac Groote Moost). (A) Spectres d'absorption lumineuse par les molécules d'eau pure (ligne bleue), et l'ensemble gilvin + tripton (ligne rouge). Dénommés de façon différente par les auteurs (acides humiques, acides fulviques, acides organiques jaunes, etc.), Kirk a proposé le terme « gilvin » (du latin gilvus: jaune pâle) pour représenter l'ensemble de ces composés (Kirk, 1994). Le tripton représente quant à lui l'ensemble des particules non vivantes, minérales ou organiques, en suspension dans la colonne d'eau (Kirk, 1994). L'absorption lumineuse due au gilvin et tripton est faible dans l'environnement oligotrophe de la station ALOHA, intermédiaire en mer Baltique et élevée dans les eaux turbides du lac Groote Moost. (B) Qualité du spectre lumineux pénétrant dans la couche photique en fonction de la profondeur. (C) Couleurs des environnements aquatiques résultant des propriétés optiques de chacun des environnements, expliquant la couleur bleue de l'océan ouvert, verte des environnements côtiers et rouge à brun des eaux turbides lacs eutrophes. Modifié d'après Stomp (2008) selon données de Stomp et al. (2007).

### II.1.2.3. *La disponibilité en nutriments*

Jack Myers a qualifié les algues unicellulaires "d'organismes petits en taille, composés d'une machinerie cellulaire riche en protéines et potentiellement capables d'un taux métabolique élevé" (Myers, 1980). Ces microorganismes sont en effet composés majoritairement de protéines, qui peuvent représenter de 30 à 60% de la masse cellulaire (Geider et Roche, 2002). Ces composants contiennent la majeure partie de l'azote de la cellule (Lourenço et al., 1998), pouvant même atteindre 85% en cas de carence en azote (Laws, 1991). Les complexes de capture de la lumière des cyanobactéries (ou phycobilisomes, PBS) requiert à eux seuls une grande quantité d'azote et fournirait une source d'azote en cas de carence en cet élément (Wyman et al., 1985). Le phosphore est quand à lui un élément essentiel dans la composition de l'ADN et de la machinerie protéique ribosomale, alors que le fer constitue un élément essentiel au niveau de l'appareil photosynthétique (Falkowski, 2000). Ces trois éléments sont donc indispensables au métabolisme du phytoplancton et, en cas de faible disponibilité dans le milieu, peuvent limiter la croissance de ces organismes (Falkowski, 1997, 2000; Morel et Price, 2003; Arrigo, 2004; Mills et al., 2004).

En 1934, Redfield remarqua que le rapport entre l'azote inorganique et le phosphore était remarquablement constant dans le milieu marin (Redfield, 1934). Par la suite étendu au carbone, le rapport C:N:P est ainsi de 106:16:1 (Redfield, 1934, 1963). Selon les environnements océaniques ce rapport varie selon la disponibilité en azote et phosphore. La majorité des environnements oligotrophes sont caractérisés par une faible teneur en azote, induisant un rapport N:P inférieur à 16:1. Certains écosystèmes sont en revanche pauvres à la fois en azote et en phosphore, qui peuvent alors être co-limitants, comme c'est le cas en Méditerranée Est et dans la mer des Sargasses (Krom et al., 1991; Cotner et al., 1997; Thingstad et al., 2005). Les couches supérieures de la couche photique sont généralement pauvres en nutriments, mais la concentration de ces derniers augmente en profondeur, à la base du maximum profond de Chl. En zone tempérée, un enrichissement nutritif des eaux de surface se produit lors du brassage hivernal et est généralement suivi d'efflorescences phytoplanctoniques, généralement observées au printemps (Lindell et Post, 1995; Coale et al., 1996). Ce phénomène saisonnier n'est cependant pas ou peu observé en zone tropicale et subtropicale, dont l'oligotrophie est quasi permanente, exception faite du brassage du aux cyclones. Diverses études suggèrent que la surface des zones oligotrophes serait actuellement en expansion, notamment du fait de l'augmentation du degré de stratification du au réchauffement climatique (Behrenfeld et al., 2006; Polovina et al., 2008; Irwin et Oliver, 2009), ce qui ne devrait faire que renforcer l'importance relative du picophytoplancton dans la production primaire océanique (Li, 2002; Irwin et al., 2006; Finkel et al., 2007).

#### II.1.2.4. La température

Comme pour la lumière, l'effet de la seule température n'est pas simple à étudier dans l'environnement océanique, car la modification de cette variable induit une réponse multifactorielle, se traduisant à la fois au niveau cellulaire, mais également sur les conditions physico-chimiques du milieu environnant, notamment sur la densité ou la viscosité (Finkel et al., 2010). A une échelle plus globale, la température influe sur la circulation océanique, le degré de stratification et contrôle ainsi les apports nutritionnels auxquels le phytoplancton est soumis.

La température influe directement sur les capacités métaboliques et enzymatiques des cellules, une augmentation modérée induisant jusqu'à 25% d'augmentation du taux de croissance (Eppley, 1972; Moore et al., 1995). Au dessus de l'optimum de température, une augmentation de ce paramètre peut par contre induire des dommages au niveau cellulaire, altérer la capacité photosynthétique, entraîner une inhibition de la croissance, voire la mort cellulaire.

L'effet de la température sur le métabolisme est variable selon les genres, espèces ou écotypes étudiés. *Synechococcus* semble globalement plus sensible aux fortes températures que *Prochlorococcus* (Fu et al., 2007), et ces deux genres possèdent des écotypes adaptés à différentes gammes de température (Zinser et al., 2007; Zwirgmaier et al., 2008). La réponse métabolique des espèces marines varie également selon les niveaux trophiques et les écosystèmes étudiés, et est susceptible de modifier la structure des communautés phytoplanctoniques (Eppley, 1972; Moisan et al., 2002; Edwards et Richardson, 2004) et la dynamique des efflorescences (Rose et Caron, 2007). Une baisse de la capacité d'assimilation de l'azote a été rapportée chez différentes espèces phytoplanctoniques face à une augmentation de température, mais ne semble pas affecter significativement les cyanobactéries *Trichodesmium*, *Synechococcus* et *Prochlorococcus* chez qui la composition élémentaire C:N:P reste stable (Fu et al., 2007; Hutchins et al., 2007).

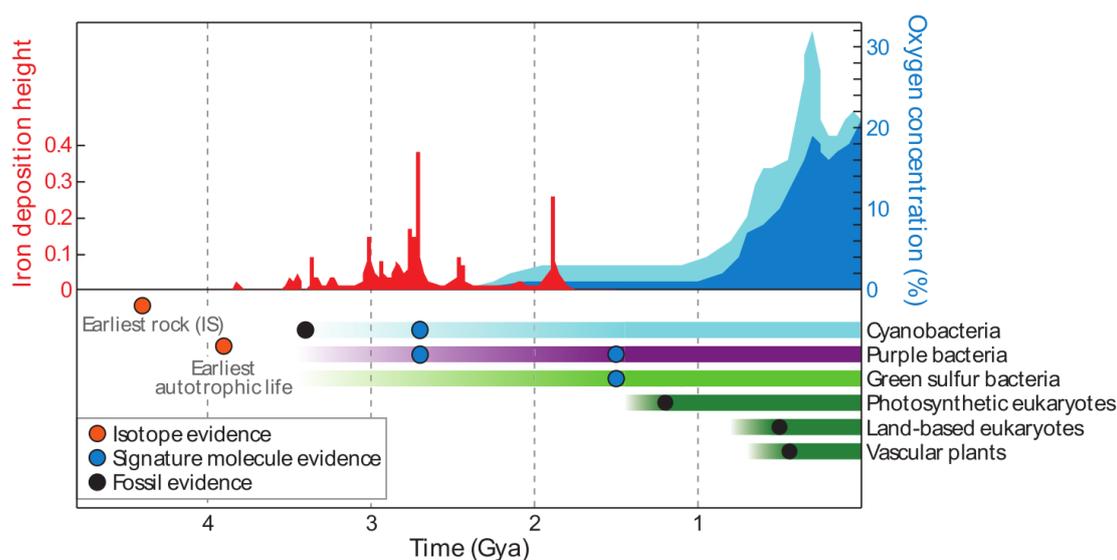
## **II.2. La compétition pour la lumière dans l'océan**

### **II.2.1 La photosynthèse oxygénique**

L'utilisation de la lumière comme source principale d'énergie (phototrophie) est un des processus biologiques les plus importants sur Terre, comme en attestent la signature spectrale de notre planète (Kiang et al., 2007) et la teneur élevée en oxygène de son atmosphère (Raven, 2009). Parmi les organismes phototrophes, plusieurs lignées bactériennes sont capables de photosynthèse, c'est-à-dire de convertir l'énergie lumineuse en énergie de liaison chimique. On distingue ainsi les bactéries oxyphototrophes, les bactéries vertes sulfureuses et filamenteuses, les héliobactéries, les bactéries pourpres sulfureuses et non sulfureuses, ainsi que les phototrophes aérobies anoxygéniques

et les bactéries à protéorhodopsine. En fonction de la source d'électrons utilisée pour la photosynthèse, qui peut être de l' $H_2O$ ,  $H_2S$ ,  $H_2$  ou d'autres composés réduits inorganiques, on peut distinguer deux grands types de photosynthèse. Les cyanobactéries sont les seuls procaryotes pratiquant le type de photosynthèse aujourd'hui le plus répandu sur Terre, la photosynthèse oxygénique, *i.e.* qui s'accompagne d'un dégagement d'oxygène. Les autres organismes cités précédemment pratiquent une photosynthèse anoxygénique, considérée comme "primitive". Les mécanismes permettant la capture de la lumière, les pigments associés aux centres réactionnels, ainsi que la structure et le fonctionnement de ces derniers diffèrent entre ces organismes. Les complexes collecteurs de lumière utilisés par les cyanobactéries seront décrits en détail dans les paragraphes §II.2.4 et §II.2.5.

La période exacte d'apparition de la photosynthèse oxygénique est une question encore vivement débattue, et varie selon les études et les marqueurs utilisés (traces fossiles, indicateurs chimiques, études génétiques, etc.). On estime aujourd'hui qu'elle se situerait entre 3,5 et 2,7 milliards d'années (Fig. 18; Awramik, 1992; Canfield et al., 2000; Farquhar et al., 2007; Buick, 2008). Pendant des millions d'années, les microorganismes capables de fixation d'azote atmosphérique, de réduction de sulfate, de méthanogénèse et de photosynthèse oxygénique ont contribué à un changement radical des conditions physico-chimiques de la planète. La Terre est ainsi passée d'une atmosphère riche en dioxyde de carbone à une atmosphère riche en dioxygène, dans laquelle nous vivons actuellement (Bekker et al., 2004). Ces changements ont contribué à l'émergence de nouveaux environnements, terrestres et aquatiques, rapidement colonisés par les organismes utilisant la respiration oxygénique.

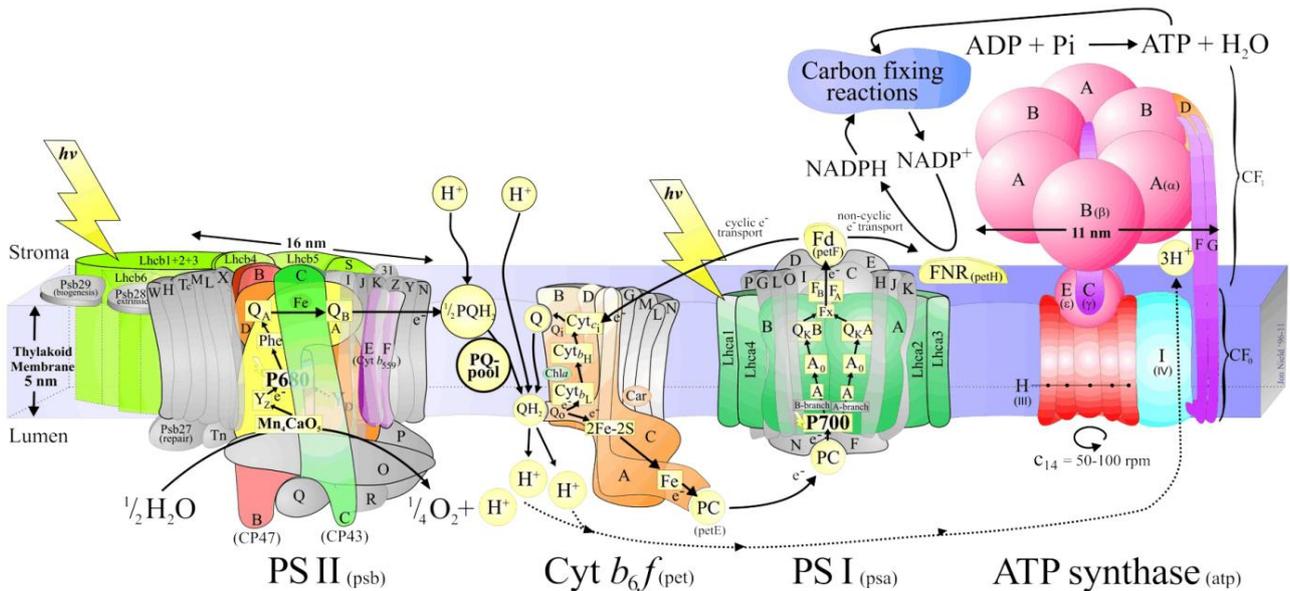


**Figure 18** : Apparition de la photosynthèse, augmentation de la teneur en oxygène atmosphérique et évolution du vivant. Les estimations minimales et maximales de la concentration d'oxygène sont indiquées en bleu foncé et bleu clair respectivement. D'après Hohmann-Marriott et Blankenship (2011).

L'accumulation d'oxygène dans l'atmosphère à des taux significatifs aurait débuté il y a 2,4 milliards d'années (Bekker et al., 2004; Buick, 2008). Deux phases peuvent être distinguées au cours desquelles la teneur en oxygène (1) serait restée stable pendant des millions d'années autour de 1 à 2 % puis (2) aurait augmenté considérablement il y a 1 milliard d'années pour atteindre des concentrations similaires à celles d'aujourd'hui, c'est-à-dire autour de 20 %. Cette seconde période aurait suivi l'apparition des organismes eucaryotes photosynthétiques et des plantes supérieures (Fig. 18; Holland, 2006).

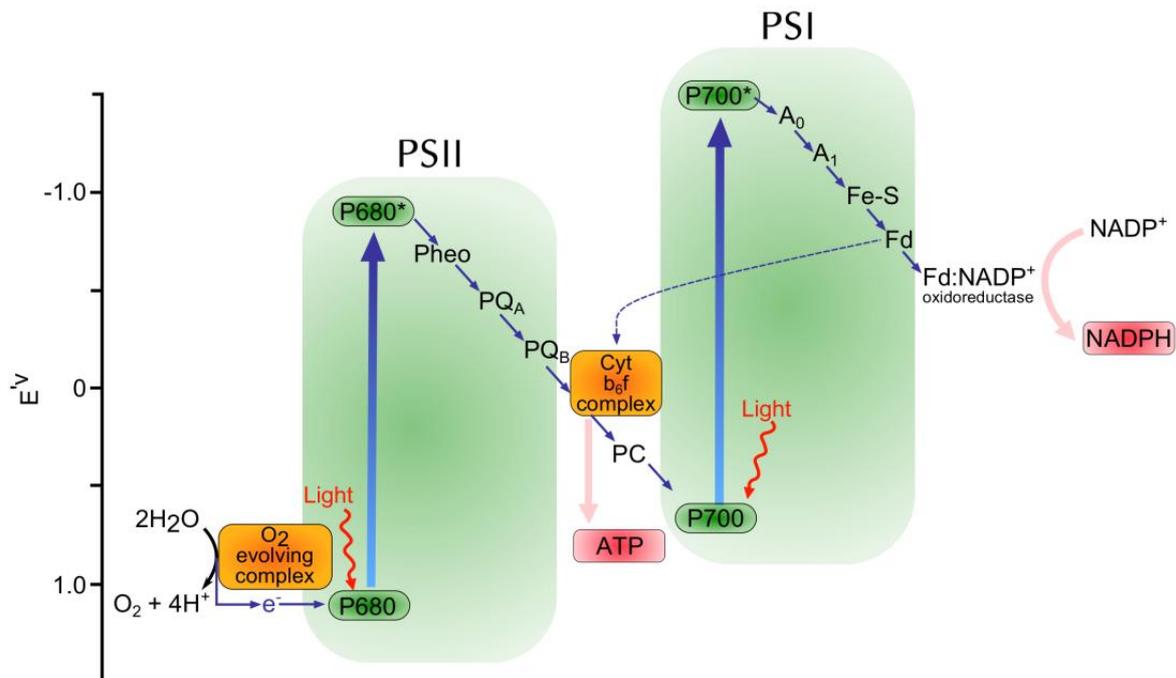
### II.2.2 L'organisation générale de l'appareil photosynthétique oxygénique

Les réactions photosynthétiques s'effectuent au niveau des thylacoïdes, délimitant un compartiment interne, le lumen, au sein du stroma du chloroplaste ou du cytosol chez les procaryotes. Ces membranes contiennent quatre complexes multiprotéiques à l'origine d'un transfert linéaire d'électrons ("non cyclique") : deux photosystèmes I et II (PSI et PSII), le cytochrome *b<sub>6</sub>f* et l'ATP synthase (Fig. 19). Les deux PS diffèrent par leur composition en protéines et en cofacteurs, par leurs états d'agrégation, mais leur fonctionnement est néanmoins similaire. Chacun de ces PS est constitué d'un centre réactionnel (CR) se composant de deux protéines majeures liant de la Chl et d'une antenne interne, formée d'un assemblage de protéines et de pigments. Les deux systèmes agissent en série pour catalyser les mouvements d'électrons, transportés entre les deux PS via le cytochrome *b<sub>6</sub>f*, ce qui peut être représenté par un schéma en Z (Fig. 20).

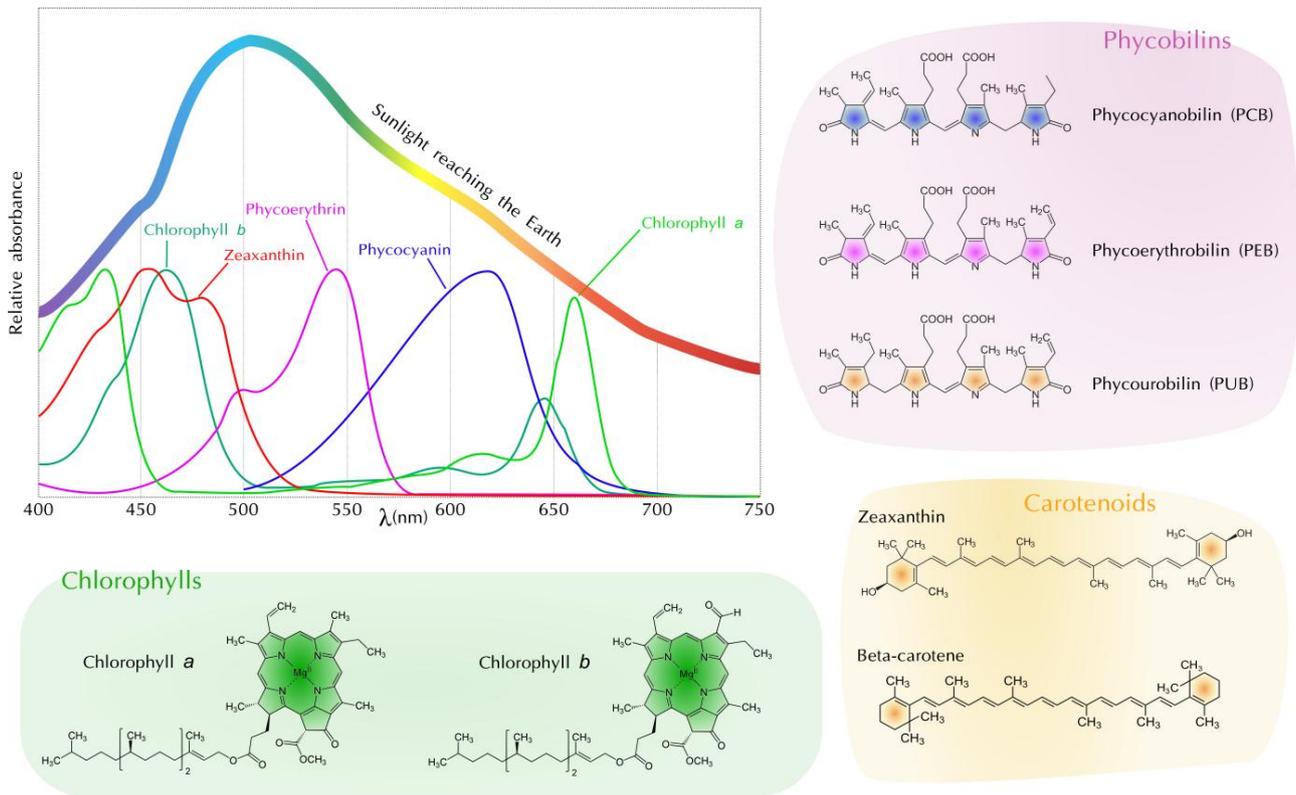


**Figure 19 : Organisation de l'appareil photosynthétique d'une plante vasculaire.** Les quatre complexes multiprotéiques impliqués dans les réactions de photosynthèse oxygénique sont les deux photosystèmes I et II, le cytochrome *b<sub>6</sub>f* ainsi que l'ATP synthase. D'après <http://www.queenmaryphotosynthesis.org/nield/psIIimages/oxygenicphotosynthmodel.html>

Au sein du spectre électromagnétique émis par le soleil, le spectre utile pour les organismes photosynthétiques, ou PAR pour « Photosynthetic Active Radiation », est compris entre 400 nm et 900 nm (Fig. 21). Parmi les pigments associés aux complexes photosynthétiques permettant la capture et le transfert de l'énergie lumineuse, la Chl *a* est considérée comme le pigment essentiel de la photosynthèse car il réalise la séparation de charge. Cette molécule est capable d'absorber à la fois les radiations bleues (maximum absorption  $\lambda_{\max} \sim 440\text{nm}$ ) et rouges ( $\lambda_{\max} \sim 680\text{nm}$ ). Afin de maximiser la capacité d'absorption de la lumière et de coloniser de nouvelles niches écologiques, les cellules algales et en particulier les cellules phytoplanctoniques ont développé ou acquis une grande variété de pigments, considérés comme « accessoires » comme les Chl *b*, *c* et *d*, les PBPs et de nombreux caroténoïdes (Fig. 21). Cette diversité de pigments permet ainsi d'élargir la gamme spectrale de lumière pouvant être exploitée, et certains de ces composés sont aussi très impliqués dans des mécanismes de photoprotection.



**Figure 20** : Schéma en "Z" représentant le transfert d'électrons lors de la photosynthèse oxygénique conduisant à la photolyse de l'eau conduisant à la formation de dioxygène, d'ATP et de NADPH. Les chlorophylles P680 et P700 sont les donneurs primaires des photosystèmes II et I, respectivement. Pheo : phéophytine;  $PQ_A$  : plastoquinone A;  $PQ_B$  : plastoquinone B; PC : plastocyanine;  $A_0$  et  $A_1$  : accepteurs d'électrons; Fe-S : protéine fer-soufre;  $F_d$  : ferredoxine. La ligne discontinue représente la voie de transfert cyclique des électrons, qui n'implique que le photosystème I.



**Figure 21** : Spectres d'absorption de la lumière visible des pigments majoritaires retrouvés au sein des cellules phytoplanctoniques et structures chimiques associées. Trois familles de photopigments sont montrées avec les caroténoïdes, les chlorophylles ainsi que les chromophores fixés au niveau des phycobiliprotéines.

## II.2.3 Les photosystèmes

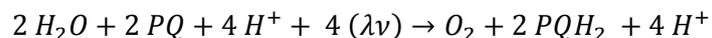
### II.2.3.1. *Le photosystème II*

L'énergie lumineuse capturée par les antennes collectrices de lumière est transférée au CR du PSII. Une vingtaine de protéines, auxquelles sont associées divers cofacteurs et pigments, forment le CRII (Umena et al., 2011). Toutes les protéines liant de la Chl lient également des caroténoïdes (Green et Durnford, 1996), qui jouent un rôle dans la photoprotection (Niyogi, 2000; Telfer, 2005). Le CR proprement dit est formé par un hétéroduplex protéique, D1 et D2. Une antenne interne composée par les protéines CP43 et CP47, lui est associée (Bricker, 1990). Ces quatre protéines sont pigmentées, mais la plupart des molécules de Chl *a* et de  $\beta$ -carotène sont fixées à l'antenne interne (Pakrasi, 1995; Rhee et al., 1998). L'énergie lumineuse capturée par les antennes collectrices est transmise au donneur primaire du CR, constitué d'une paire de Chls spéciales, appelée P680 (Kühlbrandt et al., 1994; Durrant et al., 1995). En modifiant leur potentiel redox, ces dernières se retrouvent dans un état excité (P680\*), leur permettant ainsi de céder un électron à des molécules de phéophytine. Les électrons sont ensuite transférés de cofacteur en cofacteur à travers des réactions d'oxydoréduction. La phéophytine transmet ainsi son électron à la plastoquinone A puis à

la plastoquinone B. Une fois totalement réduite, en ayant accepté deux électrons de la plastoquinone A et deux protons à partir de l'H<sub>2</sub>O, la plastoquinone B initie le transfert d'électrons au sein des membranes thylacoïdales. Ces derniers sont alors transmis au PSI *via* le cytochrome *b<sub>6</sub>f* et par l'intermédiaire d'une chaîne de transporteurs constituée notamment du pool de plastoquinone, du cytochrome *b<sub>6</sub>f* et de la plastocyanine.

Pour revenir à un état réduit, P680\* récupère un électron *via* un résidu tyrosine (TyrZ) de la protéine D1. Chez les organismes pratiquant la photosynthèse oxygénique, cet électron est issu de la photolyse de l'H<sub>2</sub>O, au niveau de l'OEC (pour "Oxygen Evolving Complex"; parfois appelé WOC pour "Water oxidizing complex"; Ghanotakis et Yocum, 1990; Nugent et al., 2001). Le site catalytique de l'OEC est formé par un cluster manganèse (Mn<sub>4</sub>Ca) générant 4 électrons ainsi qu'une molécule d'O<sub>2</sub> pour deux molécules d'H<sub>2</sub>O protolysées (Umena et al., 2011). Le cluster manganèse est indirectement stabilisé par différentes protéines extrinsèques, situées dans la lumière des thylacoïdes, participant au maintien des ions requis comme cofacteurs pour l'oxydation de l'H<sub>2</sub>O (Seidler, 1996; De Las Rivas et al., 2004; Roose et al., 2007).

Le PSII peut ainsi être décrit comme une H<sub>2</sub>O-plastoquinone oxidoréductase, nécessitant l'énergie de quatre photons et effectuant la réaction biochimique suivante :



### II.2.3.2. *Le photosystème I*

Le PSI absorbe dans le rouge lointain, grâce à une paire de Chls, appelée P700 (Kok et Gott, 1960; Krauss et al., 1993). Ce macrocomplexe pigments-protéines est composé d'une quinzaine de sous-unités protéiques. Dans le PSI, le CR et l'antenne interne sont fusionnés et constitués par les deux protéines principales PsaA et PsaB, portant un grand nombre de molécules de Chl *a* (~90) et de  $\beta$ -carotène (10-15). Il est intéressant de noter qu'il existe une forte similarité structurale d'une part entre les extrémités N-terminales de PsaA/B et CP43/CP47 qui forment l'antenne interne du PSII, et d'autre part entre les domaines C-terminaux de PsaA/B et D1/D2 formant le CRII (Barber et al., 2000; Ferreira et al., 2004), ces différents complexes ayant probablement une origine évolutive commune (Olson et Blankenship, 2004).

L'hétéroduplex PsaA/B lie les différents transporteurs d'électrons et de cofacteurs que sont la paire de Chl *a* P700, la Chl *a* (A<sub>0</sub>), la phylloquinone (A<sub>1</sub>) et les trois clusters Fer-Soufre (Fe-S; Brettel, 1997; Scheller et al., 1997; Brettel et Leibl, 2001). Les électrons sont ainsi transportés jusqu'à la ferrédoxine liant deux groupements Fer-Soufre, puis à la ferrédoxine-NADP<sup>+</sup> réductase (FNR). Cette

dernière est responsable du passage d'électrons de la ferrédoxine jusqu'au  $\text{NADP}^+$ , produisant ainsi une molécule au puissant pouvoir réducteur, le NADPH (Fig. 20). Alternativement au flux d'électrons "non cyclique" décrit précédemment, il existe une deuxième voie de transfert, dite "cyclique", impliquant uniquement le PSI (Bendall et Manasse, 1995; Joliot et Joliot, 2002). Dans cette voie, les électrons arrivés à la ferrédoxine retournent vers le cytochrome  $b_6f$  qui les retransmettra au PSI via la plastocyanine au niveau du PSI. Cette voie permet au PSI de fonctionner en continu, indépendamment du PSII. Cette voie contribue au gradient de protons par le cytochrome  $b_6f$  mais ne génère pas de NADPH.

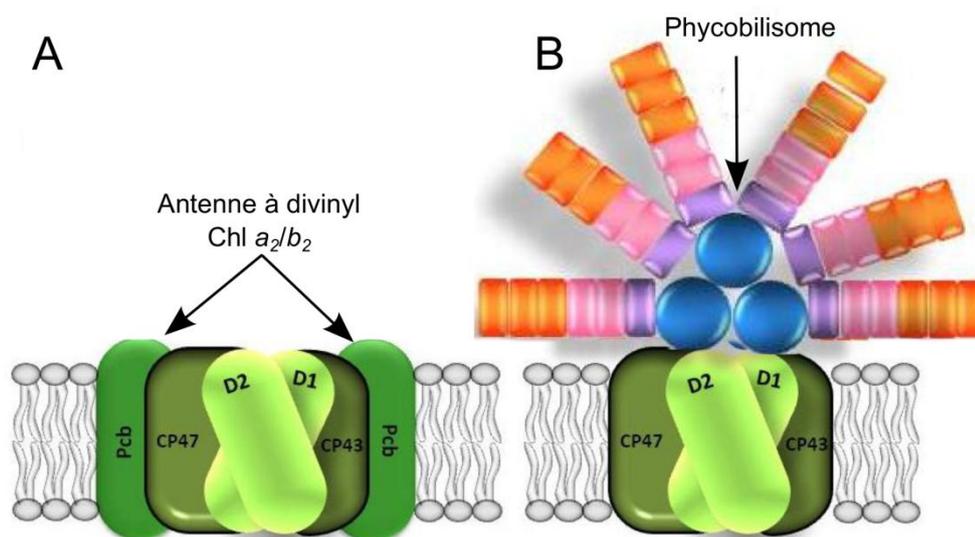
### II.2.3.3. Production d'énergie au sein des membranes thylacoïdales

Le transfert d'électrons du PSII vers le PSI est réalisé par l'intermédiaire du cytochrome  $b_6f$ , catalysant l'oxydation de la plastoquinone et la réduction de la plastocyanine. Ce complexe joue un rôle central dans la transformation et le stockage d'énergie en une forme de plastoquinone réduite ( $\text{PQH}_2$ ). Puisque la plastoquinone prélève les deux protons dans le stroma (ou le cytosol chez les procaryotes), cette réaction participe à la création d'un gradient de pH électrochimique transthylacoïdal. L'acidification de la lumière des thylacoïdes va générer un flux de protons au travers de complexes ATP synthétase (Fig. 19), permettant la phosphorylation de l'ADP en ATP dans le stroma ou le cytosol chez les procaryotes. La photosynthèse oxygénique permet ainsi la production de molécules ATP et NADPH possédant un haut pouvoir réducteur et qui seront utilisées pour incorporer une molécule de  $\text{CO}_2$  sur le ribulose 1, 5 biphosphate grâce à la Rubisco au sein du cycle de Calvin-Benson.

## II.2.4 Les antennes collectrices de *Prochlorococcus*

Contrairement à la majorité des cyanobactéries, *Prochlorococcus* ne possède pas de PBS pour capturer l'énergie lumineuse (Fig. 22). Cette antenne extrinsèque aux membranes thylacoïdales a été remplacée par une antenne intrinsèque originale, composée d'un assemblage de protéines fixant des dérivés divinyles de la Chl  $a/b$  (DVChl), uniques à ce genre (Chisholm et al., 1988). Ce type d'antenne est également trouvé chez plusieurs autres cyanobactéries atypiques, notamment *Prochloron* et *Prochlorothrix*, qui fixent de la Chl  $a/b$  'normale' (monovylée), comme les plantes (Lewin et Withers, 1975; Burger-Wiersma et al., 1986). Cette similarité avait conduit à regrouper ces cyanobactéries dans la sous-classe des « Prochlorophyta » (*pro* pour avant, et *chlorophyta* pour plantes vertes; Lewin, 1976), par opposition aux Cyanophyta possédant des phycobilines. Néanmoins le caractère polyphylétique de ces trois genres au sein de la radiation des cyanobactéries (Turner et al., 1999), a

rendu cette dénomination obsolète et a été remplacée par le terme « oxyphotobactérie verte » qui désigne les procaryotes photosynthétiques possédant une antenne intrinsèque à DVChl *a* et *b*, sans assignation phylogénétique particulière (Partensky et Garczarek, 2003).

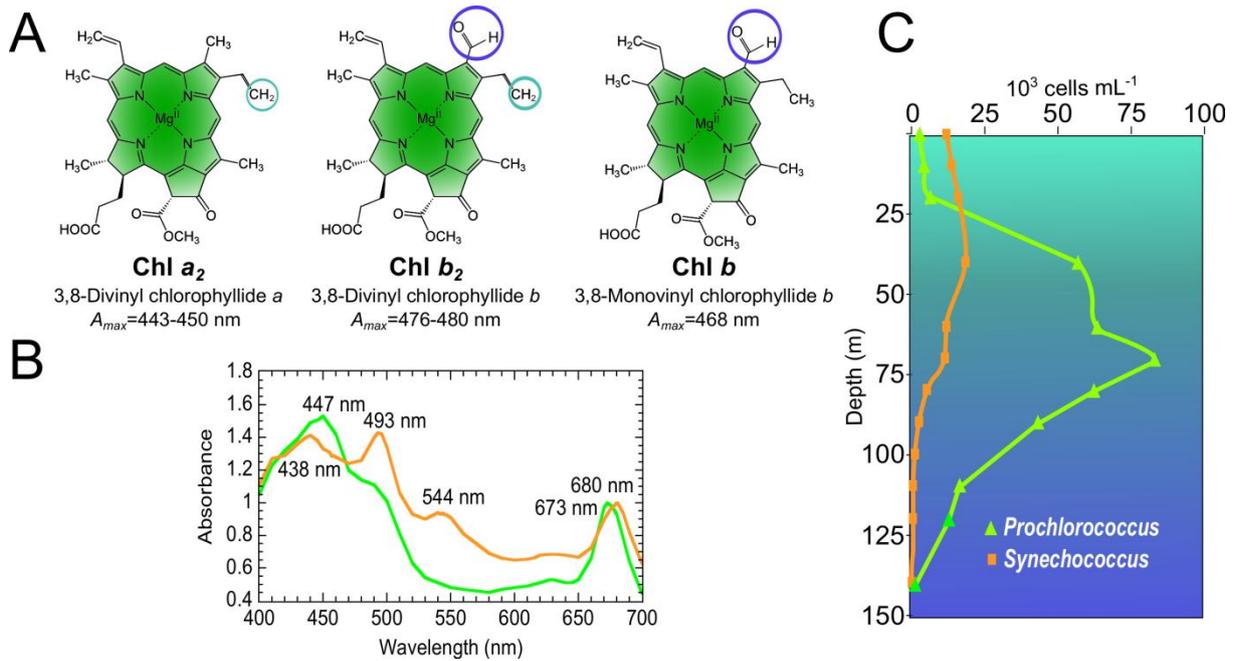


**Figure 22 : Antennes collectrices de lumière retrouvées au sein des picocyanobactéries marines au niveau du photosystème II. (A) *Prochlorococcus* possède une antenne intrinsèque à la membrane thylakoïdale, alors que (B) *Synechococcus* possède une antenne extrinsèque, similaire à la majorité des cyanobactéries : le phycobilisome. Modifié d'après Mella-Flores (2011).**

Les protéines d'antenne des oxyphotobactéries vertes, nommées Pcb (pour *Prochlorophyte* Chl *b* binding protein) possèdent six hélices transmembranaires liant des molécules de Chl *a* et *b* (LaRoche et al., 1996; Barber et al., 2000), ce qui les distingue des protéines LHC (pour Light Harvesting Complex) des plantes et algues vertes qui ne possèdent que trois hélices (Kühlbrandt et al., 1994; Barber et al., 2000). Le nombre de résidus Histidine, impliqués dans la fixation des pigments sur les Pcb, a été évalué entre 8 et 11 et serait donc comparable à celui des constituants majeurs de l'antenne interne du PSII des organismes oxyphototrophes, CP47 (PsbB) et CP43 (PsbC) qui en portent respectivement 12 et 10 (Barber et al., 2000; Garczarek et al., 2000; Ting et al., 2002). Phylogénétiquement les protéines Pcb sont classées au sein d'une famille protéique incluant CP43, CP47, ainsi que la protéine IsiA, une protéine spécifique des cyanobactéries et qui lie de la Chl *a*. IsiA qui est également nommée CP43' en raison d'une forte similarité de séquence et structurale avec la protéine CP43 (Bibby et al., 2001b; Bricker et Frankel, 2002), est abondamment synthétisée en cas de carence en fer chez certaines cyanobactéries d'eau douce (Laudenbach et Straus, 1988; Leonhardt et Straus, 1992, 1994). Le contenu protéique et l'efficacité des centres réactionnels des PS se trouvent ainsi réduits en cas de carence en fer (Öquist, 1974; Sandström et al., 2002; Küpper et al., 2008). Le rôle de la protéine CP43' a longtemps été controversé. Initialement il était suggéré que la protéine n'était pas directement impliquée dans la photosynthèse, mais servait à dissiper l'énergie excédentaire arrivant au niveau du PSII (Park et al., 1999; Cadoret et al., 2004; Sarcina et Mullineaux,

2004; Singh et Sherman, 2007). L'isolement d'un supercomplexe IsiA-PSI, dans lequel 18 monomères d'IsiA sont couplés à un trimère de PSI a conduit à suggérer un rôle compensatoire de cette protéine dans la capture des photons par le PSI, réduit lors d'une carence en fer (Bibby et al., 2001b, a; Bibby et al., 2001c; Boekema et al., 2001). Il a récemment été démontré que l'accumulation d'IsiA en cas de carence ferrique induisait une augmentation de 60% de l'efficacité du PSI, ce qui tend à confirmer le rôle important de cette protéine dans la capture lumineuse au niveau du PSI (Ryan-Keogh et al., 2012).

Au niveau évolutif, les protéines Pcb seraient issues de duplications et divergences de protéines ancestrales de type IsiA-like ou CP43. Compte tenu du rôle central du fer dans le métabolisme des cyanobactéries, cet élément a pu être une force sélective majeure dans l'évolution de l'antenne des oxyphotobactéries vertes (Martin et al., 1994; Coale et al., 1996; Mann et Chisholm, 2000). De manière intéressante chez *Prochlorococcus* l'antenne du PSII est constitutive alors que celle associée au PSI semble être plus variable. Si cette dernière est constitutive chez la souche SS120 (LL), elle n'est synthétisée qu'en cas de carence en fer (comme les protéines IsiA) chez la souche LL MIT9313, et est absente chez MED4 (HL; Bibby et al., 2001a, 2003a). Cette adaptation physiologique s'explique par des différences au niveau génomique entre les différents écotypes de *Prochlorococcus*. Les souches HL possèdent une (e.g. cas spécifique de la souche MED4) à deux copies du gène *pcb* (Garczarek et al., 2000), une pour les protéines associées au PSII, l'autre codant pour celles associées au PSI (Garczarek et al., 2007). En revanche chez les souches LL, on trouve davantage de copies du gène *pcb*, variant de 6 à 8 copies (Garczarek et al., 2000; Garczarek et al., 2001; Bibby et al., 2003a; Bibby et al., 2003b). Le scénario évolutif actuellement retenu propose qu'il y ait eu initialement une copie d'un gène *isiA*-like chez l'ancêtre des picocyanobactéries marines et que des séries de duplications géniques auraient ensuite eu lieu durant l'histoire évolutive du genre *Prochlorococcus*. Le premier évènement de duplication du gène *pcb* a conduit à la différenciation entre une copie associée constitutivement au PSII, alors que l'autre copie continuait à coder une protéine, induite en cas de carence en stress et associée au PSI. La multiplication des gènes d'antenne au sein des souches LL seraient à l'origine de l'adaptation évolutive à un environnement faiblement éclairé.



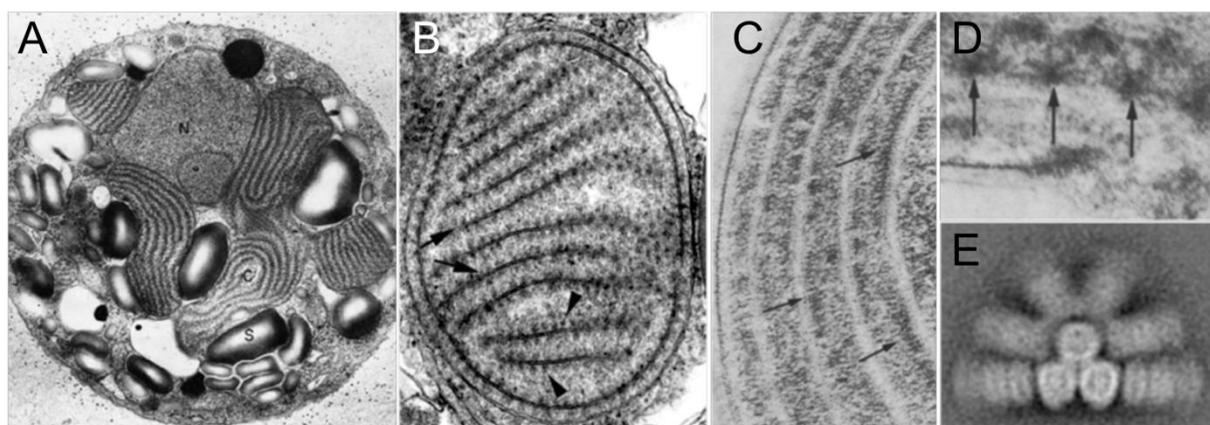
**Figure 23** : Pigments majoritaires de la picocyanobactérie marine *Prochlorococcus* et implications écologiques. **(A)** Structure des pigments chlorophylliens majoritaires, présents dans les cellules du genre *Prochlorococcus*. Les chlorophylles  $a_2$  et  $b_2$  sont des dérivés divinylés de la chlorophylle  $a$  et  $b$  respectivement. Les groupements chimiques différant entre ces Chl sont entourés. Les chaînes aliphatiques de phytol ne sont pas représentées. **(B)** Spectre d'absorption in vivo de *Prochlorococcus* MED4 (HLI, vert) et de *Synechococcus* WH8102 (orange) cultivées sous les mêmes conditions lumineuses. **(C)** Distribution verticale des cellules picocyanobactériennes en Atlantique subtropical. Modifié d'après Ting et al. (2002).

L'antenne originale à divinyl Chl  $a$  de *Prochlorococcus* permet à ces cellules de croître en surface à de fortes intensités lumineuses, mais également en profondeur près du bas de la couche photique, où la lumière bleue prédomine. Les DVChl, dont le pic d'absorbance est décalé de 8-10 nm vers le rouge absorbent plus efficacement la lumière bleue que les Chl monovinylées (MVChl; Fig. 23). *Prochlorococcus* est en revanche capable de croître à de fortes intensités lumineuses, soulignant que cette évolution pigmentaire a été concomitante avec une évolution des antennes photosynthétiques et/ou des centres réactionnels liant ces DVChl (Moore et al., 1998; Ito et Tanaka, 2011). Récemment il a été démontré que la protéine D1 de *Prochlorococcus* présentait deux acides aminés modifiés comparés aux autres organismes photosynthétiques oxygéniques (Ito et Tanaka, 2011). Une de ces deux modifications est également présente chez *Acaryochloris marina*, qui capture la lumière infrarouge en utilisant la Chl  $d$  comme pigment majoritaire (Chen et Blankenship, 2011). Ces modifications, en permettant un transfert énergétique plus efficace de CP43 à CP47, pourraient induire une réduction du nombre des photons transférés de l'antenne aux CRs, et permettrait à *Prochlorococcus* de croître à de plus fortes intensités lumineuses (Yokono et al., 2012). Au niveau évolutif, en raison de l'apparente toxicité de la DVChl sur des microorganismes possédant une protéine D1 ancestrale, il semble probable que l'ancêtre de *Prochlorococcus* possédait des protéines D1 de type DVChl, et que l'acquisition de DVChl ait été ultérieure (Ito et Tanaka, 2011).

## II.2.5 Les phycobilisomes de *Synechococcus*

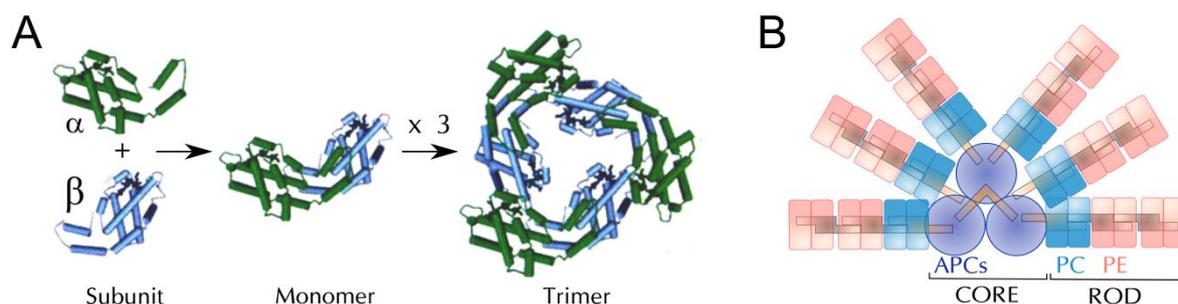
### II.2.5.1. Structure des phycobilisomes

La plupart des cyanobactéries, dont les *Synechococcus* marins, ont des antennes collectrices de la lumière constituées par des macrostructures de grande taille (7,000 à 15,000 kDa), observables en microscopie électronique (Fig. 24). Ces structures, localisées à la surface cytoplasmique des membranes thylacoïdales, sont connues depuis très longtemps et ont fait l'objet de nombreuses études d'un point de vue structural, fonctionnel, physiologique et génétique. En 1836, Nees Esenbeck décrivait un pigment bleu soluble très fluorescent extrait de de la cyanobactérie *Oscillatoria* (Esenbeck, 1836). Quelques années plus tard, Kützing isolait à son tour un nouveau pigment, mais arborant cette fois-ci une fluorescence orange, à partir de différentes algues rouges (Kützing, 1843). Le début du XIXe siècle fut marqué par les nombreux travaux sur les PBP et la proposition du terme « phycobiline » pour désigner les chromophores de ces protéines, par homologie aux pigments retrouvés dans la bile animale (Lemberg, 1928, 1930; Lemberg et Bader, 1933). Ce n'est qu'avec le développement de la microscopie, et notamment des travaux d'Elizabeth Gantt, que l'étude de ces complexes eu un essor important. Gantt et Conti mirent en évidence la présence de petits granules attachés aux lamelles chloroplastiques de l'algue rouge *Porphyridium cruentum* (Gantt et Conti, 1966a). Deux fois plus grands que les ribosomes, ces granules furent dénommés PBS (Gantt et Conti, 1966a, b), et furent par la suite décrits dans un grand nombre d'algues rouges et de cyanobactéries (voir revue de Gantt, 1980).



**Figure 24** : Photos en microscopie électronique à transmission montrant la présence des phycobilisomes (PBS) au niveau des thylacoïdes. (A) Cellule de l'algue rouge *Porphyridium cruentum* vue au MET (Gantt et Conti, 1966a). (B) Chloroplaste de l'algue rouge *Griffithsia pacifica* (Florideophyceae). Les flèches pointent les PBS (Photographie en MET de C.M Pueschel). (C-D) Espaces interthylacoïdaux de la cyanobactérie *Pseudoanabaena* sp. PCC7409 et PBS associés (pointés par les flèches), d'après Bryant et al. (1979). (E) PBS hémidiscoïdal isolé avec un cœur tricylindrique de *Synechocystis* sp. PCC 6803, observé par microscopie électronique à une seule particule, d'après Arteni et al. (2009).

Ces complexes sont présents chez la plupart des cyanobactéries, les glaucocystophytes, les Cryptophytes et chez les Rhodophytes (ou algues rouges; Figs. 1, 24A et B; Toole et Alnutt, 2003). Tous les PBS présentent une structure commune et consistent en un assemblage de PBPs, composés d'un cœur fixé à la surface membranaire des thylacoïdes, sur lesquels sont fixés des projections radiales, ou bras (Fig. 24E; voir revues de Glazer, 1985; MacColl, 1998). Ces deux sous-structures lient de façon covalente des chromophores et des protéines le plus souvent dépourvues de pigments, appelées protéines de liaison (ou *linkers*). Les PBS peuvent être classés en trois catégories en se basant sur la structure du cœur : bicylindrique, tricylindrique ou pentacylindrique (Yamanaka et al., 1980; Glauser et al., 1992a; Sidler, 1994; Ducret et al., 1998). D'autres structures encore plus complexes ont également été décrites (Wehrmeyer, 1983; Wehrmeyer et al., 1988). Malgré cette apparente diversité, la formation du PBS est soumise à des contraintes structurales précises (Anderson et Toole, 1998). Les PBS à cœur pentacylindrique possèdent huit bras radiaux, comme cela a été observé chez *Anabaena* (Ducret et al., 1998). Les PBS de type hémidiscoïdal à cœur tricylindrique et six bras radiaux constituent cependant le cas le plus fréquent chez les cyanobactéries (Fig. 25B).



**Figure 25: Structure du phycobilisome (PBS).** (A) Assemblage des sous-unités  $\alpha$  et  $\beta$  en trimère. Le modèle montré provient de la seule structure d'un complexe phycobiliprotéine-protéine de liaison (i.e. allophycocyanine (APC) et protéine de liaison  $L_c$  de *Mastigocladus laminosus*; Reuter et al. (1999)). (B) Structure d'un PBS comprenant un cœur tricylindrique d'APC et de six bras radiaux contenant un seul hexamère de phycocyanine (PC) et deux de phycoérythrine (PE). Les protéines de liaison permettant l'assemblage de ces différentes structures sont indiquées au centre des bras et du cœur.

Les PBPs sont composées de deux peptides, d'environ 15 à 20 kDa, appelées sous-unités  $\alpha$  et  $\beta$  (Fig. 25; Ficner et Huber, 1993). Les deux sous-unités présentent une structure similaire avec un cœur globulaire de type globine formé par au moins six hélices, responsable de l'attachement des chromophores, et une extension N-terminale impliquée dans l'aggrégation de ces sous-unités en hétérodimères, notamment en monomère ( $\alpha\beta$ ) qui correspond au plus petit état d'aggrégation connu chez les PBPs (Schirmer et al., 1986). Néanmoins les monomères présentent une faible stabilité en solution et forment alors des trimères ( $\alpha\beta$ )<sub>3</sub> ou, plus correctement, des hétérohexamères. Ces trimères présentent une structure discoïdale de 110-120 Å (11-12 nm) de diamètre et de 30 Å d'épaisseur, laissant apparaître un canal central de 35 Å de diamètre, où se situe la majorité des

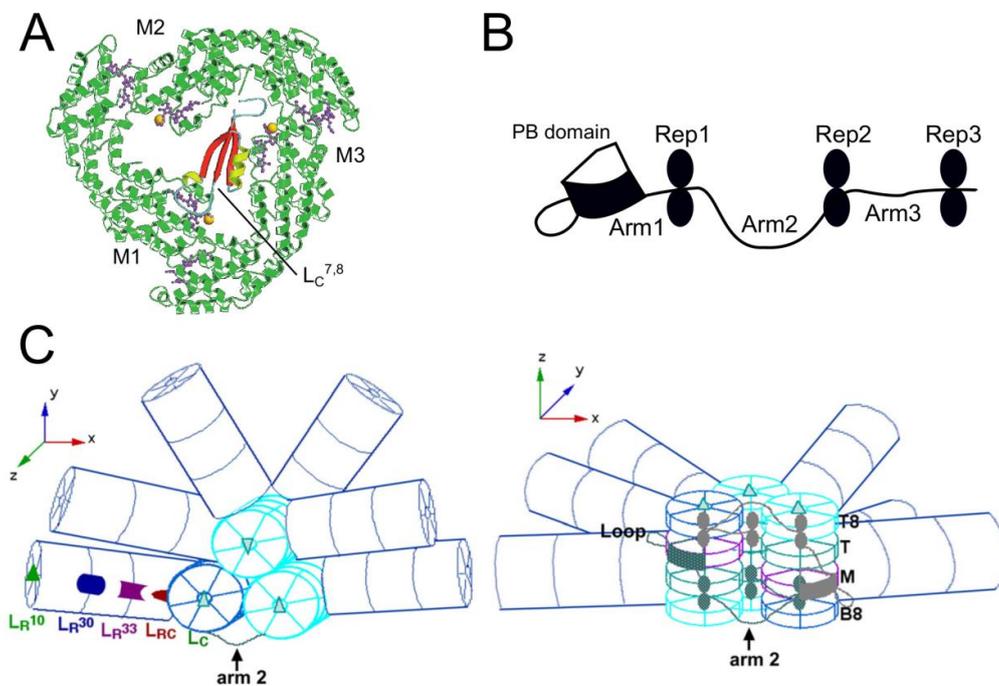
chromophores (Schirmer et al., 1986; Brejc et al., 1995). Ces trimères sont ensuite assemblés en face à face pour former des hexamères  $[(\alpha\beta)_3]_2$  ou, plus exactement, des hétérododécamères, qui constituent les « briques » du PBS. Ces structures s'empilent les unes sur les autres en laissant apparaître un cylindre creux à l'intérieur de chacune de ces projections radiaires. Quatre classes de PBPs ont été décrites : la PC, la PE, la phycoérythrocyanine (PEC) et l'allophycocyanine (APC). Si les capacités photochimiques et spectroscopiques diffèrent largement entre ces différentes classes, les nombreuses études structurales menées à ce jour ont souligné une remarquable similarité entre la PC (Fisher et al., 1980; Schirmer et al., 1986; Duerring et al., 1991), la PE (Ficner et al., 1992), la PEC (Duerring et al., 1990) et l'APC (Brejc et al., 1995) et même une PE retrouvée chez une cryptophyte ne formant pas de PBS (Wilk et al., 1999).

### ***Les protéines de liaison ou "linkers"***

Le niveau maximal d'aggrégation des PBPs s'arrête au stade de trimères ou d'hexamères. L'organisation en complexe supramoléculaire est assurée par des protéines de liaison ou "linkers", représentant de 12 à 15% des composés protéiques du PBS (Tandeau de Marsac et Cohen-Bazire, 1977; Gantt, 1980; Lundell et al., 1981a). Au niveau des bras, ces protéines assurent le maintien et la cohésion des disques de PBS, en formant une "colonne vertébrale" au niveau des canaux centraux, et favorisent ainsi le transfert d'énergie des extrémités distales vers les centres réactionnels (Fig. 25B; Glazer, 1994; Sidler, 1994; Apt et al., 1995; Reuter et al., 1999). En se basant sur les fonctions et poids moléculaires, un système d'abréviations proposé par Glazer a été utilisé pour classer ces protéines de liaison (Glazer, 1985). Ces dernières sont ainsi nommées par un "L" suivi en exposant de leur poids moléculaire et en indice leur localisation (C pour le cœur, R pour les bras ("rod"), RC pour la jonction RC) et CM (pour la jonction entre le cœur et la membrane). En utilisant ce système de classification, 4 classes de protéines de liaison ont ainsi pu être définies : les  $L_R$ , dont le poids moléculaire varie entre 27 et 35 kDa, intervenant uniquement dans l'arrangement des bras; les  $L_{RC}$  (25-27 kDa) assurant l'attachement des projections radiales au cœur du PBS, les  $L_C$  de plus faible poids moléculaire joue un rôle dans la structuration du cœur et enfin les  $L_{CM}$ , dont le poids moléculaire varie entre 70 et 120 kDa suivant le type de PBS, responsables de l'attachement du supracomplexe à la membrane thylacoïdale (Glazer, 1982; Zilinskas et Greenwald, 1986; de Lorimier et al., 1990; Glauser et al., 1992b; MacColl, 1998; Pizarro et Sauer, 2001).

Le rôle principal de ces polypeptides est de connecter les différentes PBPs et ainsi d'assurer la stabilité du supracomplexe protéique. La basicité de ces protéines de liaison ( $pI > 8.0$ ), comparée à la relative acidité ( $pI < 5.0$ ) des PBPs, participe ainsi à ces interactions/liaisons entre les différentes structures protéiques (Lundell et Glazer, 1980). D'autre part, les alignements des séquences de ces protéines ont permis de mettre en évidence un plus faible niveau de conservation comparé à celui

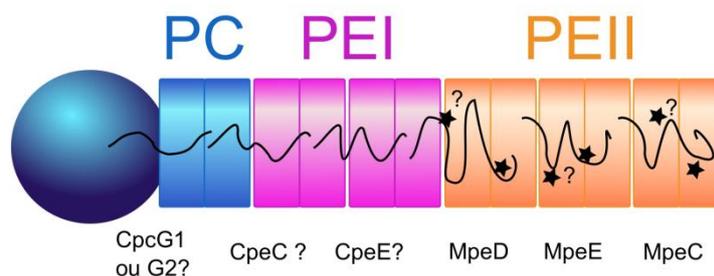
observé pour les PBPs, à l'exception du polypeptide  $L_{CM}$  qui présente une forte similarité entre les différents organismes (Liu et al., 2005). Six domaines conservés ont été mis en évidence au niveau des polypeptides  $L_R$ , près de l'extrémité N-terminale et joueraient un rôle dans la jonction des différentes PBPs au niveau des canaux centraux, alors que les autres régions serviraient d'interface pour l'assemblage en hexamères (Yu et Glazer, 1982; Sidler, 1994; Anderson et Toole, 1998). Les protéines  $L_{RC}$ , possèdent également six domaines conservés, et occuperaient le canal central de la PC (Glauser et al., 1992c), alors que les régions en C-terminal, dont le niveau de conservation est plus faible, interagiraient avec différentes régions du cœur du PBS. En se basant sur des comparaisons de séquence et des expériences de protéolyses, il a été proposé que les protéines de liaison possédaient deux domaines : le domaine N-terminal qui est enfoui dans le canal central du trimère et le domaine C-terminal qui est situé en dehors de l'hexamère (Fig. 26B; Parbel et Scheer, 2000).



**Figure 26** : Rôle des protéines de liaison dans l'assemblage et l'architecture du phycobilisome. **(A)** Structure du complexe allophycocyanine (APC)-protéine de liaison  $L_C^{7.8}$  de *Mastigocladus laminosus*. Les monomères d'APC, notés M, sont représentés en vert, et les chromophores en violet. La structure du polypeptide de liaison est quant à elle représentée en rouge, jaune et bleu (PDB 1B33). D'après Reuter et al. (1999). **(B)** Représentation schématique du polypeptide d'ancrage  $L_{CM}$  et ses différents domaines associés. Le nombre de domaines Rep conditionne l'architecture du cœur du phycobilisome (PBS). D'après Ajlani and Vernotte (1998). **(C)** Représentation d'un PBS hémidiscoïdal vu de face (à gauche) et de la membrane thylacoïdale (à droite). Les différentes phycobiliprotéines des PBS, comportant un cœur tricylindrique et six projections radiales, sont assemblées et maintenues grâce aux polypeptides de liaison. Le maintien du PBS est assuré par le domaine Arm2. D'après Arteni et al. (2009).

La majorité des protéines de liaison sont dépourvues de chromophores et ont un rôle principalement structural. Néanmoins deux types de protéines assurant cette fonction attachent des chromophores. Parmi eux, le  $L_{CM}$  ou polypeptide d'ancrage, codé par le gène *apcE*, est la protéine de

plus haut poids moléculaire du PBS (Lundell et al., 1981a; Zhao et al., 2005b). En fonction du type de PBS (i.e. cœur bicylindrique, tricylindrique ou pentacylindrique), la masse moléculaire du  $L_{CM}$  varie de 70 à 75 kDa, 92 à 99 kDa et de 115 à 128 kDa (MacColl, 2004). Deux copies de ce polypeptide multifonctionnel sont présentes au sein de chaque PBS. Les analyses de séquence ont permis de mettre en évidence différents domaines conservés, dont l'extrémité C-terminale qui possède 2 à 4 domaines répétés (Rep) d'environ 120 résidus. Ces derniers présentent environ 60 % de similarité avec les domaines conservés des protéines  $L_R$ , et jouent ainsi un rôle structurel identique, en intervenant entre les différents disques de PBPs du cœur (Bryant, 1988; Capuano et al., 1991). Le nombre de ces domaines Rep, variant de 2 à 4, conditionne le nombre de cylindres d'APC que le  $L_{CM}$  peut lier, et ainsi l'architecture globale du cœur du PBS (Capuano et al., 1993). L'extrémité N-terminale en revanche présente une très grande similarité avec les PBPs, et est ainsi appelée le domaine PB. Cette similitude a d'ailleurs conduit à proposer que la protéine  $L_{CM}$  était l'ancêtre de toutes les PBPs et protéines de liaison actuelles (Capuano et al., 1991). Le domaine PB se caractérise par la présence d'un résidu Cys, où un PCB est fixé (position 186 chez *Synechococcus* sp. PCC 7002). Il est interrompu par une boucle, d'environ 50 à 65 résidus, qui interviendrait dans l'ancrage à la membrane thylacoïdale (Bryant, 1988; Capuano et al., 1991). Néanmoins, le rôle de cette boucle fait encore débat et est dispensable pour le fonctionnement du PBS (Ajani et Vernotte, 1998). Elle serait ainsi retrouvée effectivement à la surface thylacoïdale mais n'interviendrait pas dans l'ancrage du PBS, cette fonction étant assurée par le domaine Arm2, présent entre les différents domaines REP (Ajani et Vernotte, 1998). Le  $L_{CM}$  a également un rôle dans le transfert d'énergie. Le  $L_{CM}$  modifie les propriétés spectrales des trimères d'APC. Le  $L_{CM}$  absorbe en effet dans le rouge lointain, entre 668 et 676 nm (Zilinskas et al., 1980; Lundell et al., 1981b), et permet ainsi de transmettre l'énergie émise par les trimères d'APC vers la Chl *a* du PSII. Il sert d'accepteur terminal de l'énergie, capturée par les centaines de chromophores du PBS (Gindt et al., 1994; Gao et al., 2012).

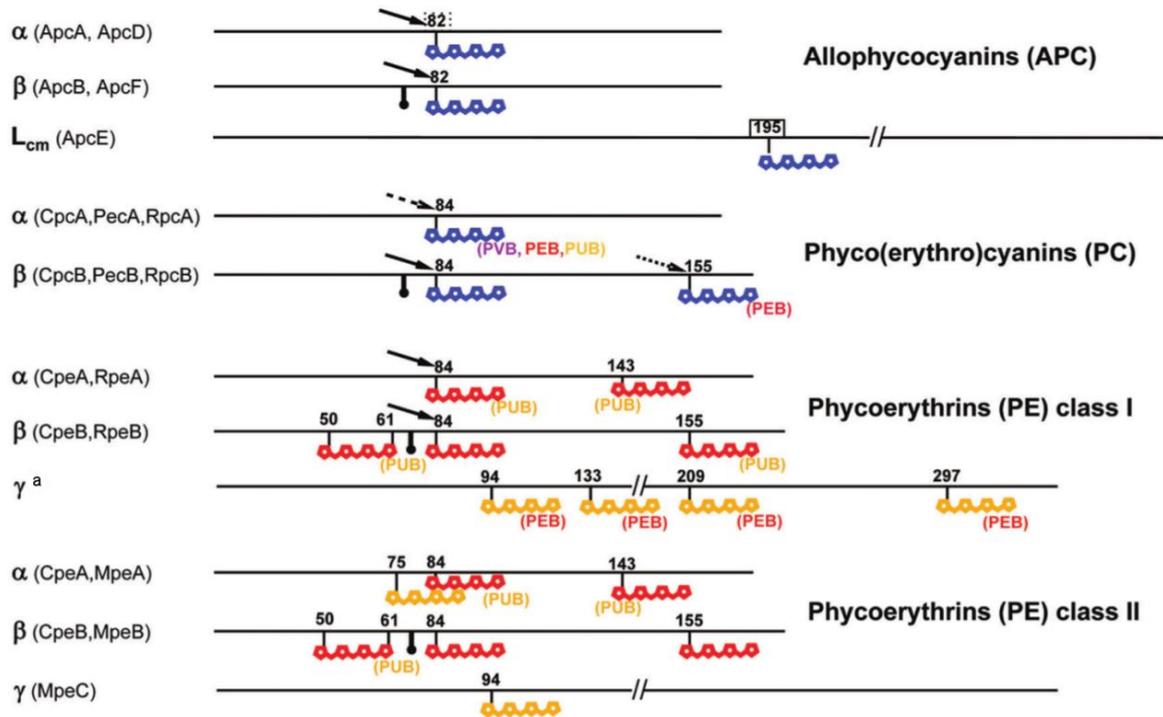


**Figure 27** : Représentation schématique des localisations probables des protéines de liaison au sein d'un phycobilisome de *Synechococcus* sp. WH8102, comportant de la phycoérythrine II. Les positions de CpeC et CpeE sont incertaines. Les étoiles indiquent les positions possibles des chromophores de phycourobiline. Modifié d'après Six et al. (2005).

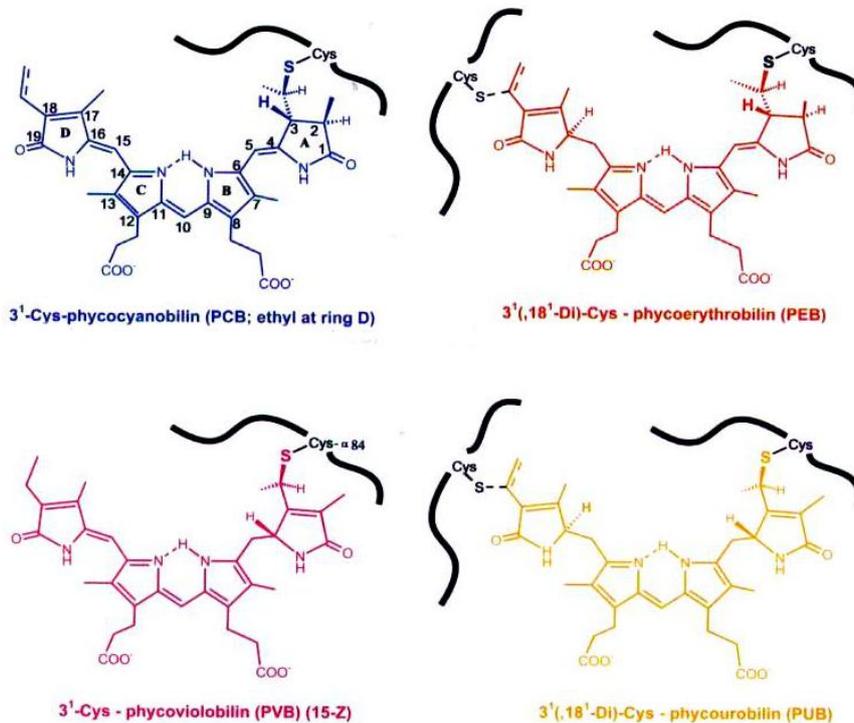
Les algues rouges possèdent également un autre type de linker, associé à la PE, appelé sous-unité  $\gamma$ , pouvant être au nombre de 2 (Glazer et Hixson, 1977; Klotz et Glazer, 1985; Bernard et al., 1996). Ses caractéristiques biochimiques ( $pI > 8$ ) sont similaires aux linkers classiques, et la partie N-terminale occuperait également le canal central de la PE conduisant à la formation d'un complexe stable  $(\alpha\beta)_6\gamma$  (Wilbanks et Glazer, 1993a; Ritter et al., 1999). Cependant, ces sous-unités, codées au niveau du noyau, ont un faible degré de similarité avec les linkers classiques codés par le génome plastidial (Egelhoff et Grossman, 1983; Apt et al., 1993). Grâce aux chromophores qu'elles lient, ces sous-unités participent à la capture de la lumière et au transfert énergétique. Chez les *Synechococcus* marins, plusieurs polypeptides de type  $\gamma$ , MpeC et MpeE, dont le chromophore est la PUB ont également été mis en évidence au niveau de la PEII (Wilbanks et Glazer, 1993a; Six et al., 2005). MpeC assurerait la stabilité du disque distal des bras des PBS, et disparaîtrait à forte lumière afin de réduire la taille des PBS (Six et al., 2005). Un autre linker, deux fois plus long que les précédents, MpeD, s'est avéré être la fusion entre un linker non chromophorylé lié à la PEI et un linker chromophorylé lié à la PEII et il assurerait donc la liaison entre ces deux PBPs dans les bras (Fig. 27).

#### II.2.5.2. *Phycobiliprotéines et chromophorylation*

Les différentes PBPs se distinguent par leurs propriétés spectroscopiques. Malgré leur homologie au plan structural, les PBPs peuvent lier un nombre différent de chromophores, allant de un à quatre, le nombre de sites de fixation allant croissant depuis l'APC vers la PE (Fig. 28; Sidler, 1994). Les phycobilines sont des groupements tetrapyrroles linéaires, maintenus en conformation ouverte *in vivo*, alors que leur conformation native est cyclique. Ces isomères diffèrent par leur nombre de double liaisons. On distingue ainsi la phycocyanobiline (PCB,  $A_{\max} \sim 620$  nm), la phycoviolobiline (PVB,  $A_{\max} \sim 570$ nm), la phycoerythrobiline (PEB,  $A_{\max} \sim 550$  nm) et la phycourobiline (PUB,  $A_{\max} \sim 495$  nm; Fig. 29). Elles sont toujours fixées aux PBPs au niveau de résidus cystéine (Cys) par une liaison thioéther, souvent en C-3, bien que des double liaisons en C-3 et C-18 existent chez certaines PE, entre les Cys- $\beta$ 50 et Cys- $\beta$ 61 (Fig. 29; Bishop et al., 1987; Ficner et Huber, 1993). En plus de la liaison covalente thioéther, d'autres interactions faibles existent entre l'apoprotéine et le chromophore et sont essentielles pour le fonctionnement du PBS. Les phycobilines libres absorbent très peu l'énergie lumineuse, et leurs états d'excitation ne provoquent pas de rayonnement mais est rapidement perdue sous forme de chaleur (Falk, 1989). A l'état libre les phycobilines adoptent une conformation flexible hélicoïdale, correspondant à l'état le plus stable thermodynamiquement (Falk, 1989). Lorsqu'elles sont liées aux PBPs, une conformation linéaire et rigide est imposée aux phycobilines, permettant d'accroître la durée de vie de l'état excité (Ficner et Huber, 1993; Ritter et al., 1999; Doust et al., 2004; Wagner et al., 2005; Schmidt et al., 2007).



**Figure 28** : Chromophorylation des différentes phycobiliprotéines. Les cystéines où sont fixés les chromophores sont indiqués par leur position consensus. Les chromophores sont indiqués de différentes couleurs (phycocyanobiline (PCB) en bleu, phycoviolobiline (PVB) en violet, phycoérythrobilin (PEB) en rouge, et phycourobilin (PUB) en orange). Lorsque d'autres chromophores peuvent être fixés, ceux-ci sont mentionnés entre parenthèses. Les flèches indiquent les sites pour lesquels les lyases ont été identifiées. <sup>a</sup> Uniquement chez les algues rouges. D'après Scheer et Zhao (2008).



**Figure 29** : Liaison des différentes phycobilines aux résidus cystéines des phycobiliprotéines par des liaisons thioéthers. Les couleurs utilisées se rapprochent de la couleur des différents pigments. D'après Roy et al. (2011).

Chez toutes les PBP, sauf le domaine PBP du  $L_{CM}$ , un site de fixation de chromophore est systématiquement présent en position 84 (position consensus, le site variant de la position 81 à 84) des deux sous-unités. Des sites de fixation additionnels sont présents autour de la position 150 au niveau de l'extrémité C-terminale, ou autour de la position 50 en N-terminal. La Cys- $\beta$ 84 est l'accepteur final d'énergie au sein de chaque trimère et est toujours chromophorylée avec de la PCB chez l'APC et la PC et de la PEB chez la PE.

En fonction de la chromophorylation de chacune de ces sous-unités, les PBP présentent des propriétés spectrales nettement distinctes. Le nombre de chromophores distincts varie généralement de 1 à 3 (Tableau 2). Différents types de PBP ont ainsi été définis. Initialement cette classification prenait en compte l'origine des organismes dont les PBP étaient isolées en utilisant le préfixe C pour une origine cyanobactérienne et R pour les PBP d'algues rouges. Ces lettres sont toujours utilisés, bien qu'elles fassent à présent référence aux propriétés spectrales et non à l'origine taxonomique. La lettre C fait alors référence aux PC dépourvues de PEB ou PUB. A ce jour, sept types de PC ont ainsi été décrits (Tableau 2). La PEC correspond à une PC portant une PVB, et n'est présente que chez les cyanobactéries d'eau douce. La R-PCI est spécifique des algues rouges et la R-PCIV de la cyanobactérie marine *Crocospaera watsonii* (Swanson et al., 1991). Chez les *Synechococcus* marins, 4 classes de PC ont été répertoriées : la C-PC ne liant que de la PCB (retrouvée chez les souches vertes de type 1), la R-PCII liant à la fois de la PCB et de la PEB avec un rapport 1:2 (*Synechococcus* sp. WH7803), la R-PCIII liant ces mêmes chromophores mais avec un rapport de 2:1 (*Synechococcus* sp. WH7805) et enfin la R-PCV liant 3 phycobiline différentes : PCB, PEB et PUB (*Synechococcus* sp. WH8102; Ong et Glazer, 1987; Six et al., 2007b; Blot et al., 2009).

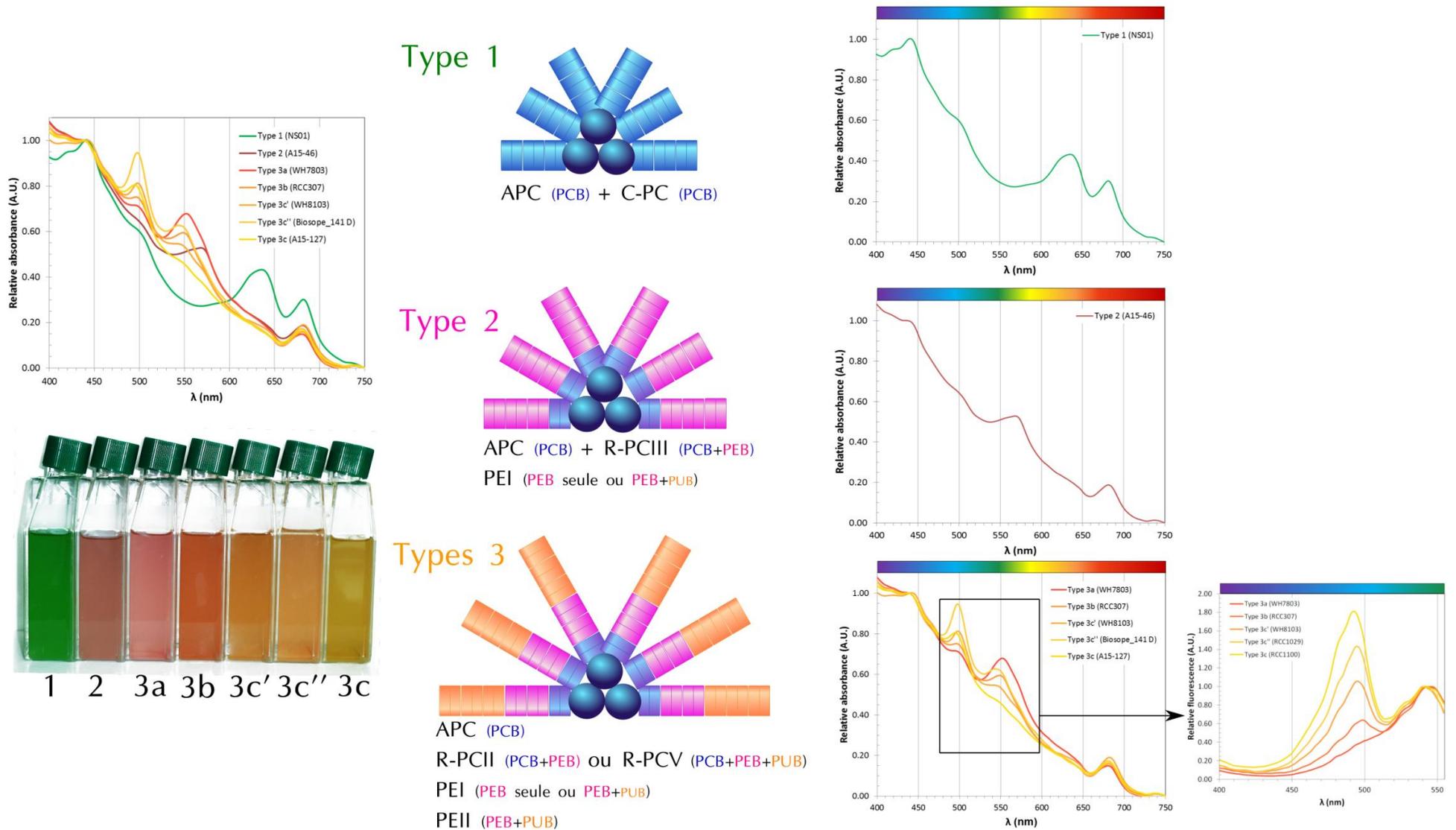
**Tableau 2: Classification des différents types de phycocyanine selon leur chromophorylation.** Nomenclature d'après Ong et Glazer (1987, 1988); Swanson et al. (1991); Sidler (1994); Blot et al. (2009).

	C-PC	PEC	R-PCI	R-PCII	R-PCIII	R-PCIV	R-PCV
$\alpha$ 84	PCB	PVB	PCB	PEB	PCB	PUB	PUB
$\beta$ 84	PCB	PCB	PCB	PCB	PCB	PCB	PCB
$\beta$ 155	PCB	PCB	PEB	PEB	PEB	PCB	PEB

Les PE possèdent plus de sites de fixation que les PC, mais ont toujours de la PEB en Cys- $\beta$ 84. Chez les algues rouges, deux types de PE, nommés B-PE ou R-PE, portent une majorité de PEB, induisant la couleur rougeâtre de ces organismes. Contrairement à la B-PE, de la PUB en Cys- $\alpha$ 84 est fixée à la R-PE (PUB:PEB 1:4; Glazer et al., 1982). De nombreuses cyanobactéries d'eau douce possèdent de la C-PE, qui comme la B-PE ne comporte que de la PEB. En revanche, une plus grande diversité a été mise en évidence au sein des cyanobactéries marines. *Crocospaera watsonii* est ainsi particulièrement adaptée à la capture de la lumière bleue et sa PE comporte 4 PUB pour 1 PEB

(Swanson et al., 1991). Au sein des *Synechococcus* marins, on distingue 2 types de PE : la PEI et la PEII (Alberte et al., 1984; Six et al., 2007b). La PEI lie soit uniquement de la PEB, soit de la PEB et de la PUB, la PEB restant majoritaire. La PEII lie toujours les deux chromophores mais en proportion variable.

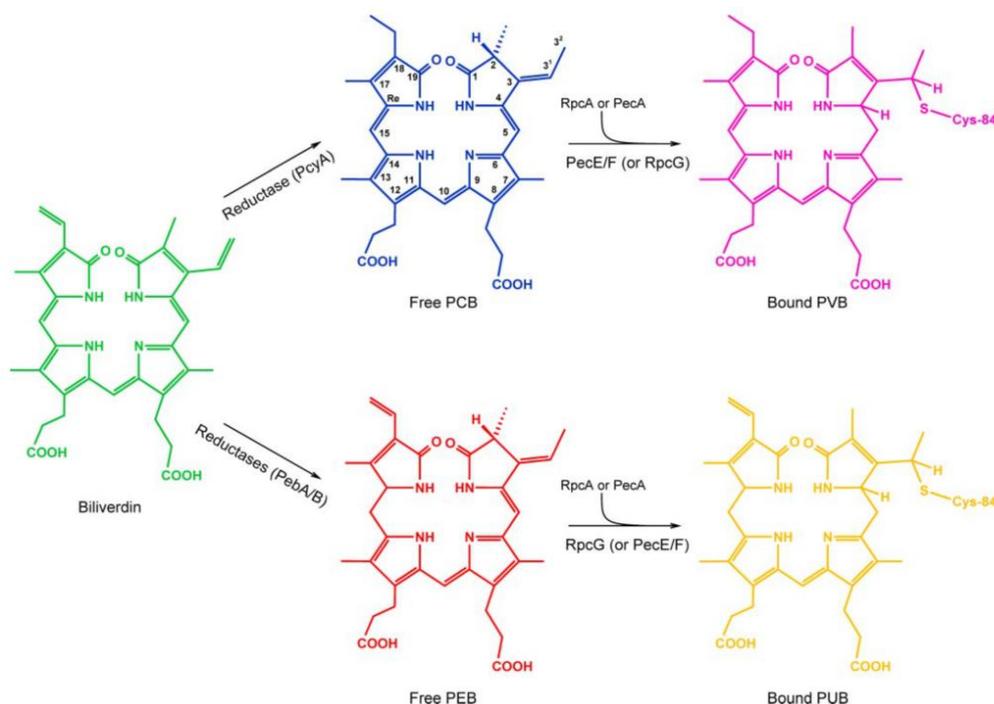
Les analyses biochimiques et génétiques menées à ce jour sur les PBS de *Synechococcus* ont permis de mettre en évidence six types et sous-types pigmentaires (Fig. 30; Six et al., 2007b). On peut distinguer les souches bleues-vertes, correspondant au type 1, dont les bras des PBS sont uniquement composés de C-PC. Les souches rose fuschia correspondent au type 2 et possèdent de la R-PCIII à proximité du cœur alors que les extrémités distales sont composées de PEI ne fixant que de la PEB. Le type 3 est représenté par les souches rouge-orange contenant de la R-PCII ou R-PCV et les deux sortes de PE (PEI et PEII). Ce dernier groupe peut être divisé en quatre sous-types en se basant sur le rapport d'excitation de fluorescence  $Ex_{495:545nm}$  utilisé pour évaluer le rapport PUB:PEB. Ce rapport peut-être faible ( $Ex_{495:545nm} \sim 0.4$ , 3a), moyen ( $Ex_{495:545nm} \sim 0.8$ , 3b), élevé ( $Ex_{495:545nm} \sim 1.7$ , 3c) ou variable (3d). Les souches du groupe 3d présentent la pigmentation la plus sophistiquée des *Synechococcus* marins et sont capables de changer leur pigmentation en fonction de la qualité de lumière, grâce au phénomène d'acclimatation de type IV (voir §II.3.4; Palenik, 2001; Six et al., 2007b).



**Figure 30:** Types pigmentaires rencontrés au sein des *Synechococcus* marins et spectres d'absorption des cellules en culture. Les souches capables d'acclimatation chromatique sont capables de modifier la chromophorylation de leurs phycobilisomes (F. Humily, non publié).

## II.2.5.3. Biosynthèse et attachement des phycobilines

Les voies métaboliques de biosynthèse des phycobilines et des Chl sont communes jusqu'à l'étape de la protoporphyrine IX où elles divergent par une métallation différentielle, par du fer (phycobilines) ou du magnésium (Chl; Beale, 1993; Frankenberg et Lagarias, 2006). Les phycobilines sont des dérivés de l'hème, qui est converti en biliverdine IX $\alpha$  (BV $\alpha$ ). Ce composé est ensuite réduit par des enzymes de la famille des FDBRs (ferredoxin-dépendent bilin reductase), famille catalysant la même réaction que les hème oxygenases des mammifères, et également retrouvé chez les plantes pour la biosynthèse des phytochromes (Cornejo et Beale, 1997; Cornejo et al., 1998; Davis et al., 1999; Muramoto et al., 1999). Ces derniers se caractérisent par une régiospécificité de doubles liaisons au niveau de ces composés cycliques, induisant ainsi des propriétés spectroscopiques distinctes.



**Figure 31 : Biosynthèse des phycobilines à partir de biliverdine IX $\alpha$ .** La biosynthèse de phycourobiline (PUB) à partir de phycoréthrobiline (PEB), et de phycoviolobiline (PVB) à partir de phycocyanine (PCB) résulte d'une isomérisation qui ne peut avoir lieu qu'après attachement sur les phycobiliprotéines. D'après Blot et al. (2009).

A partir de la BV, trois voies majeures peuvent être distinguées. La synthèse du chromophore des phytochromes des plantes supérieures, la phytochromobiline (P $\Phi$ B), est catalysée par la phytochromobiline synthase (HY2; Terry et al., 1995; Frankenberg et al., 2001; Kohchi et al., 2001). La PCB quant à elle est issue de deux étapes de réduction au niveau du cycle A, menant à la formation d'un composé intermédiaire, assurées par l'enzyme PcyA (phycocyanobiline:ferredoxine oxidoreductase; Frankenberg et al., 2001). PcyA est l'enzyme la mieux caractérisée biochimiquement

et structurellement de la famille FDBR (Frankenberg et Lagarias, 2003; Tu et al., 2004; Hagiwara et al., 2006). La biosynthèse de PEB requière également deux étapes de réductions, au niveau du cycle D en C15, mais contrairement à la PCB, celles-ci sont catalysées par deux enzymes distinctes : PebA et PebB (Frankenberg et al., 2001; Dammeyer et Frankenberg-Dinkel, 2006). Récemment, un homologue de PebA appelé PebS (pour PEB synthase) a été mis en évidence dans le génome du myovirus P-SSM2 infectant *Prochlorococcus* (Sullivan et al., 2005). Cette enzyme combine les activités de PebA et PebB en étant capable de réduire seule de la BV en PEB (Dammeyer et al., 2008; Busch et al., 2011). Cette enzyme est spécifique aux génomes de cyanophages et permettraient d'accroître leurs capacités d'infection (Dammeyer et al., 2008). La PCB et la PEB sont les seules phycobilines à avoir été retrouvées libres au sein du cytoplasme. La biosynthèse de la PVB et de la PUB ne peut avoir lieu que par isomérisation après fixation préalable d'un chromophore (PCB et PEB, respectivement). Si chez les *Synechococcus* marins, l'isomérisation de la PEB en PUB peut-être effectuée par différentes enzymes selon la PBP et le site sur laquelle elle a lieu (Blot et al., 2009; Shukla et al., 2012), c'est l'hétéroduplexe PecE/F qui assure le passage de la PCB à la PVB sur la PEC chez les cyanobactéries d'eau douce (Jung et al., 1995; Zhao et al., 2000; Storf et al., 2001).

Les phycobilines PCB et PEB peuvent s'attacher de manière spontanée *in vitro* sur les PBPs au niveau des résidus Cys natifs (Arciero et al., 1988; Fairchild et Glazer, 1994). Néanmoins, ces réactions produisent une mauvaise stéréochimie au niveau des carbones asymétriques et les PBPs ainsi chromophorylées n'existent pas naturellement (Scheer et Zhao, 2008). Certains attachements spontanés conduisent cependant à la formation de PBPs natives. Bien que principalement observé au niveau des phytochromes, ces attachements dits "autocatalytiques" se produisent aussi au niveau du L<sub>CM</sub> (Zhao et al., 2005b). L'attachement des chromophores serait principalement sous le contrôle d'une famille d'enzymes spécialisées, appelées phycobilines lyases. Cependant le rôle de ces enzymes serait davantage une participation au contrôle de l'attachement, notamment au niveau régio- et stéréospécificité, et elles auraient donc un rôle de chaperone (Schluchter et Glazer, 1999).

Les phycobilines lyases constituent un groupe particulièrement diversifié (voir revues de Scheer et Zhao, 2008; Schluchter et al., 2010). Par convention, les enzymes Cpc catalysent la chromophorylation de la PC, tandis que les lyases Cpe interviennent sur la PEI (Cobley et al., 2002) et Mpe sur la PEII (Bretaudeau et al., 2012). Actuellement, trois clans (protéines partageant une structure tridimensionnelle commune) et trente familles (groupes d'orthologues) ont été décrits et sont répertoriés dans la base de données Cyanolyase (<http://cyanolyase.genouest.org/>), décrite dans un article auquel j'ai participé (voir Bretaudeau et al., 2012; Annexe B).

Parmi ces trois clans, le clan E/F constitue le groupe le plus diversifié et rassemble la majorité des phycobilines lyases décrites à ce jour. Les lyases sont soit mono- soit hétérodimériques. Le

premier hétérodimère à avoir été caractérisé biochimiquement est CpcE/CpcF, une enzyme spécifique de la Cys-84 de l' $\alpha$ -PC et qui peut aussi bien catalyser l'attachement de la PCB que son détachement (Fairchild et al., 1992; Zhou et al., 1992). Chez la plupart des cyanobactéries, ces deux gènes (ou leurs homologues *rpcE/F* dont le produit fixe une PEB au même site), sont retrouvés au sein d'un opéron en aval de l'opéron *cpcBA* (ou *rpcBA*), codant pour les deux sous-unités de la PC (de Lorimier et al., 1993; Wilbanks et Glazer, 1993b; Kaneko et al., 1996; Six et al., 2007b). Deux autres hétéroduplexes, PecE/PecF qui est spécifique des cyanobactéries d'eau douce possédant de la PEC et RpcG qui est présent chez certaines cyanobactéries marines, ont également été caractérisés au sein de cette famille. RpcG est une protéine de fusion dont le N-terminal présente une forte homologie avec PecE et le C-terminal avec PecF (Blot et al., 2009). Ces deux hétéroduplexes sont bi-fonctionnels puisqu'ils assurent à la fois l'attachement de PCB et PEB au niveau de Cys-84 de l' $\alpha$ -PC et leur isomérisation en PVB et PUB, respectivement (Jung et al., 1995; Zhao et al., 2000; Zhao et al., 2002; Blot et al., 2009). La chromophorylation par de la PEB de la Cys-82 de l' $\alpha$ -PE est assuré par le complexe CpeY/Z, où CpeY est la lyase et CpeZ un "facilitateur" (Biswas et al., 2011). Toutes les protéines du clan E/F décrites à ce jour contiennent des domaines répétés HEAT, variant de 2 à 6, impliqués dans les interactions protéines-protéines (Andrade et al., 2001; Takano et Gusella, 2002).

Le second clan de phycobilines lyases est le clan S/U, dont les principaux représentants sont les homologues CpcS et CpeS. CpcS est active à l'état d'homodimère chez certaines souches (Zhao et al., 2006; Zhao et al., 2007a; Wiethaus et al., 2010) et d'hétérodimère en association avec CpcU chez d'autres souches (Saunée et al., 2008; Shen et al., 2008) et il en est probablement de même pour CpeS, homodimérique chez *Prochlorococcus* (Wiethaus et al., 2010) et sans doute hétérodimérique avec CpeU chez les *Synechococcus* marins (F. Partensky, comm. pers). Les lyases de type S/U présentent une grande spécificité pour la Cys-82 (position consensus), mais peuvent fonctionner avec une plus grande gamme de substrats (chromophore et apoPBP) que les lyases E/F (Scheer et Zhao, 2008; Kupka et al., 2009). En effet, l'enzyme CpcS de *Nostoc* sp. PCC7120 est (au moins *in vitro*) une lyase quasi-universelle, capable de chromophoryler avec de la PCB ou de la PEB les Cys-82 de toutes les sous-unités de PBP, à l'exception de l' $\alpha$ -PC (chromophorylée par CpcE/F; Zhao et al., 2007a). Cependant, *in vivo*, c'est CpeS qui catalyse la chromophorylation de la Cys-82 de la  $\beta$ -PE.

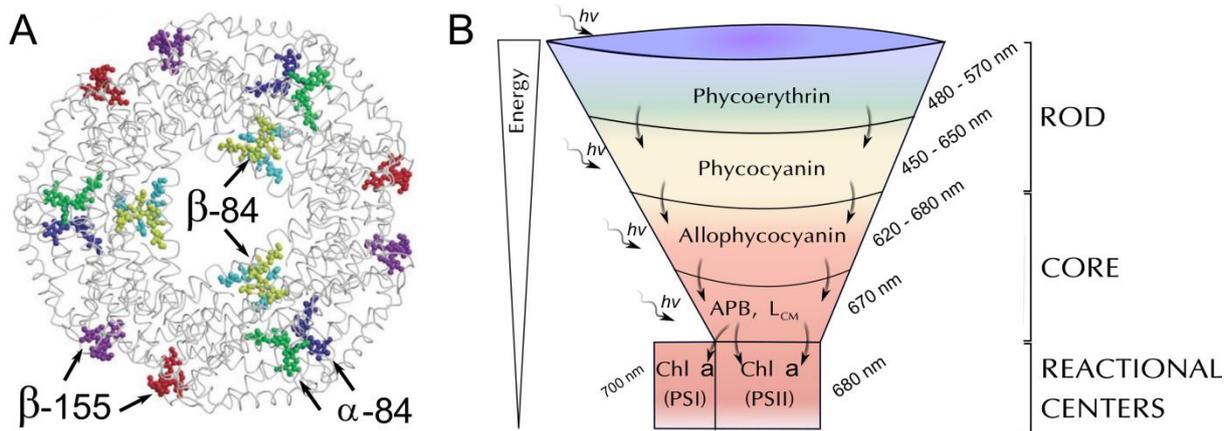
Le clan T constitue le troisième clan des phycobilines lyases. Contrairement aux deux clans précédents, ses représentants ne semblent pas former d'hétérocomplexe. CpcT est capable de fixer de la PCB en Cys-153 de la  $\beta$ -PC (Shen et al., 2006; Zhao et al., 2007b). CpeT, un homologue de CpcT (Cobley et al., 2002), dont la fonction supposée serait de fixer une PEB en Cys-153 de la  $\beta$ -PE, est la seule lyase à avoir été mise en évidence au sein de myovirus infectant *Synechococcus* (Dammeyer

et al., 2008), dont la phase d'infection est accompagnée d'une augmentation de la PEI et PEII chez la souche WH7803 (Shan et al., 2008).

Le nombre de lyases d'un organisme dépend directement de la complexité de sa pigmentation et, par extension, la connaissance du contenu en lyases d'une souche entièrement séquencée peut théoriquement permettre de prédire sa pigmentation, comme cela est proposé dans la base de données Cyanolyase (Bretaudeau et al., 2012). Ainsi, trois lyases suffisent pour la chromophorylation des PBP des cyanobactéries du type pigmentaire 1 (souches à PC) : CpcS (+/- CpcU), CpcT et CpcE/F. Au moins quatre autres phycobilines lyases sont nécessaires pour chromophoryler la PEI (CpeF, CpeS+/-CpeU, CpeT et CpeY/Z). Plusieurs de ces lyases sont probablement capables de chromophoryler aussi la PEII, mais d'autres sont très probablement nécessaires (notamment pour les sites fixant de la PUB) et à ce jour seule une lyase spécifique de la PEII a été caractérisée biochimiquement: la lyase-isomérase MpeZ, qui est spécifique des adaptateurs chromatiques de type IV (voir § II.3.4 et Shukla et al., 2012).

#### II.2.5.4. *Transfert orienté d'énergie*

Le PBS est un système particulièrement complexe et original par rapport aux antennes intrinsèques des plantes supérieures. La particularité du PBS réside dans la forte distance entre les complexes collecteurs et les centres réactionnels. Les contraintes structurales mises en évidence au sein du PBS témoignent d'optimisations dans le but d'accroître l'efficacité du transfert d'énergie. Les niveaux d'énergie ne sont pas équivalents entre les différents niveaux du PBS, et entre les différents chromophores d'une même PBP (Glazer, 1989a). Au sein du PBS l'énergie lumineuse migre toujours des extrémités des bras vers le cœur d'APC avant d'être délivrée à la Chl  $\alpha$  au niveau des membranes thylacoïdales (Fig. 32B). Au fur et à mesure que l'on s'approche du cœur, les photons capturés sont de plus faible énergie. Ce transfert est extrêmement rapide et ne nécessite que de 50 à 100 ps (Holzwarth, 1986; Wendler et al., 1986; Gillbro et al., 1993; Yamazaki et al., 1994). Ces temps de transfert sont ainsi plusieurs ordres de magnitude plus rapides que les temps de vie moyen de la fluorescence d'une PBP. En se basant sur ces deux variables, Glazer a estimé que l'efficacité d'un tel système est d'environ 99% (Glazer, 1989a).



**Figure 32 : Transfert orienté d'énergie au sein du phycobilisome (PBS).** (A) Structure d'un hexamère de C-phycocyanine de *Synechococcus elongatus*. Les chromophores du même trimère sont indiqués dans la même couleur. D'après Nield et al. (2003). (B) Illustration du transfert orienté d'énergie au sein du PBS, de l'extrémité des bras vers les photosystèmes. Modifié d'après Sidler (1994); Roy et al. (2011).

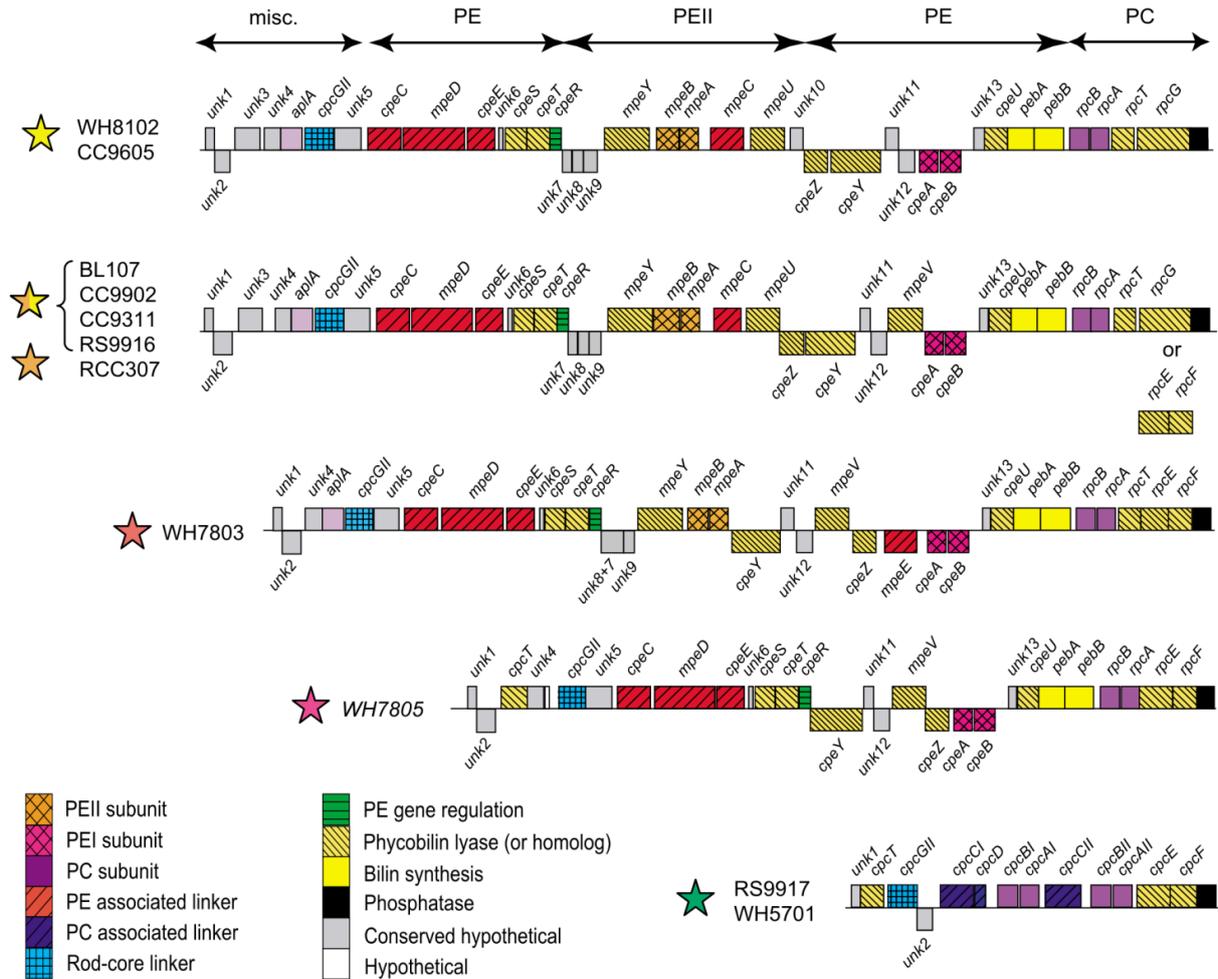
Le transfert d'énergie au sein du PBS est non radiatif et résulte d'une interaction dipôle-dipôle entre deux molécules. Les études spectroscopiques ont ainsi permis de classer les chromophores en deux types : le type *s* (sensible) et le type *f* (fluorescent; Teale et Dale, 1970). Les chromophores *s* (donneurs) absorbent l'énergie puis la transfèrent aux chromophores *f* qui fluorescent (accepteurs). Le couplage de ces études spectroscopiques avec les données structurales a permis d'identifier ces différents chromophores au sein des PBPs. Une caractéristique importante des trimères, commune à toutes les PBPs, est la présence d'un canal central. Parmi les différentes cystéines chromophorylées, les Cys- $\alpha$ 84 et Cys- $\beta$ 155 se retrouvent en périphérie alors que la Cys- $\beta$ 84 est systématiquement en relation avec le canal et correspond à l'accepteur (Fig. 32A; Schirmet et al., 1986). Compte tenu des distances entre les résidus Cys- $\alpha$ 84 et Cys- $\beta$ 155, le transfert énergétique se produit principalement de ces deux chromophores vers Cys- $\beta$ 84, augmentant ainsi l'efficacité du système (Mimuro et al., 1986; MacColl, 1998). La position et la fonction de la Cys- $\beta$ 84 est conservée au sein des différentes PBPs. Bien qu'elle suive les mêmes lois physiques, la situation semble être plus complexe au niveau de la PE en raison du plus grand nombre de chromophores ainsi que des distances plus courtes relevées entre ces derniers (Ficner et al., 1992; Ritter et al., 1999). La détermination de la position des chromophores au sein des trimères de PE a conduit à proposer les voies de transfert énergétiques inhérentes à cette protéine (Jiang et al., 1999). Ces processus restent à découvrir au niveau de la PEII des *Synechococcus* marins, qui possède un chromophore de plus que les PEI, et pour laquelle aucune structure n'est disponible à ce jour. Un schéma de transfert énergétique entre les différentes PBPs du PBS a été proposé par Jiang (Jiang et al., 2001).

## II.2.6 Gènes codant pour les polypeptides des phycobilisomes

### II.2.6.1. Nature et organisation génique

La connaissance des gènes codant pour les différentes unités du PBS a permis de faire des avancées considérables dans la compréhension de la structure des PBS, du fonctionnement du PBS ainsi que des processus évolutifs ayant agi sur ces gènes (de Lorimier et al., 1984; Conley et al., 1985; Conley et al., 1986; Houmard et al., 1986; Houmard et al., 1988; Wilbanks et Glazer, 1993a). De ces différents travaux, complétés depuis par le séquençage de régions génomiques et de génomes complets, a émergé une nomenclature relativement claire afin de désigner les gènes homologues d'organismes distincts impliqués dans la synthèse et régulation des PBPs. Ainsi, les gènes impliqués dans la synthèse de l'AP, la C-PC (ne fixant que de la PCB) et la R-PC (fixant différents chromophores) sont respectivement désignés *apc*, *cpc* et *rpc*. Chez les cyanobactéries d'eau douce, les gènes relatifs à la PEC sont nommés *pec*. Les gènes cyanobactériens de PE sont désignés *cpe*. Au sein de la radiation marine des *Synechococcus* où deux formes de PE sont rencontrées, les gènes de la PEI sont désignés *cpe*, alors que ceux impliqués dans la biosynthèse de la PEII sont désignés *mpe*. Ces gènes sont suivis de lettres désignant le polypeptide pour lequel ils codent. Les lettres A et B réfèrent aux sous-unités  $\alpha$  et  $\beta$  des PBPs. *cpeA* est ainsi le gène codant pour la sous-unité  $\alpha$  de la PEI chez les *Synechococcus* marins. Les autres lettres réfèrent dans la majorité des cas à des linkers ou des sous-unités de lyase de phycobilines.

Les gènes codant pour les deux-sous unités ( $\alpha$  et  $\beta$ ) des PBPs majeures (APC, PC, PE et PEII chez les *Synechococcus* marins) sont toujours contigus et cotranscrits (de Lorimier et al., 1984; Conley et al., 1985; Houmard et al., 1986; Conley et al., 1988; Mazel et al., 1988; Swanson et al., 1992; Grossman et al., 1993; Six et al., 2007b). De même, des gènes de linkers sont souvent adjacents et cotranscrits avec ceux codant pour les sous-unités de PBP de laquelle ils dépendent (Grossman et al., 1993). Cette organisation a d'ailleurs permis de déterminer la spécificité des différents polypeptides de liaison (Liu et al., 2005). Chez la plupart des cyanobactéries, les polypeptides intervenant dans l'assemblage du cœur, comme le  $L_C$  et le  $L_{CM}$  (encodés respectivement par *apcC* et *apcE*) sont ainsi retrouvés à proximité des gènes codant pour les sous-unités de l'APC, *apcE* étant retrouvé en amont, *apcC* en aval (Grossman et al., 1988; Houmard et al., 1988). Cependant chez certains organismes, comme *Synechococcus* sp. PCC 7002 et *Synechocystis* sp. PCC 6803, *apcE* est retrouvé isolé au sein du génome (Bryant, 1988; Grossman et al., 1993). Les deux autres sous-unités du cœur,  $\alpha^B$  et  $\beta^{18}$ , intervenant dans la formation de l'APC B et l'APC I respectivement, sont codées par les gènes *apcD* et *apcF* toujours retrouvés isolées au sein du génome. De manière similaire aux linkers, les gènes de certaines phycobilines lyases sont également retrouvées à proximité des gènes de PBPs qu'elles chromophorylent (Jung et al., 1995; Kahn et al., 1997; Schluchter et al., 2010).



**Figure 33 : Comparaisons des régions génomiques impliquées dans la synthèse et la régulation des bras des phycobilisomes au sein des différentes souches marines de *Synechococcus*.** Ces gènes sont localisés au sein d'îlots génomiques, et pourraient avoir été transférés latéralement entre les différentes lignées (Dufresne et al., 2008). Les souches vertes RS9917 et WH5701 constituent le type pigmentaire 1 (uniquement de la phycocyanine (PC)), WH7805 le type pigmentaire 2 (PC et phycoérythrine I (PE I)), les autres souches correspondant aux types 3a-d (PC + PE I + PE II). Les souches BL107, CC9902, CC9311 et RS9916 (type 3d) sont capables d'adaptation chromatique de type IV (d'après nomenclature proposée par Palenik (2001)). D'après Six et al. (2007b).

La majorité des gènes impliqués dans la structure et la biosynthèse des PBPs, des linkers et des phycobilines est regroupée au sein de régions plus au moins spécialisées. Cette organisation a initialement été décrite chez *Fremyella diplosiphon*, cyanobactérie filamenteuse d'eau douce, chez qui les gènes d'APC et de PC sont regroupés en un vaste cluster génomique (Conley et al., 1986; Grossman et al., 1993). L'organisation génomique retrouvée chez cette dernière est en réalité originale car les clusters impliqués dans la synthèse du cœur et des bras ne sont éloignés que de 4 Kpb, alors que ceux-ci sont généralement retrouvés à différents endroits du génome chez les autres cyanobactéries (Conley et al., 1988; Grossman et al., 1993). L'étude de son génome a également permis de mieux comprendre les bases moléculaires du processus d'acclimatation chromatique (voir §II.3.2). Trois opérons codant pour des sous-unités  $\alpha$  et  $\beta$  de PC ont été mis en évidence chez *F. diplosiphon*. Les deux premières séries de gènes *cpcB1A1* et *cpcB2A2* sont contigues, mais possèdent

des propriétés transcriptionnelles distinctes (Conley et al., 1986; Grossman et al., 1993). L'expression de *cpcB1A1* est constitutive, alors que l'expression de *cpcB2A2* est significativement plus forte en lumière rouge qu'en lumière verte (voir §II.3.2; Conley et al., 1986; Lomax et al., 1987; Conley et al., 1988; Houmard et al., 1988; Mazel et al., 1988). Le troisième opéron n'est induit qu'en cas de carence en soufre (Mazel et Marliere, 1989).

Cette organisation est également retrouvée au sein des *Synechococcus* marins pour lesquels plusieurs clusters ont été mis en évidence (Wilbanks et Glazer, 1993a; Six et al., 2005). Comme pour la majorité des autres cyanobactéries, un premier cluster comprend 4 gènes impliqués dans la synthèse d'APC, dans l'ordre *apcEABC*, alors que deux autres gènes, *apcD* et *apcF* sont retrouvés isolés au sein des différents génomes. D'autre part, la majorité des gènes impliqués dans la synthèse et régulation des projections radiaires sont localisés au sein d'une région génomique beaucoup plus large, dont la taille augmente avec le degré de complexité du PBS, d'environ 9 Kpb pour les souches du type 1 à 27-28.5 Kpb chez les souches capables d'acclimatation chromatique (3d; Fig. 33). La synténie de ces régions est très similaire entre les souches du même type pigmentaire. Dans la plupart des génomes, l'extrémité 5' de cette région commence par un gène court dont la fonction reste inconnue à ce jour (*unk1*) et se termine à l'extrémité 3' par un gène conservé codant probablement pour une phosphotyrosine phosphatase de faible poids moléculaire.

Chez les souches de type 1, l'essentiel de la région est principalement composée de deux opérons *cpcBA* codant pour les sous-unités de la PC-C, 3 linkers associés aux bras, 1 linker associé au cœur d'APC et par deux types de phycobilines lyases (CpcT et CpcE/F). Chez la seule souche de type 2 actuellement séquencée, WH7805, une partie de la région est remplacée par un groupe de 19 gènes, impliqués dans la synthèse et régulation de la PEI. Les gènes *pebA* et *pebB*, situés à l'extrémité 3', sont impliqués dans la synthèse des chromophores PEB. La région PE est aussi présente chez les souches contenant de la PEII, mais celle-ci est interrompue par une autre sous-région additionnelle (région PEII), contenant de 5 à 9 gènes et située entre le régulateur de la PE *cpeR* et la lyase isomérase *cpeY* (ou *cpeZ* chez d'autres souches). Les gènes codant pour les sous-unités  $\alpha$  et  $\beta$  de la PEII, deux phycobilines lyases, un polypeptide linker ainsi que deux ou trois protéines de fonction inconnue sont présents au sein de la région PEII. En amont de tous ces gènes, toutes les souches possédant de la PEII, contiennent le gène de linker *cpcGII* et *aplA*, dont le produit présente une homologie avec la sous- et appartiendrait à une nouvelle classe de photosenseurs cyanobactériens (Montgomery et al., 2004). Il est important de noter que plusieurs gènes codant pour des protéines de fonction inconnue (nommés *unk*) sont présents au sein de ce cluster génomique, et pourraient

avoir eu un rôle dans le transfert latéral de gènes en servant de points de recombinaison (Six et al., 2007b).

#### II.2.6.2. *Processus évolutifs des phycobiliprotéines*

Les PBPs sont des protéines retrouvées à la fois dans le domaine bactérien, au sein de l'embranchement cyanobactérien, mais aussi au sein des plastes d'organismes eucaryotes comme les rhodophytes et les cryptophytes. De la PC et des sous-unités  $\beta$  d'APC sont également présentes chez *Acaryochloris*, qui utilise de la Chl *d* comme pigment principal (Miyashita et al., 1996; Marquardt et al., 1997; Miyashita et al., 2003). De la PE a été reportée chez les écotypes LL de *Prochlorococcus*, même si son rôle reste encore à déterminer (Hess et al., 1996; Steglich et al., 2003a; Steglich et al., 2005). Cependant chez ces deux cyanobactéries, les PBPs ne sont pas organisées en PBS.

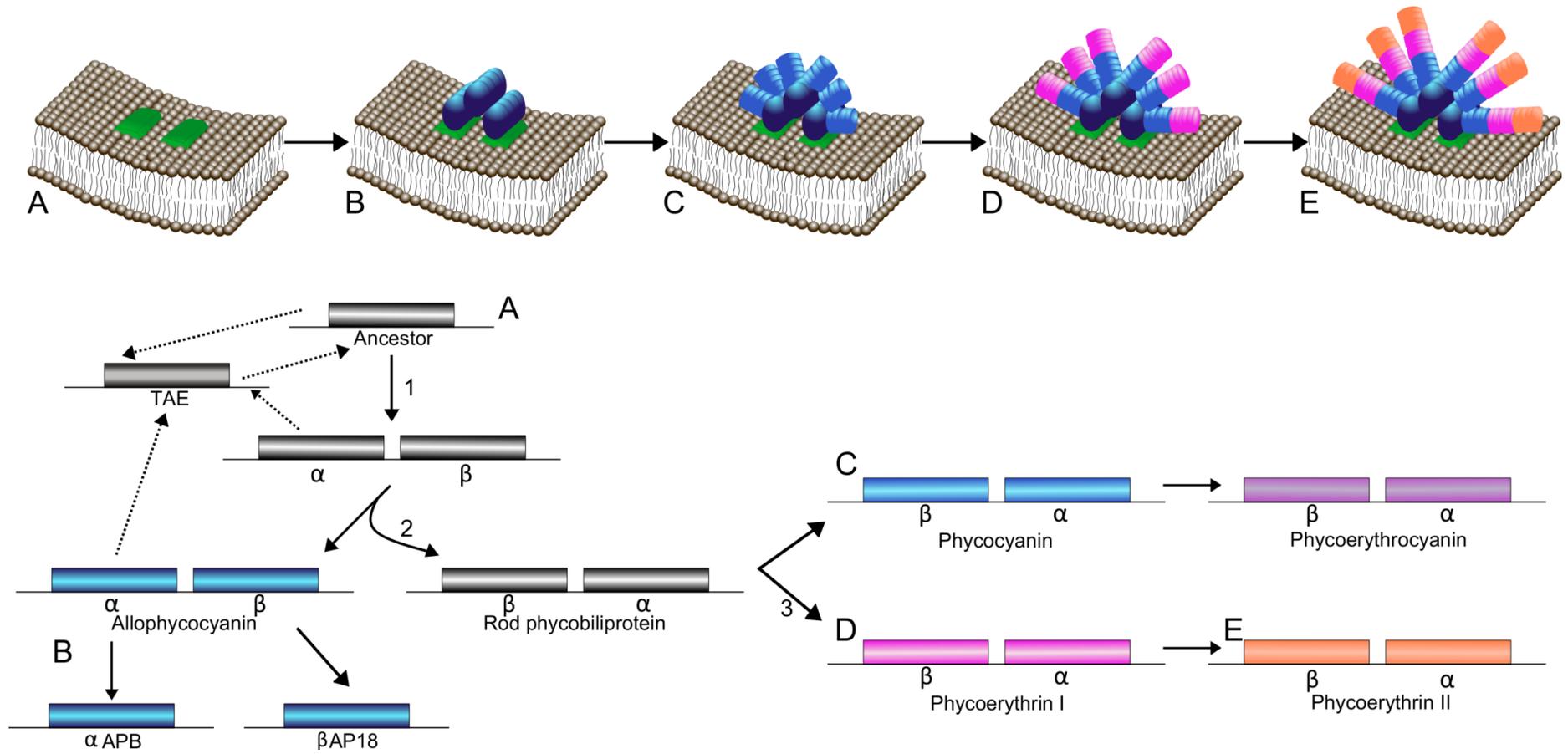
La similarité entre les PBPs cyanobactériennes et d'algues rouges a été observée pour la première fois en 1967 par des approches immunochimiques (Berns, 1967). De ces travaux ont émergé la théorie endosymbiotique d'un organisme proche des cyanobactéries actuelles avec un organisme non photosynthétique eucaryote, entraînant ainsi l'apparition du chloroplaste des rhodophytes. Un évènement endosymbiotique secondaire serait à l'origine du chloroplaste des cryptophytes (Ludwig et Gibbs, 1985; Palmer et Delwiche, 1996). En effet, de nombreuses études d'immunochimie et la comparaison des séquences de PBPs ont démontré que ces protéines étaient toutes liées (Guard-Friar et al., 1986; Sidler et al., 1990).

De plus, les PBPs dérivent clairement d'une protéine ancestrale commune. Les comparaisons de séquences de PBPs ont mis en évidence un grand nombre d'acides aminés conservés (Ducret et al., 1994; Apt et al., 1995; Zhao et Qin, 2006). Ces résidus qui sont retrouvés tout au long de la molécule sont néanmoins concentrés dans les domaines d'attachement des chromophores et dans les domaines X et Y (Zhao et Qin, 2006). Ces deux domaines participeraient à la stabilité des chromophores ou du monomère  $\alpha\beta$  (Schirmer et al., 1985; Schirmer et al., 1986; Kikuchi et al., 2000). Selon le modèle proposé par Apt en 1995 (Fig. 34), la protéine ancestrale devait contenir ces acides aminés, ainsi que ceux nécessaires pour la trimérisation et l'héxamérisation, une structure tridimensionnelle proche des PBPs actuelles, et un seul site d'attachement de chromophore, probablement la Cys-84 (Apt et al., 1995).

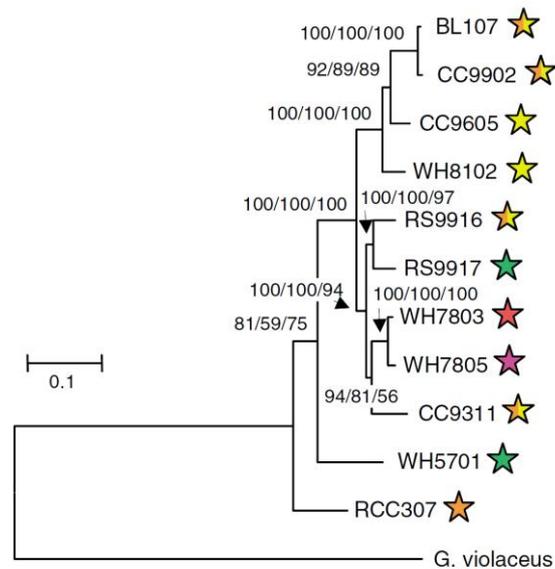
Le gène codant pour cette protéine aurait au cours d'une première étape été dupliqué, conduisant à la génération d'un tandem de gènes de PBP. Les gènes de ce tandem auraient divergé et coévolué pour former les deux sous-unités  $\alpha$  et  $\beta$  (Shi, 2011). Les sous-unités  $\alpha$  et  $\beta$  étant virtuellement toutes transcrites, cette co-transcription a dû participer à la conservation de la stoechiométrie 1:1 nécessaire à l'hétérodimérisation. La différenciation entre sous-unités  $\alpha$  et  $\beta$  a été une étape critique dans l'évolution des PBPs, ayant aidé à établir un complexe protéique avec un

transfert unidirectionnel d'énergie, des sous-unités  $\alpha$  vers  $\beta$ . Par la suite la duplication de ce tandem a donné lieu à deux lignées évolutives distinctes, celle du cœur du PBS (APC) et celle des bras (PC et PE; étape 2). Dans une de ces deux lignées, l'ordre des gènes a été inversé, probablement lors de la duplication.

Au sein de la lignée de l'APC, une nouvelle duplication génique a permis l'obtention des sous-unités  $\alpha^B$  et  $\beta^{18}$ . Le tandem précurseur des projections radiales a été dupliqué pour former respectivement la PC et la PE. Ces gènes ont co-évolué avec les gènes du cœur et des différentes protéines linkers, permettant d'obtenir la structure du PBS que l'on connaît actuellement. Le nombre de chromophores au sein de chaque unité a été augmenté, de même que leur nature, leur permettant de capturer de manière plus efficace la forte énergie des faibles longueurs d'onde. Ces différents processus évolutifs ont mené à l'obtention de caractéristiques spectrales distinctes au sein des différentes lignées de PBP.



**Figure 34:** Modèle d'évolution des gènes des phycobiliprotéines et de développement du phycobilisome. Adapté de Apt et al. (1995).



**Figure 35 :** Arbre en maximum de vraisemblance produit à partir des séquences d'acides aminés concaténées de 51 protéines ribosomales des *Synechococcus* marins. D'après Six et al. (2007b).

Les PBS actuels sont donc issus de deux forces évolutives majeures: i) le développement d'un transfert unidirectionnel de l'énergie lumineuse, des extrémités des bras radiaires vers le cœur d'APC, et ii) la capacité d'absorber les faibles longueurs d'onde. D'un point de vue évolutif, les PBS les plus récents sont donc ceux possédant à la fois de l'APC, de la PC et deux types de PE, comme ceci est le cas chez les cyanobactéries marines du genre *Synechococcus* du type pigmentaire 3. Il semblerait que les différents composants des PBS de *Synechococcus* aient évolué de manière presque indépendante. La phylogénie des gènes d'APC est similaire à celle des gènes de ménage (type ARNr ou protéines ribosomales), suggérant que les systèmes de transfert d'énergie lumineuse du cœur jusqu'au PS II est un mécanisme ancien et particulièrement conservé, ne permettant pas de variabilité phénotypique majeure au cours des processus évolutifs (Six et al., 2007b). En revanche, les composants des bras radiaires auraient évolué à travers des séries de duplication. L'acquisition de la PC et de la PE (ou PE I) a vraisemblablement eu lieu avant la différenciation de la branche *Synechococcus/Prochlorococcus/Cyanobium* des autres cyanobactéries (voir Fig. 7). Par contre, la différenciation de la PE II, un complexe spécifique des *Synechococcus* marins, serait postérieure à cette séparation (Everroad et Wood, 2012). Le fait que les types pigmentaires ne sont pas restreints à des clades spécifiques de *Synechococcus* (Fig. 35) suggère que les gènes impliqués dans la synthèse et régulation de la PE I et PE II, auraient été transférés d'une lignée à une autre, probablement par transfert latéral de gènes via des virus ou par transformation naturelle.

### **II.3. L'acclimatation chromatique chez les cyanobactéries**

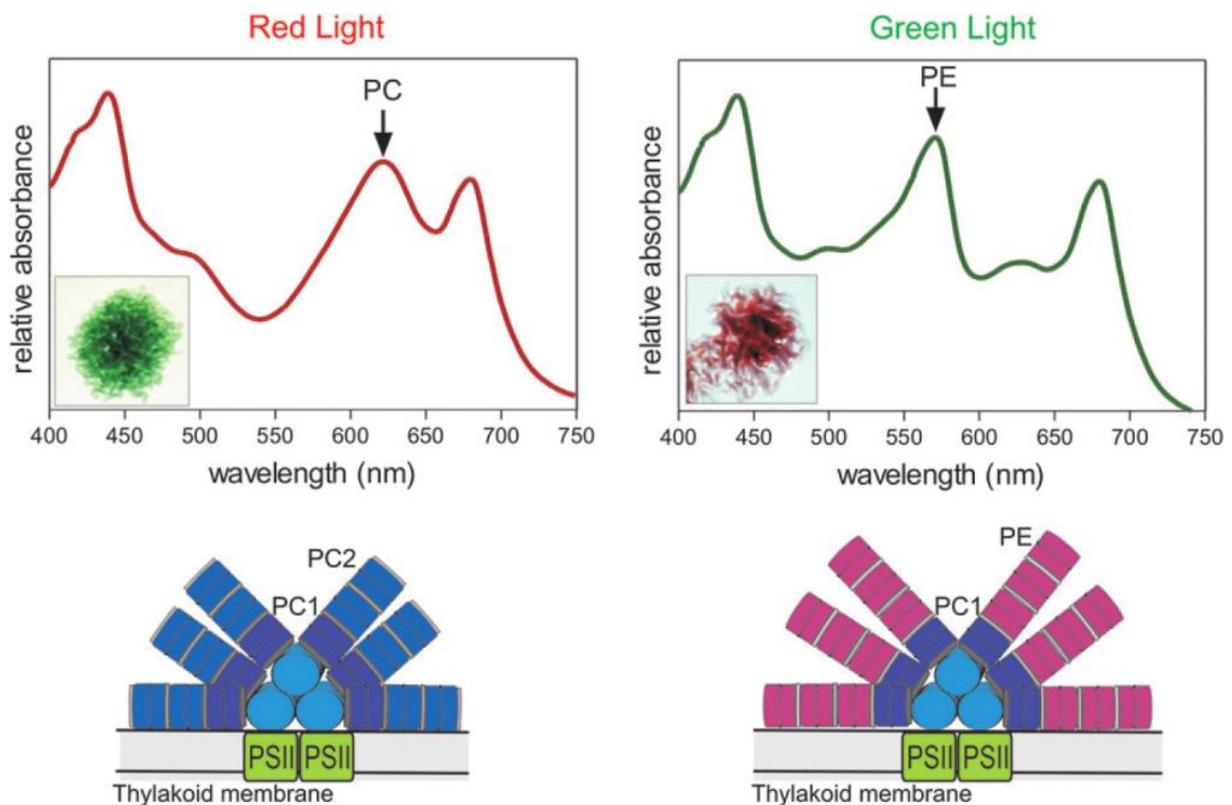
Dans l'environnement, l'appareil photosynthétique des cyanobactéries est confronté aux fluctuations quantitatives et qualitatives des conditions lumineuses (voir § II.1.2.1 et II.1.2.2). Un des premiers niveaux de réponse face à une augmentation de l'éclairement est la diminution de la quantité de complexes d'antenne et des pigments associés (Kana et Glibert, 1987). Sans cette modulation de la quantité de pigments, les processus photosynthétiques s'accompagneraient d'une forte production d'espèces réactives de l'oxygène (ROS) à forte intensité lumineuse. La production intracellulaire de ROS, générés à la fois au niveau des PSI et II et de l'antenne photocollectrice, est en effet particulièrement exacerbée lorsque le transport d'électrons photosynthétiques excède le taux de consommation de ces électrons pour la fixation du CO<sub>2</sub> (Vass et al., 1992; Asada, 1999; Zolla et Rinalducci, 2002). Les ROS réagissent avec de nombreux composés cellulaires et peuvent être toxiques pour la cellule. Une diminution de la capacité photosynthétique des cellules est donc observée à forte lumière (Kana et Glibert, 1987; Six et al., 2004) et/ou en réponse à l'exposition aux UV (Garczarek et al., 2008). En plus d'une réduction du nombre total de PBS, une réduction de taille des bras peut également être observée (de Lorimier et al., 1992; Six et al., 2007a). Un deuxième niveau de réponse aux changements d'intensité lumineuse implique la modulation du rapport entre le nombre de PSI et de PSII de manière à moduler le flux d'électrons en sortie du PSII et ainsi d'éviter les effets phototoxiques (Kawamura et al., 1979; Barlow et Alberte, 1985).

En plus de la variation d'intensité lumineuse, Engelmann notait dès la fin du XIX<sup>ème</sup> siècle que certaines cyanobactéries étaient capables de changer de couleur en fonction de la lumière à laquelle elles étaient exposées (Engelmann, 1883a, b). Les travaux de Gaidukov, son étudiant, permirent par la suite de mettre en évidence que les filaments de la cyanobactérie d'eau douce *Oscillaria* avaient cette capacité de changer leur couleur en fonction des filtres, et donc de la qualité spectrale de la lumière ambiante, sous lesquelles elles étaient cultivées (Gaidukov, 1902, 1903). Les cellules présentaient une coloration bleu-vert sous lumière orange, et rouge sous lumière verte. Les auteurs ont alors proposé le terme "d'adaptation chromatique complémentaire" (ACC). Ces observations furent contestées par de nombreux scientifiques, incapables de reproduire ces résultats en utilisant d'autres souches de cyanobactéries. Cette controverse fut résolue par Boresch, qui fut le premier à confirmer les résultats de Gaidukov, précisant même que ce phénomène impliquait un changement au niveau du rapport PE/PC (Boresch, 1922), modifiant ainsi la composition des PBS (Bennett et Bogorad, 1973; Tandeau de Marsac, 1977; Kehoe et Gutu, 2006). Bien que le terme adaptation ait été utilisé initialement, ce type de réponse n'implique aucun changement au niveau génétique, mais ferait appel à un changement au niveau des réseaux de régulation génique. Gutu et Kehoe ont proposé d'utiliser plutôt le terme « acclimatation chromatique » (Gutu et Kehoe, 2012).

### II.3.1 Diversité des souches capables d'acclimatation chromatique chez les cyanobactéries

Si toutes les souches de cyanobactéries ne sont pas capables d'AC, ce phénomène n'est pas circonscrit à une seule lignée et apparaît largement distribué au sein de ce phylum (Tandeau de Marsac, 1977). L'AC a été observée chez des souches d'eau douce mais également marines, et ne semble pas être restreinte à des niches écologiques ou des zones géographiques spécifiques (Perry et al., 1981; Tandeau de Marsac, 1983; Postius et al., 2001; Grossman, 2003; Six et al., 2007b; Acinas et al., 2008; Duxbury et al., 2009). A l'exception d'*Acaryochloris marina*, l'AC n'est retrouvée qu'au sein de cellules produisant de la PE (Boresch, 1922). La diversité des cyanobactéries capables d'AC continue de croître avec l'exploration de nouveaux environnements.

La connaissance sur les différentes couleurs de lumière pouvant être différemment perçues et les mécanismes impliqués ont énormément progressé ces dernières années. Initialement ce phénomène avait été décrit à partir de cellules capables de modifier leur contenu pigmentaire en réponse à des changements entre lumière rouge (RL) et verte (GL), mais des cellules capables de percevoir des variations entre bleu (BL) et GL, ou encore rouge et infrarouge ont récemment été décrites (Bennett et Bogorad, 1973; Palenik, 2001; Duxbury et al., 2009).



**Figure 36** : L'acclimatation chromatique de type III chez *Fremyella diplosiphon*. Images des différents phénotypes en fonction de la qualité de lumière et spectres d'absorption associés en lumière rouge et lumière verte. La chlorophylle *a* absorbe à 440 et 680 nm. En lumière rouge les bras des phycobilisomes sont courts et constitués uniquement de phycocyanine (PC), alors qu'ils sont plus longs d'un hexamère et constitués d'une majorité de phycoérythrine (PE) en lumière verte. D'après Gutu et Kehoe (2012).

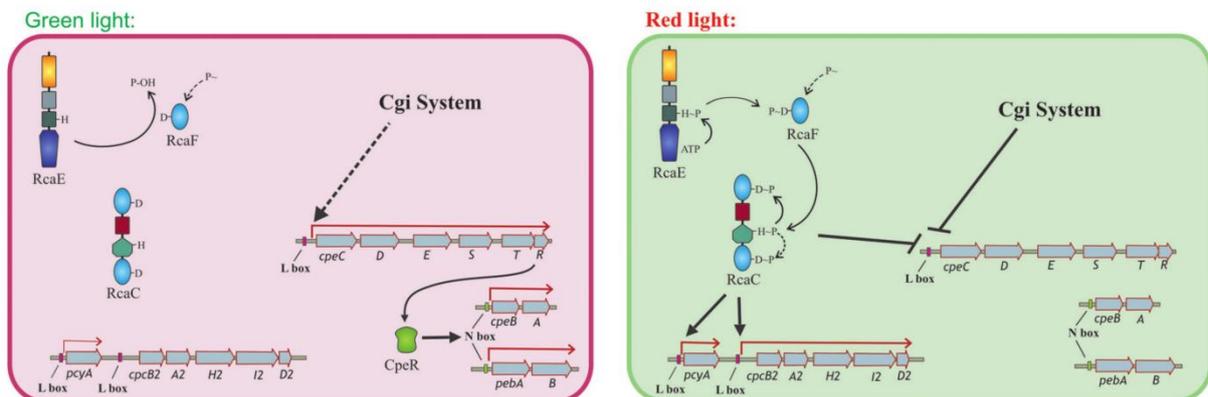
Les cyanobactéries possédant à la fois de la PC et de la PE, et donc potentiellement capables d'AC rouge-vert peuvent être réparties en différents types selon leur réponse (Tandeau de Marsac, 1977). Cette classification valable pour toutes les cyanobactéries distingue ainsi les souches n'étant pas capables de modifier leur composition pigmentaire (type I). D'autres souches ont des niveaux de PE plus importants en GL qu'en RL (type II). Les niveaux de PC ne sont quant à eux pas influencés par ce phénomène. Les souches du type III (CA3 ou ACC) ont des niveaux de PE plus importants en GL, mais accumulent davantage de PC en RL (Fig. 36). En effet en RL, les cellules ayant accumulé la PC prennent une coloration bleu-verte, et réciproquement en GL (Fig. 36). Le type IV, découvert plus récemment chez les cyanobactéries marines du genre *Synechococcus*, regroupe les cellules capables de percevoir les changements entre GL et BL (Palenik, 2001). Ce type d'AC implique une modification du rapport des deux chromophores liés à la PE I et la PE II (PEB et PUB) mais pas de la composition protéique du PBS (Everroad et al., 2006).

### II.3.2 Bases moléculaires de l'acclimatation chromatique de type III

*Fremyella diplosiphon* est la cyanobactérie la plus étudiée vis-à-vis des mécanismes moléculaires impliqués au niveau de l'AC. Placée en RL, cette cyanobactérie ne fabrique quasiment plus de PE, et les bras des PBS sont composés de 3 hexamères de PC (Rosinski et al., 1981; Siegelman et Kycia, 1982). En revanche en GL, les bras ne sont plus composés que d'un hexamère de PC et 3 disques de PE (Fig. 36). Des variations au niveau structural par rapport à ce modèle ont été observées chez d'autres souches (Ohki et al., 1985; Westermann et al., 1993; Westermann et Wehrmeyer, 1995). Dans la majorité des cas, la PC composant les extrémités des bras en lumière rouge, nommée PC2, n'est pas la même que celle composant l'hexamère situé à la base des bras, nommée PC1 (Bryant, 1981). Chez *Fremyella diplosiphon* l'opéron *cpcB1A1* (codant la PC1) est exprimé de façon constitutive, et peu de variations au niveau transcriptionnel surviennent entre RL et GL. L'expression du second opéron *cpcB2A2* (codant la PC2) est quant à elle induite en lumière rouge (Conley et al., 1986; Lomax et al., 1987; Conley et al., 1988; Mazel et al., 1988). Les deux gènes *cpeBA* codant pour les deux sous-unités de la PE sont retrouvés isolés au sein du génome (Mazel et al., 1986; Grossman et al., 1993) alors que trois polypeptides de liaison, deux phycobiline lyases et une protéine régulatrice sont encodées au niveau de l'opéron *cpeCDEST* (Federspiel et Grossman, 1990; Federspiel et Scott, 1992). Lorsque les cellules sont placées en GL, les niveaux de transcription de ces deux opérons augmentent rapidement (Federspiel et Grossman, 1990). L'AC3 implique également une régulation des gènes *pcyA* et *pebAB* impliqués dans la biosynthèse de PCB et PEB respectivement (Alvey et al., 2003; Alvey et al., 2007).

Ces différentes études ont souligné que l'accumulation de PC était maximale en RL autour de 641 nm alors que l'accumulation de PE était induite en GL autour de 541 nm. Ce processus est

photoréversible, suggérant l'implication d'un système de type phytochrome (Vogelmann et Scheibe, 1978). La régulation de l'AC3 fait en réalité intervenir deux systèmes régulateurs distincts (Fig. 37). Le premier système, nommé Rca (pour Regulator for complementary Chromatic Adaptation) est activé en RL. Il contrôle la synthèse de la PC "induite" et en même temps réprime l'expression de l'opéron *cpeCDEST*R (Sobczyk et al., 1994; Li et al., 2008). Le second système, nommé Cgi (pour Control of Green light Induction), n'a pas d'effet sur la synthèse de la PC ou de PcyA mais réprime la synthèse de PE en RL.



**Figure 37** : Bases moléculaires de la régulation de l'acclimation chromatique de type III. En lumière rouge, le système Rca est phosphorylé et active les opérons codant pour la phycocyanine 2 mais réprime les opérons impliqués dans la synthèse de phycoérythrine. D'après Gutu et Kehoe (2012).

Le système Rca est composé de trois protéines régulatrices, RcaE, RcaF et RcaC, dont les gènes forment un opéron (Chiang et al., 1992; Kehoe et Grossman, 1996, 1997). Ces trois protéines sont indispensables à la régulation de l'AC3, et tout mutant d'une de ces protéines entraîne un phénotype similaire aux cellules placées en lumière verte (Kehoe et Grossman, 1997).

RcaE est capable de percevoir les changements de condition lumineuse, alors que les deux autres composants du système, RcaF et RcaC fonctionnent comme des régulateurs de réponse (Seib et Kehoe, 2002; Alvey et al., 2003; Terauchi et al., 2004; Li et Kehoe, 2005). RcaE présente une certaine similarité avec les phytochromes retrouvés chez les plantes supérieures et possède notamment un domaine GAF, liant un chromophore (Kehoe et Grossman, 1996). Cependant les phytochromes des plantes supérieures répondent à un changement de RL/infra-rouge, alors que cette classe de protéine répond au changement RL/GL (Terauchi et al., 2004). D'autre part, seul le domaine GAF semble nécessaire pour la perception du signal contrairement aux phytochromes qui nécessitent trois domaines distincts nommés PAS, GAF et PHY, une particularité ayant amené à nommer "cyanobactériochromes" cette nouvelle classe de photorecepteurs spécifiques des cyanobactéries (Ikeuchi et Ishizuka, 2008). Ces protéines peuvent être impliquées, en plus de l'AC, dans de nombreux processus métaboliques comme le rythme circadien ou le phototactisme (Wilde

et al., 1997; Schmitz et al., 2000; Yoshihara et Ikeuchi, 2004). RcaE se caractérise aussi par la présence d'un module histine-kinase et un site de phosphorylation. RcaF contient également un site de phosphorylation, au niveau d'un résidu aspartate très conservé, et sert de transfert de phosphorylation (Kehoe et Grossman, 1997; Alvey et al., 2003). RcaC est comme RcaF une protéine régulatrice de réponse, et possède deux domaines récepteurs de phosphorylation, retrouvés aux extrémités N-terminale et C-terminale, un domaine de liaison à l'ADN de type OmpR/PhoB ainsi qu'un domaine de transfert de phosphorylation (Chiang et al., 1992; Kehoe et Gutu, 2006).

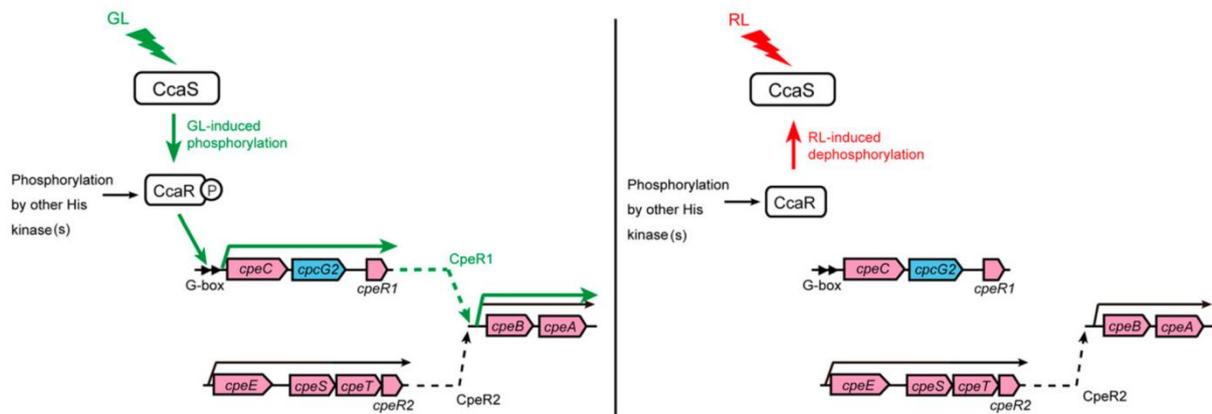
Le système Rca est activé en RL, par autophosphorylation de RcaE (Li et Kehoe, 2005). Le groupement ATP est ensuite transféré au résidu aspartate de RcaF, puis au résidu histidine de RcaC et enfin au module situé en N-terminal (Li et Kehoe, 2005). Comme les régulateurs de la famille OmpR, RcaC agit à la fois en tant qu'activateur et répresseur transcriptionnel (Feng et al., 2003; Alvey et al., 2007; Li et al., 2008). RcaC active ainsi la transcription de l'opéron codant pour la PC2 ainsi que le gène *pcyA*, et réprime transcriptionnellement l'opéron *cpeCDESTR*. RcaC régule ces gènes en se fixant sur des séquences d'ADN spécifiques de 23 à 24 pb, nommées "L Box", situées en amont des gènes *pcyA* et *cpc2* (Alvey et al., 2007; Bezy et Kehoe, 2010). Une "L Box" est également retrouvée en amont de *cpeC*, mais présente une orientation inverse (Alvey et al., 2007). Le système Rca régule également des fonctions non liées à la biosynthèse des PBS, comme la biosynthèse de la Chl (Stowe-Evans et al., 2004; Kehoe et Gutu, 2006).

Contrairement aux gènes activés en RL, les gènes activés en GL sont contrôlés par deux systèmes. On estime que le système Cgi contrôle environ 1/3 de l'expression de l'opéron *cpeCDESTR*, mais n'a aucune influence sur les gènes de PC (Kehoe et Grossman, 1997; Seib et Kehoe, 2002; Kehoe et Gutu, 2006). En RL le système Cgi agit en réprimant post-transcriptionnellement l'opéron *cpeCDESTR* (Bezy et al., 2011). Le fonctionnement du système Cgi reste néanmoins à ce jour beaucoup moins bien compris que le système Rca et ses composants restent encore à identifier (Gutu et Kehoe, 2012). En GL, les deux systèmes ne répriment plus l'expression du gène codant pour l'activateur transcriptionnel CpeR, qui agit sur l'expression des opérons *cpcBA* et *pebAB* en se fixant à l'ADN au niveau de régions conservées appelées "N-boxes" (Cobley et al., 2002; Seib et Kehoe, 2002; Kehoe et Gutu, 2006).

### II.3.3 Bases moléculaires de l'acclimatation chromatique de type II

Les adaptateurs chromatiques de type II régulent uniquement la synthèse de PE (Tandeau de Marsac, 1977). Comme pour l'AC3, ce phénomène, nommé AC2, est retrouvé au sein de plusieurs lignées de cyanobactéries (Tandeau de Marsac, 1977). Un spectre d'action mené sur *Synechocystis* sp. 6701 a permis de mettre en évidence que la synthèse de PE était contrôlée par la GL ( $\lambda=540\text{nm}$ ),

alors que la RL ( $\lambda=640\text{nm}$ ) empêchait celle-ci (Tandeau de Marsac et al., 1980). Les cultures de la cyanobactérie *Nostoc* sp. MAC, une autre souche capable d'AC2, présentent une coloration verte sous lumière blanche et un rapport équimolaire PE:PC. Les cultures deviennent brunes lorsque les cellules sont cultivées sous GL, traduisant une synthèse de PE. Le rapport PE:PC observé est alors de 2:1 (Anderson et al., 1983). Il a ainsi été suggéré qu'en GL, les PBS portaient un disque de PE en plus, par rapport aux cellules sous RL (Anderson et al., 1983; Tandeau de Marsac, 1983). D'autre part ce disque supplémentaire serait toujours ajouté à l'extrémité distale des projections radiales des PBS (Tandeau de Marsac, 1977, 1983).



**Figure 38 :** Bases moléculaires de l'acclimation chromatique de type II chez *Nostoc punctiforme*. Le système implique le cyanobactériochrome *CcaS* et la protéine régulatrice de réponse *CcaR*. D'après Hirose et al. (2010).

La régulation de l'AC2 a été beaucoup moins étudiée que l'AC3. Comme pour cette dernière, l'expression transcriptionnelle de *cpeBA* chez *Synechocystis* sp. BO 8402 est plus importante en GL qu'en RL (Neuschaefer-Rube et al., 2002). Récemment, la découverte d'un système de régulation impliquant deux protéines, *CcaS* et *CcaR*, a permis d'établir un modèle des bases moléculaires de la régulation de ce phénomène chez *Nostoc punctiforme* (Fig. 38). *CcaS* est un cyanobactériochrome, d'abord mis en évidence chez *Synechocystis* sp. PCC 6803 qui ne possède pas de PE. *CcaS* perçoit les variations de condition lumineuse, et en GL est capable de phosphoryler la protéine régulatrice *CcaR* (Hirose et al., 2008b). *CcaR* est un régulateur de réponse de la famille *OmpR* se fixant au niveau de "G Boxes" (Hirose et al., 2008b; Hirose et al., 2010). Les gènes codant pour *CcaR* et *CcaS* sont organisés en opéron (Hirose et al., 2008b; Hirose et al., 2010).

*CcaR* contrôle l'activité transcriptionnelle de l'opéron *cpeC-cpcG2-cpeR1* en activant celui-ci en GL. Par analogie avec l'opéron *cpeCDEST* de *F. diplosiphon*, le régulateur *CpeR1* active également l'opéron *cpeBA* chez *Nostoc punctiforme*. Néanmoins, contrairement aux souches capable d'AC3, la transcription de cet opéron n'est pas supprimée en RL, mais est activé de façon constitutive par le régulateur *CpeR2*, dont le niveau d'expression ne varie pas en fonction de la lumière incidente (Hirose et al., 2010). *CpcG2* est un polypeptide de liaison  $L_{CR}$ , également sous contrôle de *CcaS/R* chez

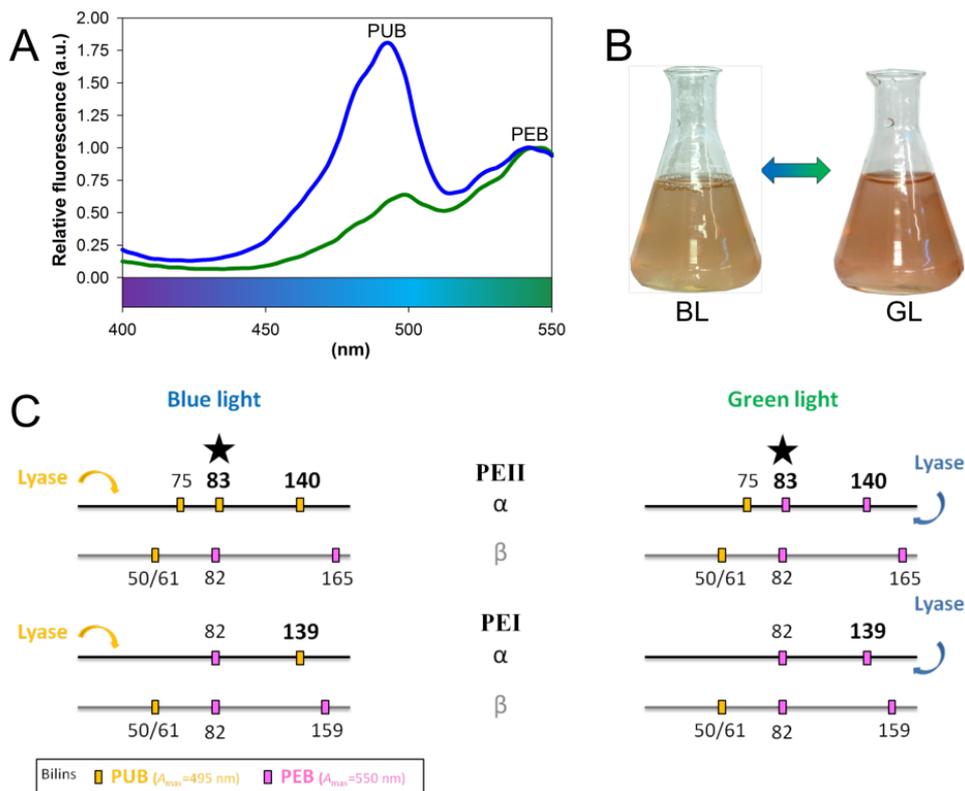
*Synechocystis* sp. PCC 6803, fortement exprimé en GL (Kondo et al., 2005; Hirose et al., 2008b). Chez *Synechocystis*, cette protéine est également impliquée dans le transfert d'énergie (Kondo et al., 2005; Kondo et al., 2009). Ainsi chez *Nostoc punctiforme* le système CcaS/R agirait sur le rapport PE/PC en GL, mais pourrait également agir sur le transfert d'énergie au niveau du PSI sous cette condition lumineuse (Hirose et al., 2010).

Les différentes études menées sur les souches capables d'AC2 et AC3 suggèrent que ces deux groupes sont régulés par deux cyanobactériochromes distincts, CcaS et RcaE. De manière intéressante, ces deux cyanobactériochromes présentent une forte similarité. Les deux protéines régulatrices impliquées dans ces deux systèmes, CcaR et RcaC sont également homologues, et appartiennent à la famille des régulateurs transcriptionnels OmpR. Ces caractéristiques ont mené à suggérer que le système CcaS/R serait un système ancestral de perception des variations de conditions lumineuses, notamment en raison de sa relative simplicité par rapport au système Rca. Ce système aurait ensuite évolué et se serait complexifié de manière à réguler davantage de gènes, entraînant alors des phénotypes plus dynamiques au niveau de l'AC (Tandeau de Marsac, 1977; Hirose et al., 2010).

#### **II.3.4 Originalité de l'acclimatation chromatique de type IV présente chez les *Synechococcus* marins**

Certaines souches de *Synechococcus* marins sont également capables de modifier leur composition pigmentaire en fonction de la lumière incidente, selon un mécanisme nommé AC4 (Everroad et al., 2006). Contrairement aux mécanismes décrits et caractérisés chez les souches d'eau douce, les *Synechococcus* marins ne répondent pas différenciellement entre GL et RL, mais entre BL et GL (Fig. 39B). L'AC4 ne répond pas non plus à un changement d'intensité lumineuse, ni différentes sources d'azote (Palenik, 2001). Contrairement à l'AC2 et AC3, l'AC4 ne semble pas impliquer de changements majeurs au niveau de la composition en PBPs du PBS, mais plutôt une chromophorylation différentielle au niveau de la PE (Everroad et Wood, 2006). En BL, les cellules ont un contenu en PUB plus important qu'en GL, leur permettant ainsi de capturer plus efficacement les photons de couleur bleue (Fig. 39A; Palenik, 2001; Everroad et al., 2006). Récemment, il a été démontré chez la souche *Synechococcus* sp. RS9916 que l'AC4 impliquait le remplacement de PEB par du PUB au niveau de trois résidus cystéine (Fig. 39C). Parmi les cinq sites de la PEI, seule la Cys- $\alpha$ 139 est affectée, alors que ce processus concerne deux des six cystéines de la PEII : Cys- $\alpha$ 83 et Cys- $\alpha$ 140 (Shukla et al., 2012). L'attachement différentiel de ces deux isomères serait sous l'action d'au moins deux couples de phycobilines lyases et lyases-isomérasés. Parmi ces différents candidats, l'enzyme MpeZ a été caractérisée chez *Synechococcus* sp. RS9916 et correspond à une enzyme bi-

fonctionnelle, capable de lier de la PEB en Cys- $\alpha$ 83 de la PEII et de changer sa conformation chimique en PUB (Fig. 39C; Shukla et al., 2012). Le gène *mpeZ* est fortement exprimé en BL par rapport en GL, mais les bases moléculaires de cette régulation restent à ce jour inconnues. Le mutant *mpeZ* a un phénotype GL dans toutes les conditions lumineuses et sa PEII fixe un PEB en Cys-83. Aussi, il a été supposé que la phycobiline-lyase responsable de cet attachement est le complexe CpeY/Z (Biswas et al., 2011; Shukla et al., 2012). Par contre, les phycobilines lyases responsables de la chromophorylation des autres sites affectés par l'AC4 restent à découvrir.



**Figure 39 : Changement de chromophorylation des phycoérythrine (PE) de *Synechococcus* sp. RS9916 lors de l'acclimatation chromatique de type IV. (A) Spectre d'excitation de fluorescence d'une culture placée en lumière bleue (BL, ligne bleue) et lumière verte (GL, ligne verte). (B) Cultures cultivées en BL et GL. (C) Chromophorylation de la PEI et PEII. Seul le changement de pigmentation de la Cys- $\alpha$ 83 est affectée par MpeZ (site désigné par l'étoile). Figure réalisée à partir des données de Shukla et al. (2012).**

Par ailleurs, les comparaisons génomiques menées sur les différents génomes de *Synechococcus* marins n'ont pas permis à ce jour d'identifier de cyanobactériochromes équivalents à ceux impliqués dans les phénomènes d'AC2 ou AC3. Un seul photorécepteur possible AplA, a pu être mis en évidence. Cette protéine avait préalablement été identifiée chez *F. diplosiphon* comme un membre d'une nouvelle classe de photorécepteurs cyanobactériens de fonction inconnue (Six et al., 2007b)

---

---

## PARTIE III

# LES APPORTS DE LA GÉNOMIQUE ET DE LA METAGÉNOMIQUE EN MICROBIOLOGIE MARINE

---

III.1. Les apports de la génomique marine .....	80
III.1.1 Génomique des espèces marines cultivées .....	80
III.1.2 Réduction génomique et succès écologique.....	83
III.1.3 Dynamique des génomes picocyanobactériens.....	84
III.2. L'étude des communautés non cultivées : l'ère de la métagénomique .....	87
III.2.1 De la classification des microorganismes à la phylogénomique.....	87
III.2.2 La naissance de la métagénomique .....	88
III.2.2.1. Métagénomique et clonage : choix du vecteur .....	89
III.2.2.2. Approches "shotgun" et nouvelles technologies de séquençage à haut-débit.....	92
III.2.2.3. Principaux problèmes méthodologiques d'une approche métagénomique .....	94
III.2.3 Intérêts de l'approche métagénomique .....	95
III.2.3.1. Classification, diversité phylogénétique et évolution .....	95
III.2.3.2. Découverte de nouvelles fonctions métaboliques .....	97
III.3. L'entrée dans l'ère de la métagénomique ciblée .....	99
III.3.1 La métagénomique ciblée .....	99
III.3.1.1. Définition et principe.....	99
III.3.1.2. Intérêts de la métagénomique ciblée.....	99
III.3.2 Un champs en pleine mutation .....	100
III.3.2.1. Isoler les microorganismes d'intérêt .....	100
III.3.2.2. Techniques de lyse cellulaire .....	102
III.3.2.3. Amplification de génomes complets : techniques et contraintes .....	102
III.3.3 Apports de la métagénomique ciblée .....	104
III.3.3.1. Obtenir de nouveaux génomes de référence .....	104
III.3.3.2. Interactions microbiennes in situ .....	105
III.3.3.3. Diversité et fonction .....	106

### **III. LES APPORTS DE LA GÉNOMIQUE ET DE LA MÉTAGÉNOMIQUE EN MICROBIOLOGIE MARINE**

L'écologie microbienne marine moderne n'a commencé que dans les années 1970 et apparaît donc comme une discipline récente au regard de bien des domaines de recherche (Pedrós-Alió, 2006). Conditionnées par les moyens techniques disponibles à cette époque, la majorité des études menées s'attachaient à démontrer l'abondance des différents microorganismes en présence, sous-évaluant généralement le rôle écologique de chacune de ces composantes. La génomique est définie comme l'étude de l'information génétique portée par les nucléotides d'un organisme, clé de ses capacités métaboliques et phénotypiques. Favorisé par le développement des techniques de biologie moléculaire, utilisées en routine dans les laboratoires de nos jours (Sanger et al., 1977; Mullis et al., 1986), ce champ de recherche a connu un grand développement depuis quelques années. L'application de ces approches génomiques a particulièrement influencé l'étude des espèces et assemblage microbiens, provoquant une véritable révolution pour interpréter les relations évolutives (Woese, 1987).

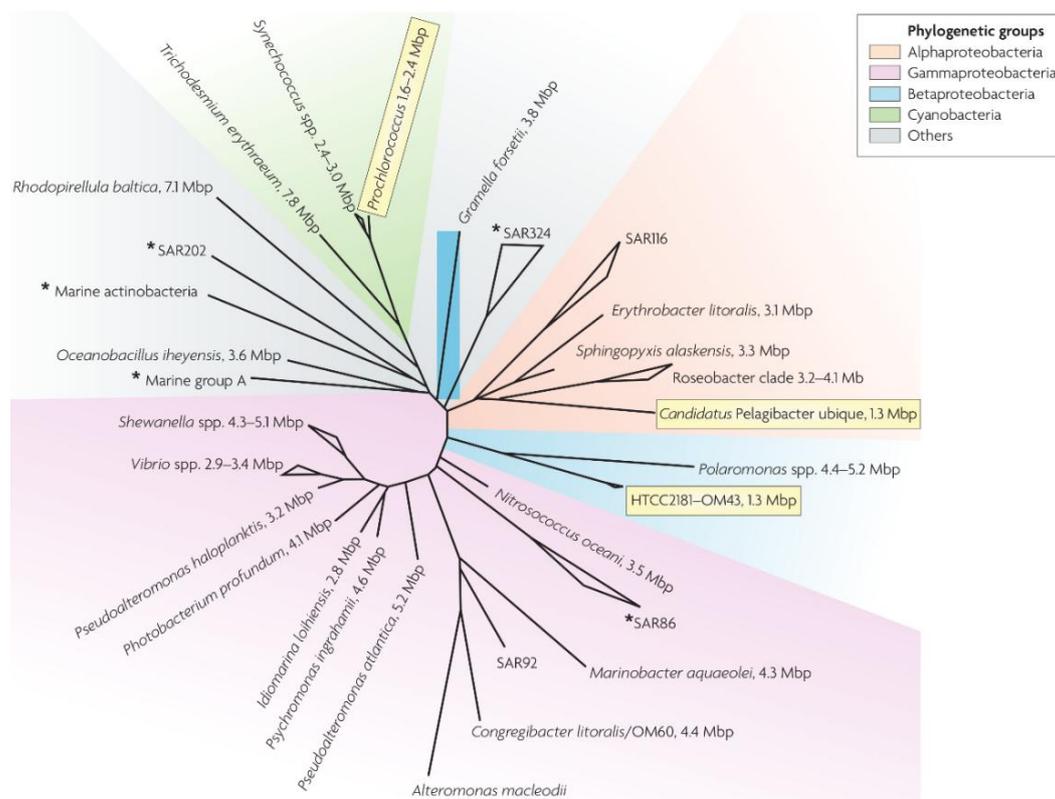
#### **III.1. Les apports de la génomique marine**

##### **III.1.1 Génomique des espèces marines cultivées**

Depuis le séquençage du premier génome bactérien en 1995 (Fleischmann et al., 1995), le nombre de génomes séquencés connaît une croissance exponentielle et plusieurs centaines d'isolats marins ont été séquencés parmi les 4,000 génomes bactériens actuellement déposés. *Methanocaldococcus jannaschii*, une archée isolée des sources hydrothermales, fut le premier microorganisme marin à être entièrement séquencé (Bult et al., 1996). D'autres projets de séquençage ont permis par la suite d'obtenir les premiers génomes de bactéries hétérotrophes (Glöckner et al., 2003; Moran et al., 2004), de cyanobactéries (Dufresne et al., 2003; Palenik et al., 2003; Roca et al., 2003) ou de protistes (Armbrust et al., 2004) (Fig. 40). Les essais de mise en culture de la bactérie la plus abondante des océans ne furent concluants que dix ans après sa découverte (Rappé et al., 2002), et permirent le séquençage en 2005 du génome de *Candidatus Pelagibacter ubique*, appartenant au clade SAR11 (Giovannoni et al., 2005b).

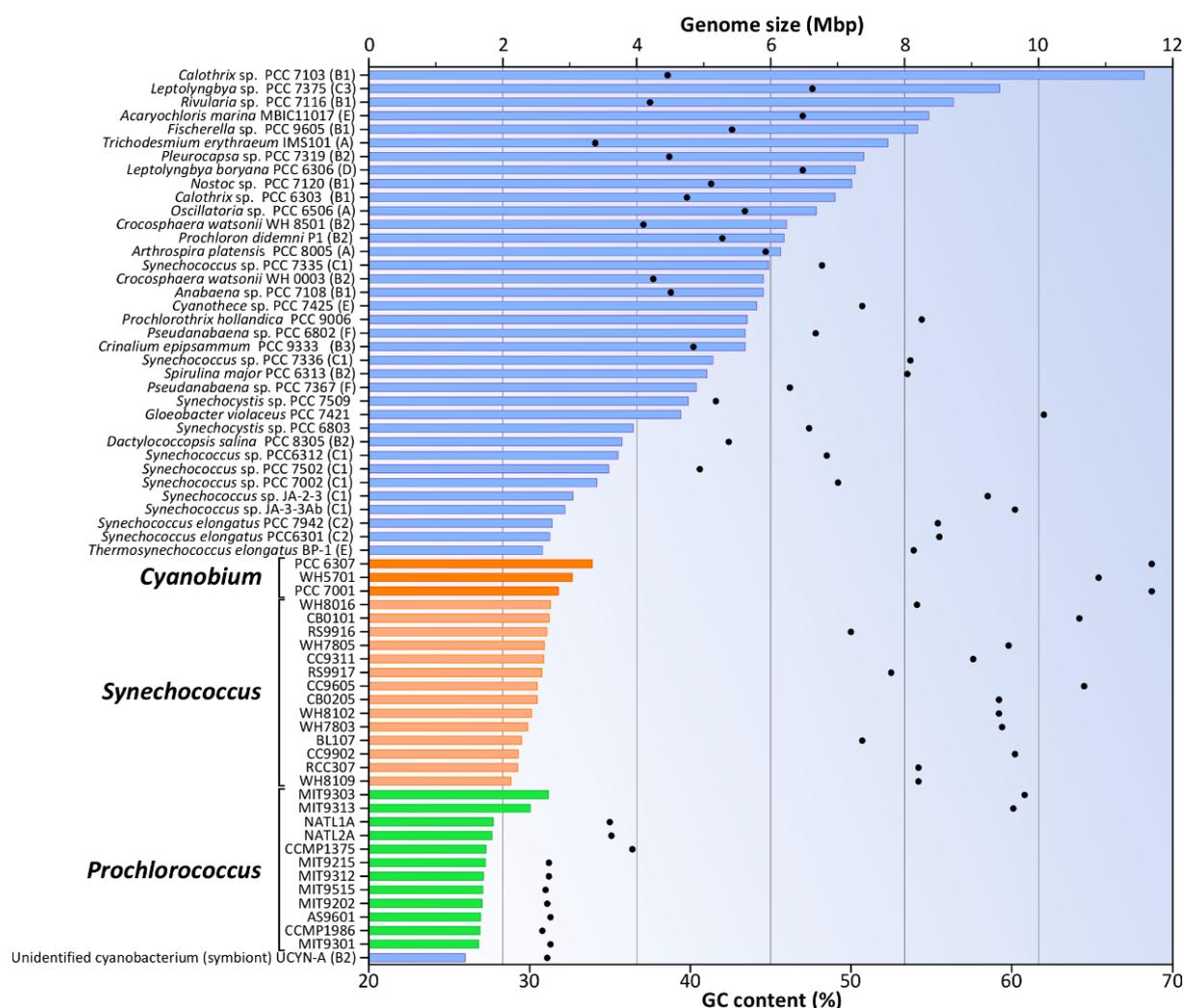
La génomique a été favorisée par le développement de nouvelles techniques de mise en culture (Connon et Giovannoni, 2002; Moore et al., 2007), des NGS (Mardis, 2008a, b; Metzker, 2009) et de nouveaux moyens bioinformatiques permettant de traiter ces vastes ensembles de données. Les programmes de financement de la fondation MOORE ([www.moore.org/](http://www.moore.org/)), du Joint Genome

Institute (www.jgi.doe.gov), ainsi que du programme GEBA (Genomic Encyclopedia of Bacteria and Archaea; Wu et al., 2009; Shih et al., 2013) ont également favorisé cette dynamique.



**Figure 40 : Microorganismes marins entièrement séquencés.** L'arbre phylogénétique représente les divisions majeures du bactérioplancton, les astérisques représentant les groupes les plus abondants. Les microorganismes possédant un génome inférieur à 2 Mbp sont encadrés. Modifié d'après Giovannoni et Stingl (2007).

Le récent projet CyanoGEBA a permis d'ajouter 54 nouveaux génomes cyanobactériens et de mieux couvrir la diversité génétique et phénotypique de cette division (Shih et al., 2013). Les cyanobactéries constituent la division bactérienne pour laquelle le plus grand nombre de génomes marins ont été séquencés (Dufresne et al., 2008; Shih et al., 2013). *Synechococcus* et *Prochlorococcus* constituent 25% de ces génomes publics ( $n=29/126$ ), et ont fait l'objet de nombreuses études de génomique comparative (Kettler et al., 2007; Dufresne et al., 2008; Scanlan et al., 2009). Ces génomes ne sont composés que d'un seul chromosome circulaire comme la majorité des cyanobactéries. *Cyanothece* sp. ATCC 51142 ou *Anabaena variabilis* ATCC 29413 possèdent quant à elles deux chromosomes, dont un linéaire (Welsh et al., 2008). La présence de plusieurs plasmides est commune à plusieurs taxons comme *Synechococcus* sp. PCC 7002 ( $n=6$ ), *Nostoc* sp. PCC 7120 ( $n=6$ ), pouvant même atteindre 9 éléments distincts et représenter 25% des ORFs chez *Acaryochloris marina* MB11017 (Kaneko et al., 2001; Swingley et al., 2008; Shih et al., 2013). Aucun plasmide n'a été mis en évidence au sein des picocyanobactéries marines en culture, même si une étude récente réalisée dans les eaux côtières de l'océan Pacifique a suggéré la possibilité que certaines cellules de *Synechococcus* possèderaient ce genre d'éléments génétiques (Palenik et al., 2009).



**Figure 41 : Taille (barres) et GC% (points) des génomes des picocyanobactéries marines, comparés à une sélection représentative d'autres cyanobactéries.** Les cyanobactéries ont été divisées en 7 sous-clades majeurs (Shih et al., 2013), et cette affiliation phylogénétique est mentionnée entre parenthèses pour chacune des souches. Les picocyanobactéries marines appartiennent au sous-clade C1.

La taille du génome varie entre 1,64 et 2,7 Mb chez *Prochlorococcus* et entre 2,12 et 2,86 Mb chez *Synechococcus* (Kettler et al., 2007; Dufresne et al., 2008). Ainsi, comparé à la taille moyenne du génome bactérien ( $3.69 \pm 1.96$  Mb) ou d'autres génomes cyanobactériens ( $5.33 \pm 3.69$  Mb, avec un record à 11,6 Mb pour *Calothrix* sp. PCC 7103), les picocyanobactéries marines possèdent de petits génomes (Fig. 41; Dufresne et al., 2008; Scanlan et al., 2009). Le pourcentage en bases GC varie entre 31% et 51% chez *Prochlorococcus* (36% en moyenne) et entre 52% et 66% pour *Synechococcus* (59% en moyenne).

### III.1.2 Réduction génomique et succès écologique

Du fait de leur génome réduit, les picocyanobactéries marines possèdent un nombre limité de gènes codant pour des protéines. Ce nombre varie de 2,358 à 3,129 pour *Synechococcus* et de 1,716 à 3,022 chez *Prochlorococcus* (Scanlan et al., 2009). Le génome commun, c'est-à-dire l'ensemble des gènes partagés par l'ensemble des souches, est composé de 1,572 gènes chez *Synechococcus* (représentant en pourcentage entre 50% pour la souche WH5701 à 66% pour CC9902; Dufresne et al., 2008). Ce génome commun est plus petit chez *Prochlorococcus*, avec 1,273 gènes (de 42% pour MIT9303 à 72% pour MED4; Kettler et al., 2007). Le génome commun à l'ensemble des picocyanobactéries marines est composé de 1,100 gènes, et le degré de paralogie est très faible. Ces cyanobactéries constituent ainsi des modèles intéressants car elles possèdent tout l'équipement génétique nécessaire pour la vie cellulaire, mais également pour la photosynthèse oxygénique.

Les comparaisons génomiques effectuées chez *Prochlorococcus* ont permis de mettre en évidence deux groupes bien distincts en terme de taille du génome, beaucoup plus variable qu'au sein du genre *Synechococcus*, et de contenu en bases GC. En effet, les souches LLIV MIT9313 et MIT9303 possèdent un génome de taille similaire à *Synechococcus*, contrairement aux autres écotypes, notamment les HL, chez qui la taille du génome est réduite de près d'un tiers. Ce phénomène de réduction génomique se serait ainsi produit après la divergence précoce de la lignée LLIV, située à la base de la radiation des *Prochlorococcus* (Kettler et al., 2007; Scanlan et al., 2009). La cyanobactérie UCYN-A, récemment renommée *Candidatus Atelocyanobacterium thalassa*, possède quant à elle un génome encore plus réduit que *Prochlorococcus* (1.44 Mb), mais est un symbionte obligatoire d'un haptophyte, du fait de l'absence de certaines voies de biosynthèse d'acides aminés, du PSII, de la Rubisco et du cycle des acides tricarboxyliques (TCA; Tripp et al., 2010; Thompson et al., 2012).

La réduction génomique observée chez *Prochlorococcus* est concomitante avec une réduction du contenu en bases GC au sein des génomes, notamment des souches HL (Dufresne et al., 2005). Ce dernier phénomène s'explique sans doute par la perte de certaines enzymes impliquées dans la réparation des tranversions G:C vers A:T (Rocap et al., 2003; Dufresne et al., 2005; Partensky et Garczarek, 2010). Cette réduction génomique conférerait un avantage évolutif à *Prochlorococcus* en diminuant ses besoins en azote, notamment pour la biosynthèse de l'ADN (Ussery et Hallin, 2004).

Par ailleurs, les cellules de *Prochlorococcus* MED4 (et des autres *Prochlorococcus* à génome réduit) sont plus petites que celles de MIT9313 et MIT9303 (clade LLIV), suggérant qu'une réduction de la taille cellulaire a accompagné la réduction du contenu génomique (Ting et al., 2007; Partensky et Garczarek, 2010). De façon analogue, *Candidatus pelagibacter ubique*, qui représente en moyenne 50% de la fraction bactérienne dans la couche superficielle de l'océan, y compris les régions les plus

pauvres, et qui comme *Prochlorococcus* a un génome réduit (1.3 Mb), est le plus petit microorganisme hétérotrophe marin. Ceci renforce l'idée que la réduction de taille cellulaire constitue un mécanisme d'adaptation aux zones oligotrophes (Morris et al., 2002; Dufresne et al., 2005; Giovannoni et Stingl, 2005). Du fait de leur rapport surface/volume plus élevé, les cellules de petite taille auraient une meilleure capacité d'absorption de nutriments (Dufresne et al., 2003; Giovannoni et al., 2005a). L'efficacité d'absorption lumineuse serait également augmentée chez les microorganismes photosynthétiques, en diminuant l'effet de diffusion lumineuse (Morel et al., 1993).

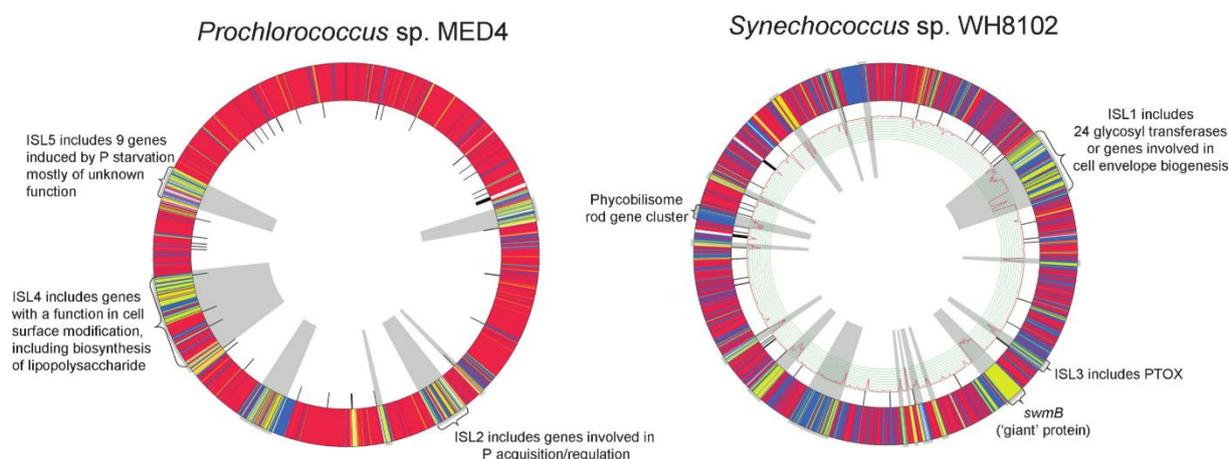
Parmi les 70 familles de gènes spécifiques de *Synechococcus*, dont la majorité aurait été perdus par *Prochlorococcus* au cours de l'évolution, 23 codent pour des protéines impliquées dans la photosynthèse (Dufresne et al., 2008). Parmi ces gènes, on peut citer par exemple les gènes impliqués dans la biosynthèse d'APC et de PC, les gènes *psbA* codant pour l'isoforme D1:2 de la protéine D1 qui confère au PSII une meilleure résistance à la photoinhibition (Garczarek et al., 2008) ou des gènes impliqués dans le transport et l'assimilation du carbone (Dufresne et al., 2008). Le gène *kaiA*, codant pour un des trois composants de l'horloge circadienne, a également été perdu au cours de l'évolution, de sorte que l'horloge de *Prochlorococcus* est minimaliste (Holtzendorff et al., 2008). Les gènes *psbU* et *psbV*, codant pour deux protéines impliquées dans la stabilisation du cluster mangénèse (ou OEC; Roose et al., 2007), ont également été perdus par les écotypes HL. Enfin, un certain nombre de transporteurs ABC, participant notamment à l'acquisition de solutés compatibles, sont absents des génomes de *Prochlorococcus* et pourrait induire une plus faible tolérance aux fortes salinités par rapport à *Synechococcus* (Hagemann et al., 1997; Scanlan et al., 2009).

A l'inverse, les comparaisons géniques ont montré que seuls 10 gènes étaient spécifiques de *Prochlorococcus* (Partensky et Garczarek, 2010). Parmi ceux-ci on retrouve également des gènes photosynthétiques, comme les gènes *pcb*, codant pour les protéines d'antenne associées au PSII, le gène *PcCao* codant pour une Chl *b* synthetase permettant la synthèse de Chl *b*<sub>2</sub> et un homologue de *crtL-e*, impliqué dans la synthèse de lycopène (Stickforth et al., 2003; Satoh et Tanaka, 2006).

### III.1.3 Dynamique des génomes picocyanobactériens

Le séquençage de nouveaux génomes de picocyanobactéries a permis de mettre en évidence que si les génomes communs de *Synechococcus* et *Prochlorococcus* se stabilisaient, le supragénome picocyanobactérien (encore appelé pan-genome; Tettelin et al., 2008) qui comprend l'ensemble des gènes d'un ensemble taxonomique, continuait de s'accroître fortement et semblait loin d'être saturé (Habib, C, comm. pers.). Des modélisations mathématiques suggèrent que le séquençage d'un nouveau génome ajoute ~277 gènes à ce pan-genome (Baumdicker et al., 2010). Pour le seul genre *Prochlorococcus*, le pan-génome contiendrait 57,792 genes, pour une taille effective de population

de  $1.01 \times 10^{11}$  cellules (Baumdicker et al., 2012), valeur cependant bien plus faible que la taille réelle estimée de la population de *Prochlorococcus* qui est de l'ordre de  $10^{27}$  cellules (Garcia-Pichel et al., 2003). La majeure partie du pan-génome est constitué de gènes non-essentiels au fonctionnement cellulaire, qui sont distribués de façon différentielle entre les différentes souches ou écotypes. Le transfert latéral de ces gènes entre les différentes lignées (Rocap et al., 2003; Coleman et al., 2006; Palenik et al., 2009; Sharon et al., 2009), constituerait ainsi un avantage évolutif notamment dans l'adaptation à de nouvelles niches écologiques (Tettelin et al., 2008).



**Figure 42** : Ilots génomiques chez *Prochlorococcus marinus* sp. MED4 et *Synechococcus* sp. WH8102. Les gènes communs, accessoires et uniques sont représentés par des couleurs rouge, bleu et jaune respectivement, au niveau du cercle extérieur. Les régions grises représentent les îlots génomiques putatifs prédits en se basant sur la fréquence tétranucléotidique pour WH8102 (Dufresne et al., 2008) et sur la synténie pour MED4 (Coleman et al., 2006). Les lignes noires au niveau du cercle intérieur représentent les ARNt. D'après Scanlan (2012).

Les gènes accessoires (partagés par au moins 2 génomes) ou uniques sont généralement concentrés dans des régions génomiques spécialisées appelées « îlots génomiques » (Fig. 42; Coleman et al., 2006; Dufresne et al., 2008). Ces îlots génomiques ont originellement été identifiés chez *Prochlorococcus* en se basant sur des interruptions de la synténie en comparant des souches phylogénétiquement proches. Chez *Synechococcus*, où la synténie apparaît moins bien conservée entre lignées distinctes, l'identification a été possible en analysant les variations de la fréquence trinuécléotique ou tétranuécléotique (Palenik et al., 2006; Dufresne et al., 2008). Ces régions sont très dynamiques et évoluent plus rapidement que le reste du génome, grâce à des phénomènes d'acquisition ou de délétions géniques, comme en atteste le faible recrutement de « reads » métagénomiques au niveau de ces régions comparées au reste du génome (Coleman et al., 2006; Kettler et al., 2007; Rusch et al., 2007; Frias-Lopez et al., 2008).

Ces îlots représentent une part très variable du génome, de 10,6 à 31,2% (Dufresne et al., 2008). La taille et la position de ces structures génétiques diffèrent largement d'un génome à l'autre.

Le rôle de ces régions reste difficile à définir car plus de 60% des gènes sont de fonction inconnue. Néanmoins il est probable que la majorité des îlots participent à l'adaptation à de nouvelles niches écologiques. En effet, chez *Prochlorococcus marinus* MED4, environ un quart des îlots sont différentiellement exprimés en cas de carence en phosphate ou de stress lumineux (Coleman et al., 2006). D'autre part l'analyse transcriptomique des communautés microbiennes au niveau du gyre subtropical du Pacifique Nord, dominé par l'écotype HLII, a mis en évidence un fort niveau d'expression des gènes retrouvés dans ces régions génomiques (Frias-Lopez et al., 2008). Parmi les gènes de fonction connue, la grande majorité est impliquée dans la biosynthèse ou la modification de la paroi cellulaire comme des glycosyl transferases ou glycosides hydrolases (Fig. 42; Dufresne et al., 2008; Scanlan, 2012). Certains de ces îlots auraient un rôle majeur dans la résistance au broutage ou aux phages (Palenik et al., 2003; Zwirgmaier et al., 2009; Avrani et al., 2011). Chez *Prochlorococcus*, des gènes impliqués dans le métabolisme nutritif comme les gènes d'assimilation du phosphate, nitrate, nitrite, acides aminés ou métaux traces sont également localisés dans les îlots génomiques (Coleman et al., 2006; Kettler et al., 2007; Martiny et al., 2009a). Certains îlots portent des gènes hypothétiques de très grande taille, parmi lesquels seul *swmB*, impliqué dans la motilité, a à ce jour été caractérisé (McCarren et Brahmsha, 2007). Les gènes impliqués dans la synthèse et régulation des bras des PBS constituent également un îlot génomique qui a la particularité d'être présents chez toutes les souches de *Synechococcus* (Six et al., 2007b). Les gènes *hli*, qui codent pour les protéines HLIP impliquées dans la réponse au stress lumineux, sont souvent retrouvés au niveau des îlots, et constitueraient des points de recombinaison (Garczarek et al., 2008). Enfin, la présence d'ARN non codants et de régulateurs transcriptionnels suggèrent que certains îlots interviendraient dans la régulation du métabolisme à un niveau plus global (Dufresne et al., 2008; Steglich et al., 2008; Georg et al., 2009; Scanlan et al., 2009).

## III.2. L'étude des communautés non cultivées : l'ère de la métagénomique

### III.2.1 De la classification des microorganismes à la phylogénomique

Les scientifiques ont depuis toujours cherché à faire l'inventaire des espèces, ou entités biologiques, qui les entourent. La caractérisation des différents organismes composant un écosystème se définit comme la structure des communautés. La taxonomie anime depuis des centaines d'années les zoologistes et botanistes. Chez les plantes et les animaux, les classifications effectuées se sont basées sur un certain nombre de critères complexes, à la fois morphologiques, phénotypiques et dans une moindre mesure physiologiques. Les scientifiques ne se sont cependant pas seulement attachés à classer les êtres vivants, mais également à établir les relations existantes entre ces différentes entités, et ainsi d'établir des scénarios d'évolution. Ces relations évolutives ont pour la première fois été illustrées sous la forme d'un arbre phylogénétique par Charles Darwin (Darwin, 1859). Ce scientifique pionnier, souvent critiqué, a réussi à faire émerger l'idée de concept phylogénétique et la systématique moderne comme en atteste les fameux arbres d'Ernst Haeckel (Haeckel, 1866). Au niveau microbien, et notamment pour les bactéries, qui apparaissent comme des systèmes biologiques plus simples d'un point de vue morphologique, les critères physiologiques ont quant à eux été préférés (van Niel, 1955).

Les premières études de la diversité bactérienne en milieu marin ont commencé à la fin du XIX<sup>ème</sup> siècle et utilisaient principalement des enrichissements ou des cultures pures (Certes, 1884; Frankland et Frankland, 1894). Plusieurs décennies plus tard, Baumann et collaborateurs fournissaient la première description et classification des bactéries marines à partir de plus de 200 isolats (Baumann et al., 1972). Avec les développements méthodologiques de la fin des années 1970, notamment de la microscopie à épifluorescence (Ferguson et Rublee, 1976; Hobbie et al., 1977), les écologues marins ont finalement découverts le fort biais dû à la mise en culture (Staley et Konopka, 1985). Par comparaison morphologique des cultures obtenues à partir d'un échantillon complexe, il fut en effet démontré que la fraction cultivable ne représentait qu'une faible partie, inférieure à 1%, de la diversité totale (Amann et al., 1995).

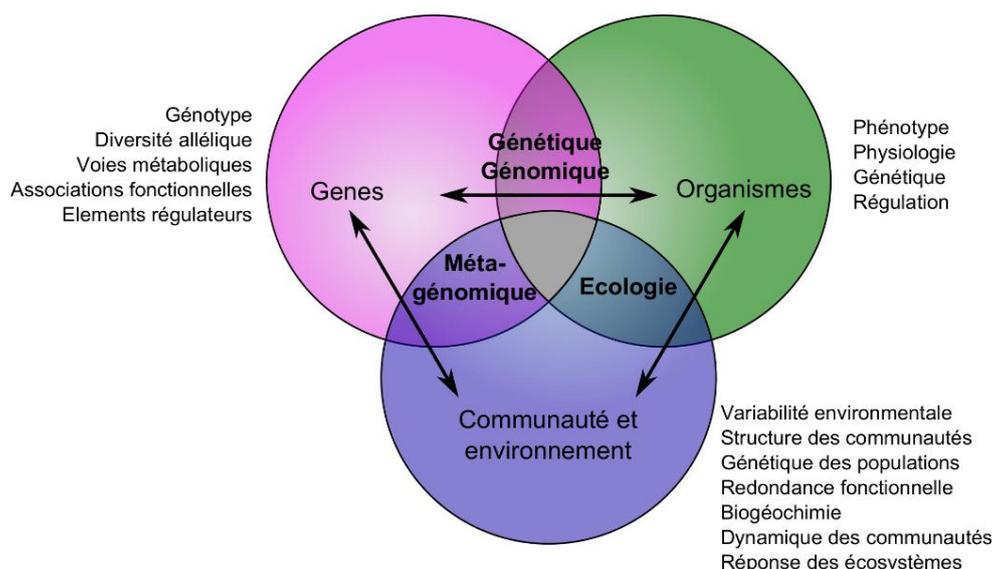
A la fin des années 1960, les biologistes prirent conscience que des molécules comme l'ADN ou les protéines fournissaient une image de l'évolution (Zuckerlandl et Pauling, 1965) et que ces données pouvaient être utilisées pour fabriquer des arbres phylogénétiques (Fitch et Margoliash, 1967). Les progrès technologiques des années 1970 au niveau des techniques de séquençage provoquèrent un changement brutal dans la perception de la complexité du monde vivant. Initialement développée sur des cultures pures, la phylogénie moléculaire se focalisant sur l'ARNr 16S devint alors l'outil de référence pour la classification taxonomique des bactéries. Le ribosome,

présent chez tous les organismes vivants, fournissait un outil idéal pour l'étude des relations entre les espèces, même très éloignées (Fox et al., 1980; Lane et al., 1985; Olsen et al., 1986). Les travaux de Carl Woese ont ainsi bouleversé les plus hauts niveaux de taxonomie du monde vivant (Woese et Fox, 1977) et ont remis en question toute la taxonomie bactérienne (Woese, 1987; Woese et al., 1990).

Au-delà de sa robustesse, la phylogénie moléculaire a surtout permis d'étudier les communautés de manière culture-indépendante. Après extraction des acides nucléiques, les gènes d'intérêt ont pu être clonés puis séquencés, directement à partir d'échantillons naturels (Stahl et al., 1984; Pace et al., 1986). De nombreux écosystèmes ont ainsi été étudiés, mettant en évidence de nombreuses lignées phylogénétiques au sein des eubactéries et confirmant l'existence d'une bien plus grande diversité au sein des écosystèmes que celle qui avait été identifiée à travers la mise en culture (Giovannoni et al., 1990; Britschgi et Giovannoni, 1991; Schmidt et al., 1991; Fuhrman et al., 1993; Wright et al., 1997; Hugenholtz et al., 1998; Rappé et al., 2000; Rappé et Giovannoni, 2003). Dans certains habitats, les organismes les plus abondants n'avaient même pas de représentant proche en culture (Rappé et Giovannoni, 2003; Schloss et Handelsman, 2004). En milieu marin, la distribution ubiquiste de *Candidatus Pelagibacter ubique* (autrement connu sous le nom de clade SAR11), une  $\alpha$ -protéobactérie présente en très forte abondance, a été révélée par ce genre d'études phylogénétiques (Giovannoni et al., 1990; Morris et al., 2002). Ces approches ont également permis d'identifier de nouvelles lignées archéennes (DeLong, 1992; Fuhrman, 1992; Karner et al., 2001) ou de picoeucaryotes (Moon-van der Staay et al., 2001).

### III.2.2 La naissance de la métagénomique

Le séquençage de génomes dans les années 1990 a permis d'accroître de manière substantielle les connaissances sur la génétique, la capacité physiologique et l'écologie microbienne. Parallèlement, les nombreuses études phylogénétiques ont souligné la nécessité d'appréhender le fonctionnement des écosystèmes de manière culture-indépendante. Dérivée de la génomique, une nouvelle approche consistant à appréhender l'information génétique des communautés microbiennes directement à partir de l'environnement naturel a ainsi vu le jour à la fin des années 1990. Cette discipline, dont la phylogénomique constituait les premiers travaux (Lane et al., 1985; Pace et al., 1986; Schmidt et al., 1991), fut dénommée « métagénomique » par Handelsman (Fig. 43; Handelsman et al., 1998) et est parfois appelée génomique environnementale, génomique des communautés, génomique des populations ou encore écogénomique (Handelsman, 2004; DeLong, 2009). Contrairement à la phylogénomique, l'approche métagénomique ne s'intéresse pas simplement à la diversité microbienne, mais permet l'étude des voies métaboliques en présence, afin d'en déduire le fonctionnement global d'un écosystème.



**Figure 43** : Intersections des disciplines traditionnelles de la microbiologie et de la métagénomique. Modifié d'après Delong (2009).

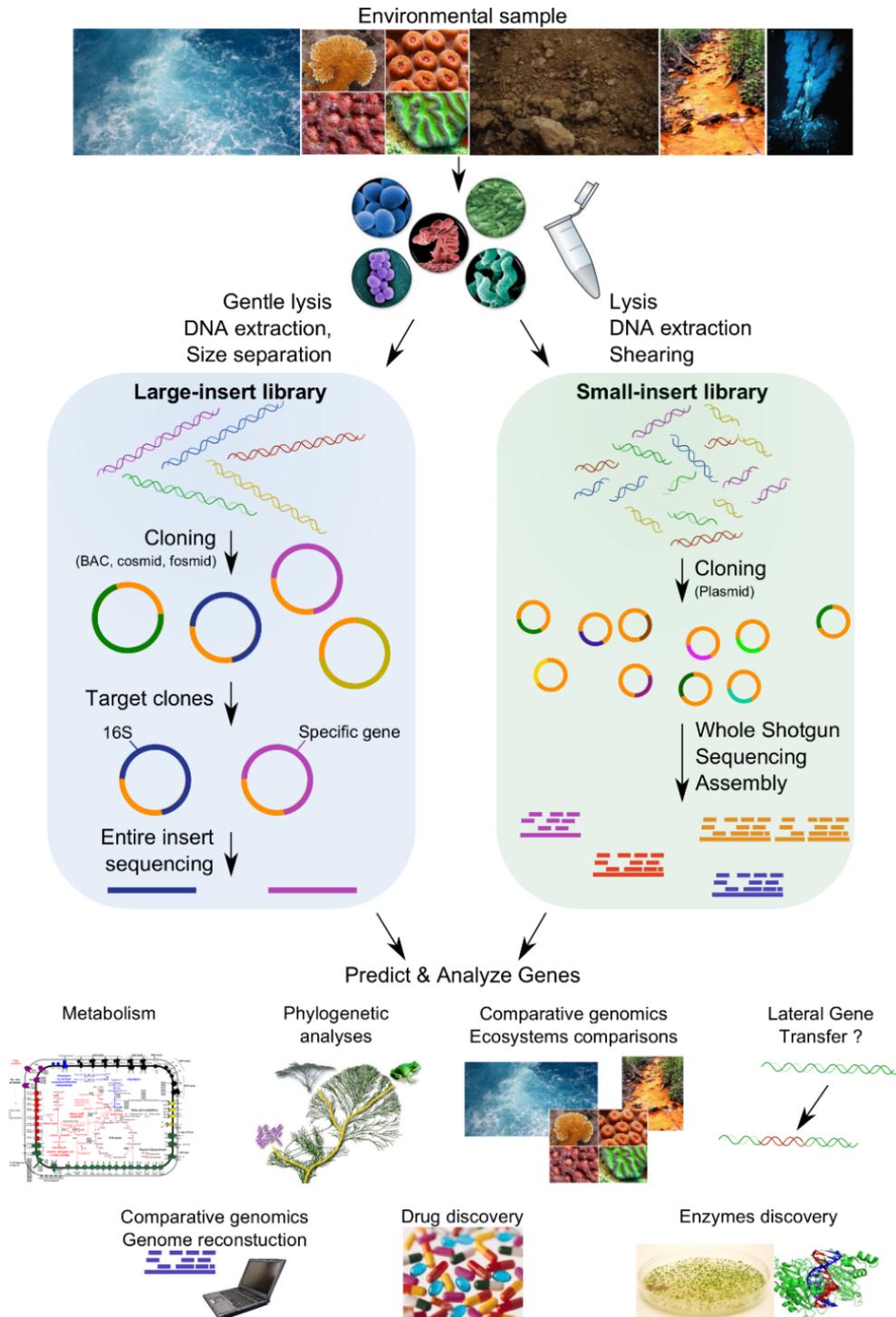
### III.2.2.1. Métagénomique et clonage : choix du vecteur

L'approche métagénomique consiste à cloner l'ADN d'un échantillon environnemental, correspondant à un mélange de différents représentants, dans des vecteurs de clonage, puis à étudier le répertoire de gènes correspondant. La métagénomique apparaît donc comme une méthode de choix pour compléter les études utilisant les approches traditionnelles de la microbiologie (Figs. 43 et 44).

L'idée de cloner des fragments d'ADN provenant d'échantillons environnementaux a d'abord été appliquée par Schmidt en 1991, qui utilisait des phages lambda pour cloner des fragments d'ADN de 10 à 20 Kpb extraits à partir d'eau de mer (Schmidt et al., 1991). Cette approche qui consiste à utiliser des petits fragments d'ADN est dénommée en anglais "shotgun". Les progrès apportés sur les vecteurs de clonage en 1992 ont marqué un réel tournant dans les études métagénomiques. L'introduction des BACs (bacterial articial chromosome) et des fosmidés (amélioration des cosmides) ont ainsi permis le clonage de larges fragments d'ADN génomique (Kim et al., 1992; Shizuya et al., 1992).

**Les BACs.** Ces vecteurs de clonage ont été mis au point en se basant sur la structure des plasmides conjugatifs retrouvés chez *E. coli*, appelés « facteurs F ». Les vecteurs BACs ne contiennent pas seulement les éléments génétiques nécessaires à leur propre répliation (*oriS* et *repE*), mais aussi les gènes *parA* et *parB*, permettant de maintenir ces éléments en un faible nombre de copies (généralement de une à deux). Un site de clonage présent au niveau du gène de la  $\beta$ -galactosidase, permettant un criblage des clones positifs par colorimétrie (Shizuya et al., 1992). Utilisés pour la première fois par Rondon afin de caractériser des inserts de *Bacillus cereus* (Rondon, 1999), les BACs

ont par la suite servi à une étude métagénomique de la flore bactérienne du sol (Rondon et al., 2000) et du picoplancton marin (Béjà et al., 2000). Si les BACs peuvent permettre le clonage de fragments supérieurs à 300 Kpb, la taille moyenne des inserts est souvent comprise entre 80 et 200 Kpb, ce qui en fait le vecteur de choix pour l'étude de voies métaboliques complètes (Béja et al., 2000), le criblage de Fogènes ou d'îlots génomiques d'intérêt biotechnologique (MacNeil et al., 2001; Schloss et Handelsman, 2003) ou encore la cartographie de génomes entiers.



**Figure 44 :** Principe de la fabrication d'une banque métagénomique utilisant des fragments d'ADN environnemental de petite ou grande taille et possibilités d'analyses.

**Les fosmidés.** Dérivés des cosmides, ces vecteurs offrent une meilleure efficacité de clonage que les BACs, mais permettent de cloner des fragments de taille plus faible, généralement comprise entre 35 et 45 Kpb. Egalement développés sur la base du facteur F, ils possèdent à la fois les propriétés d'un plasmide (origine de répllication, site de clonage, gène de résistance au chloramphénicol), les éléments régulateurs du nombre de copies des BACs, ainsi que les éléments *cos* des phages lambda (Kim et al., 1992). Ces derniers permettent l'encapsulation des éléments génétiques au sein de phages, qui seront utilisés pour la transformation des cellules hôtes. Cette étape, appelée transfection, est beaucoup plus efficace que les méthodes traditionnelles. Les fosmidés, possédant les gènes *parAB*, sont présents en une ou deux copies par cellule, permettant de s'affranchir de la toxicité observée préalablement avec les cosmides. Les fosmidés ont été les premiers éléments génétiques à être utilisés dans le cadre d'une approche métagénomique par l'équipe de DeLong qui a préparé une banque à partir d'un échantillon environnemental marin et identifié un clone contenant un gène d'ARNr 16S d'une archée jusque là incultivée (Stein et al., 1996). Les librairies de fosmidés sont devenues de loin la méthode de clonage la plus utilisée pour les études en milieu marin, notamment à cause des contraintes méthodologiques rencontrées avec les BACs (Gilbert, 2010) et avec les cosmides, dont l'utilisation reste peu répandue (López-García et al., 2004).

**Tableau 3 : Avantages et inconvénients des banques métagénomiques fabriquées à partir de petits ou grands inserts.** Modifié d'après Daniel (2005).

Type de librairie	Avantages	Inconvénients
<b>Petits inserts (plasmides)</b>	Permet de détecter les gènes exogènes faiblement exprimés grâce aux vecteurs présents en grand nombre de copies Pouvoir exprimer les gènes exogènes grâce aux vecteurs inducibles Clonage adapté pour l'ADN contaminé ou de moindre qualité Techniquement simple	Inserts de petite taille (5 à 8 Kpb) Effort de criblage important Non adapté pour le clonage des voies métaboliques encodées par de larges clusters de gènes
<b>Grands inserts (BACs, cosmides, fosmidés)</b>	Inserts de grande taille Faible effort de criblage Adapté pour le clonage des voies métaboliques encodées par de larges clusters de gènes Convient pour la reconstruction de génomes	Détection des gènes exogènes plus difficile Expression limitée des gènes exogènes par les vecteurs inducibles Techniquement difficile Requiert de l'ADN purifié de haut poids moléculaire

Le choix du vecteur de clonage dépend d'un certain nombre de paramètres, notamment de la quantité de matériel génétique disponible. La fabrication d'une librairie BAC nécessite davantage de matériel de départ que pour les librairies avec de petits inserts. La qualité du matériel est également un critère essentiel, notamment pour la fabrication de librairies BACs ou fosmidés, nécessitant du matériel génétique de haut poids moléculaire, et donc des méthodes d'extraction adaptées. Les objectifs de l'étude conditionnent aussi ce choix : une approche par « shotgun » suffira pour le séquençage massif de gènes mais ne conviendra pas au criblage de larges régions génomiques. Les avantages et inconvénients de ces approches sont résumés dans le Tableau 3.

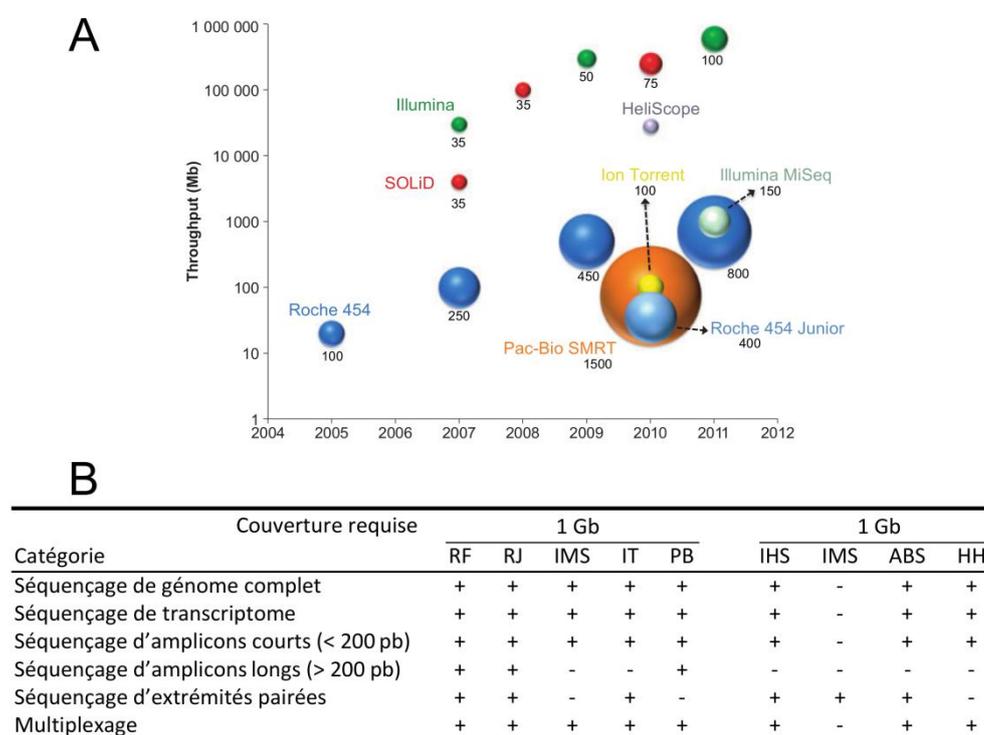
### III.2.2.2. *Approches "shotgun" et nouvelles technologies de séquençage à haut-débit*

Une étape importante de l'approche métagénomique est le séquençage massif des gènes d'un échantillon naturel par "shotgun". Autrement appelée séquençage aléatoire, cette technique consiste à séquencer les extrémités des inserts, présents le plus souvent au sein de plasmides. La reconstruction de régions génomiques de taille plus importante est par la suite permise par assemblage des différentes séquences générées (ou reads). Deux études célèbres publiées en 2004, ont utilisé cette approche et un séquençage par la méthode de Sanger: l'étude d'un site au large des Bermudes (Venter et al., 2004) et l'analyse d'un biofilm des eaux de drainage acides de la mine de Richmond (Tyson et al., 2004). La première a consisté à faire un inventaire de la diversité génétique bactérienne de cette région bien caractérisée de la mer des Sargasses (Venter et al., 2004). Profitant de la disponibilité des séquenceurs après le séquençage du génome humain, cette étude a permis d'obtenir plus d'un Gb de séquences non redondantes, 1,2 million de gènes de fonction inconnue, affiliés à 1.800 espèces différentes et 48 phylotypes (Venter et al., 2004). Cette étude a également utilisé de nouvelles méthodes de bioinformatique. Cette gigantesque masse de données, qui représentait 5 % des séquences déposées sur GenBank à la fin de l'année 2004 (Riesenfeld et al., 2004), a permis de mettre en évidence l'abondance de certaines fonctions métaboliques, comme la protéorhodopsine ou des gènes archéens intervenant dans l'oxydation de l'ammonium (voir §III.2.3.2). Néanmoins, malgré l'abondance de séquences, cette étude a aussi mis en évidence certaines limitations techniques notamment la difficulté de reconstruire des génomes complets en raison de la complexité microbienne de cet environnement océanique (Riesenfeld et al., 2004).

Ces contraintes techniques n'ont pas été rencontrées dans le cas de l'étude du drainage minier (Tyson et al., 2004) car cet environnement présente des conditions physico-chimiques extrêmes en termes d'acidité et de teneur en métaux lourds, où seul un biofilm de faible diversité peut se développer (Bond et al., 2000). Ainsi, à partir de seulement 76,2 Mpb de séquences, les auteurs ont réussi à reconstituer les génomes presque complets de deux représentants microbiens (la bactérie *Leptospirillum* et l'archée *Ferroplasma*), auxquels se sont ajoutés 3 autres génomes partiels (Tyson et al., 2004). L'analyse des voies métaboliques représentées au sein de ces génomes a permis d'identifier les voies de fixation du carbone et de l'azote, ainsi que les sources d'énergie utilisées par chaque microorganisme.

Le milieu des années 2000 a été marqué par l'apparition des nouvelles technologies de séquençage (NGS) pouvant générer des dizaines de millions de séquences en parallèle. Ces diverses technologies permettent le séquençage d'ADN, sans clonage préalable pour amplifier et séparer les différentes matrices, et permettent ainsi de s'affranchir des biais liés au clonage (voir revue de

Shokralla et al., 2012). Néanmoins, les techniques NGS possèdent leurs propres limitations techniques (voir §III.2.2.3; Mardis, 2008a). Depuis l'apparition du pyroséquençage (Margulies et al., 2005), les technologies NGS ont beaucoup évolué et su faire face à plusieurs défis à la fois techniques et financiers. Le premier a été la réduction des coûts, de plusieurs ordres de magnitude (Wetterstrand, 2013). L'autre point qui a fortement évolué concerne l'amélioration de la qualité des séquences produites, tant par la précision de la méthode que par la longueur des séquences générées (Fig. 45A). Ce point, auquel s'ajoute la quantité de données produite, est sans conteste celui qui a conditionné le succès respectif de chacune des méthodes. Le dernier défi concerne la réduction des biais techniques qui seront traités dans le paragraphe suivant.



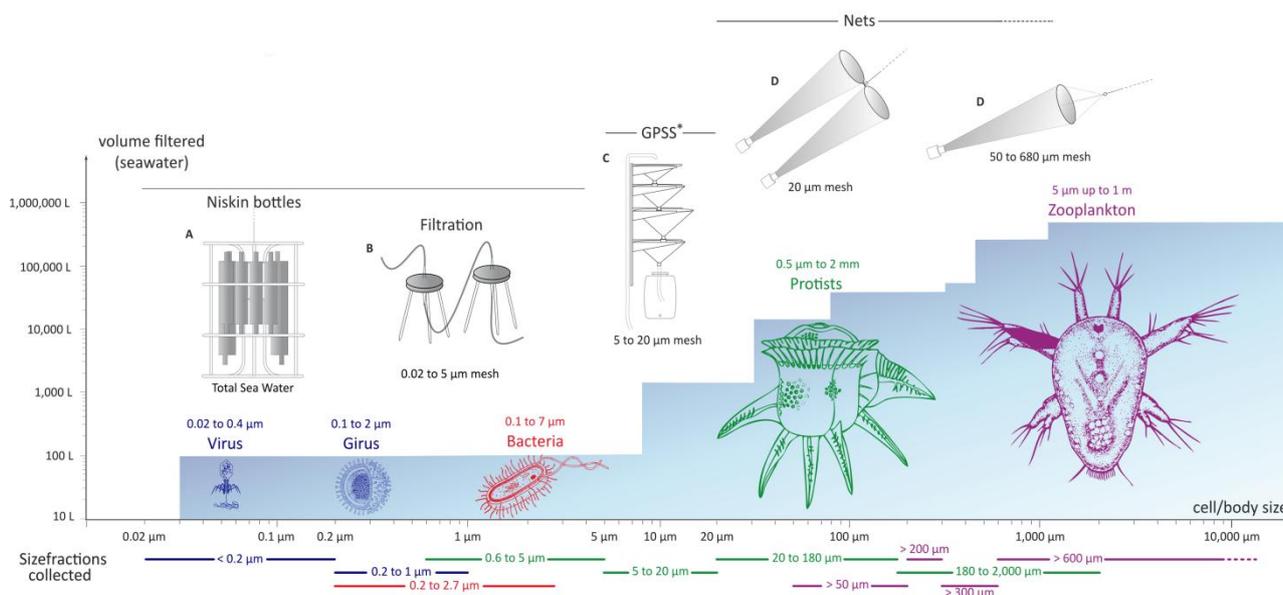
**Figure 45 : Nouvelles technologies de séquençage et domaines d'utilisation.** (A) Développement historique des différentes technologies. Le diamètre de chaque bulle représente la taille de la séquence générée (en pb). Les couleurs représentent chacune des technologies. (B) Applications possibles de chacune des méthodologies. Abréviations : RF :Roche 454 GS-FLX; RJ :Roche 454 Junior; IHS : Illumina HiSeq 2000; IMS : Illumina MiSeq; ABS : AB SOLiD system; I. T :Ion Torrent; PB :Pacific Biosciences system; HH :Helicos HeliScope. Modifié de Shokralla et al. (2012).

Malgré les particularités intrinsèques à chacune des méthodologies NGS (chimie, méthode d'incorporation des nucléotides, outils de détection), elles utilisent toutes soit une fragmentation de l'ADN génomique soit la génération d'amplicons, suivie de la génération d'une banque et de la détection des nucléotides incorporés (Glenn, 2011; Zhang et al., 2011). Les séquenceurs peuvent ensuite être classés en deux catégories: (i) ceux utilisant la technologie PCR et (ii) ceux ne nécessitant pas d'amplification préalable. La première catégorie est représentée par les séquenceurs de première génération, de loin les plus utilisés de nos jours. Le pyroséquençage (technologie 454

Roche; Margulies et al., 2005), les technologies Illumina HiSeq ou MiSeq (initialement connue sous le nom Solexa; Illumina In.), SOLiD (Applied Biosystems) ou Ion Torrent (Life Technologies) font partie de cette catégorie. La deuxième catégorie rassemble les technologies HeliScope (Helicos Biosciences) et PacBio (Pacific Biosciences). Plusieurs revues récentes décrivent et comparent chacune de ces technologies (Mardis, 2008b, a; Glenn, 2011; Shokralla et al., 2012).

### III.2.2.3. Principaux problèmes méthodologiques d'une approche métagénomique

L'étude des communautés microbiennes comporte un certain nombre de contraintes et de limitations méthodologiques. Dans toute étude s'intéressant à la structure des communautés, l'échantillonnage reste l'une des principales limitations techniques. Il dépend principalement de la question scientifique. Plusieurs paramètres, dont la complexité et la densité microbienne de l'échantillon et la présence de contaminants (Smit et al., 2006), déterminent également la stratégie d'échantillonnage à appliquer. Dans les études en milieu marin, les échantillons sont bien souvent préfiltrés pour enlever débris et cellules de grande taille (Somerville et al., 1989), éliminant ainsi les cellules qui pourraient être adsorbées ou interagir avec ces organismes (symbioses, etc.). Un fractionnement de taille est également bien souvent appliqué, permettant d'accéder aux fractions virales (Breitbart et al., 2002), procaryotes (Venter et al., 2004) et aux autres composantes de plus grande taille du plancton (Fig. 46). Les extractions d'acides nucléiques à partir d'eau de mer restent cependant moins problématiques que celles réalisées à partir d'échantillons de sols ou de sédiments, limitant ainsi les biais techniques associés à cette étape (Delmont et al., 2010; Delmont et al., 2011).



**Figure 46 : Techniques utilisées pour l'échantillonnage et le fractionnement des différentes classes de taille du plancton. D'après Karsenti et al. (2011).**

Bien que les étapes de clonage soit moins répandue depuis l'avènement des techniques NGS, leur utilisation, e.g. à des fins de sélection d'une communauté particulière, peut conduire à la sous-estimation de certains gènes (Béjà et al., 2000) ou de groupes phylogénétiques (Pham et al., 2008; Temperton et al., 2009). Même si les méthodes NGS s'affranchissent le plus souvent de ces biais méthodologiques, elles possèdent leurs propres limitations. Parmi celles-ci, la génération d'erreurs de séquençage reste la contrainte technique la plus documentée. Ces techniques, notamment le pyroséquençage, sont basées sur une amplification par PCR pouvant générer de nombreuses erreurs (Cline et al., 1996; Huse et al., 2007; Dohm et al., 2008) et induire la formation de chimères (Haas et al., 2011). Ces erreurs interfèrent au niveau de l'analyse des données notamment pour la détermination des cadres de lecture ouvertes (ORFs; Krause et al., 2006; Hoff, 2009), l'affiliation phylogénétique (Tringe et Rubin, 2005; Von Mering et al., 2007; Krause et al., 2008) ou encore les prédictions fonctionnelles (Harrington et al., 2007). Ces erreurs posent également problème dans l'interprétation et l'analyse des taxons présents en faible abondance (biosphère "rare"; Sogin et al., 2006; Huse et al., 2007; Huse et al., 2010; Kunin et al., 2010), notamment lorsqu'elle sont réalisées en ciblant des marqueurs très conservés, tel que l'ARNr 16S (Wintzingerode et al., 2006; Schloss et al., 2011).

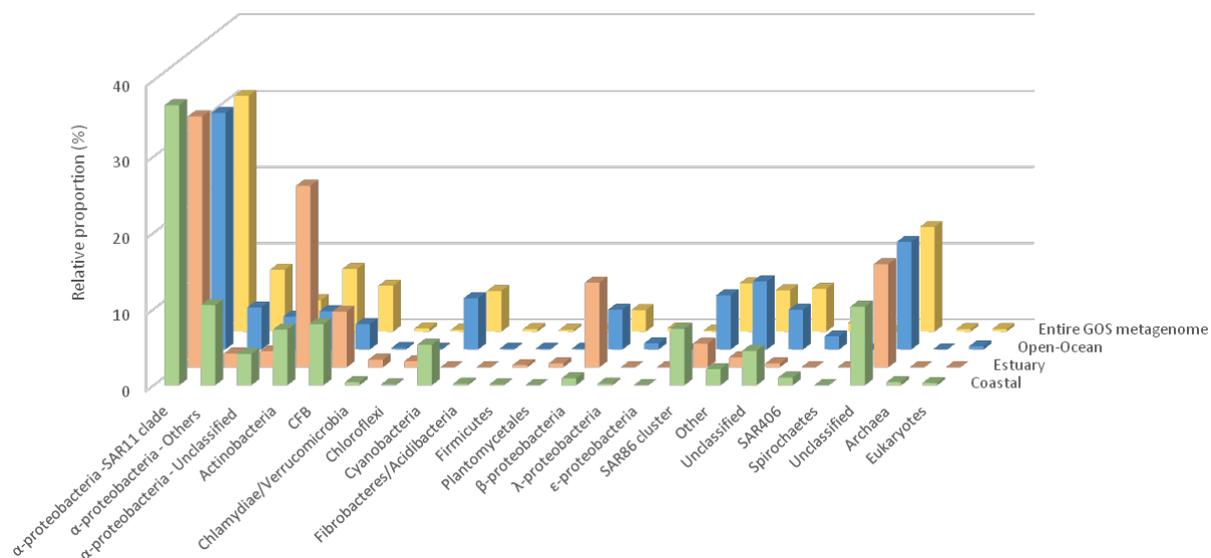
Ces contraintes techniques ont conduit au développement d'une multitude d'outils et de logiciels bioinformatiques destinés à l'analyse de ces données. Néanmoins, aucune standardisation n'existe à l'heure actuelle quant au traitement et à l'analyse de ces grands sets de données. Face à l'utilisation de plus en plus massive de ces méthodologies, des outils statistiques et l'application de méthodes standardisées seront nécessaires afin de permettre une comparaison réaliste des différents métagénomomes (DeLong, 2009).

### III.2.3 Intérêts de l'approche métagénomique

#### III.2.3.1. *Classification, diversité phylogénétique et évolution*

La métagénomique peut-être utilisée en ciblant des marqueurs phylogénétiques afin d'identifier de nouvelles lignées, d'améliorer la classification bactérienne et ainsi d'en déduire certains principes de l'évolution microbienne. Initialement, les études métagénomiques s'attachaient à identifier les clones présentant ces gènes d'intérêt puis à les séquencer (Stein et al., 1996; Béja et al., 2000; Béjà et al., 2002; Quaiser et al., 2002; Hallam et al., 2003; Liles et al., 2003; Quaiser et al., 2003; López-García et al., 2004). L'évolution des techniques NGS, notamment du pyroséquençage, permet aujourd'hui de séquencer en profondeur les amplicons générés directement à partir des échantillons environnementaux (Sogin et al., 2006), en ciblant principalement les régions hypervariables du gène codant pour l'ARNr 16S ou 18S (Huse et al., 2008). La diversité procaryote a

ainsi été explorée au niveau des écosystèmes marins (Sogin et al., 2006; Huber et al., 2007; Brown et al., 2009), d'eau douce (Eiler et al., 2011), de l'océan Arctique (Galand et al., 2009a; Kirchman et al., 2010) ou encore des sols (Roesch et al., 2007; Barberán et al., 2011; VanInsberghe et al., 2013). La connaissance des communautés eucaryotes, principalement protistes, a également fortement progressé ces dernières années (Brown et al., 2009; Cheung et al., 2010; Nolte et al., 2010; Behnke et al., 2011; Pawlowski et al., 2011), comme celle des organismes archéens en Arctique (Roesch et al., 2007; Galand et al., 2009a; Galand et al., 2009b).



**Figure 47 : Diversité microbienne observée au sein de différents habitats océaniques, côtiers et estuariens.** Données provenant du GOS, selon Biers et al. (2009).

Plusieurs nouvelles lignées bactériennes, archéennes ou protistes ont ainsi pu être mises en évidence par ce type d'approche. Les sols sont parmi les environnements qui ont le plus bénéficié de la métagénomique avec la découverte du nouveau phylum des Acidobacteria, qui est ubiquiste et abondant (Rondon et al., 2000; Quaiser et al., 2002; Liles et al., 2003). Les connaissances sur la diversité archéenne présente au niveau des sols (Quaiser et al., 2002; Treusch et al., 2004), des sources hydrothermales (Moreira et al., 2004) ou des symbioses en milieu marin (Schleper et al., 1998; Donia et al., 2011; Thompson et al., 2012) ont également énormément progressé de part les études de métagénomique.

En milieu marin, les études à grande échelle comme le GOS (Global Ocean Sampling) ont permis d'examiner la biogéographie et la diversité de groupes spécifiques (Venter et al., 2004; Rusch et al., 2007). Parmi la vingtaine de divisions majeures des eubactéries (Fig. 1), les environnements océaniques sont dominés par les α-protéobactéries, β-protéobactéries, γ-protéobactéries et cyanobactéries (Fig. 47; Venter et al., 2004; Rusch et al., 2007; Biers et al., 2009; Yooseph et al., 2010). Les bactéries phototrophes aérobies anoxygéniques (AAP) constituent une composante importante du bactérioplancton, représentant en moyenne 1 à 5% de la communauté et présentant

une grande diversité génétique notamment à l'interface entre le milieu côtier et l'océan ouvert (Yutin et al., 2007; Jiao et al., 2010). Au sein des environnements océaniques, les clades SAR11, appartenant aux  $\alpha$ -protéobactéries, et SAR86, aux  $\gamma$ -protéobactéries, sont ubiquistes et souvent retrouvés en forte abondance (Fig. 47; Morris et al., 2002; Schattenhofer et al., 2009). Néanmoins, la grande diversité de contenu et d'organisation retrouvée au sein de ces génomes a mis en évidence l'importance du phénomène de sélection naturelle au sein des populations de ces deux généralistes, résultant probablement d'un isolement biogéographique (Wilhelm et al., 2007; Dupont et al., 2012). Une approche par pyroséquençage, ciblant le gène marqueur *petB*, codant pour le cytochrome *b6*, a quant à elle été récemment menée chez les *Synechococcus* marins, permettant d'identifier de nouveaux clades et d'explorer la diversité de ces microorganismes à un niveau de résolution taxonomique beaucoup plus fin que ce qui avait été réalisé précédemment en ciblant l'ARNr 16S ou l'ITS (Mazard et al., 2012a).

La métagénomique offre l'opportunité de mieux comprendre les processus évolutifs que sont la réduction génomique, mais aussi l'acquisition de gènes. L'existence de transferts massifs de gènes a ainsi été mise en évidence chez les Crenarchaeotes (López-García et al., 2004). L'identification de gènes photosynthétiques au sein des génomes de virus marins démontrent l'importance de ces éléments dans la dynamique des populations picocyanobactériennes (Sharon et al., 2009). La mise en évidence récente de plasmides au sein des populations marines de *Synechococcus* laissent également présager leur implication dans le transfert latéral de gènes (Palenik et al., 2009).

Les études métagénomiques ont également démontré l'extraordinaire diversité des virus marins (Breitbart et al., 2002; Edwards et Rohwer, 2005; Williamson et al., 2008; Kristensen et al., 2010; Zhao et al., 2013).

### III.2.3.2. Découverte de nouvelles fonctions métaboliques

En fournissant l'accès à un grand répertoire de gènes, la métagénomique permet d'identifier certaines nouvelles fonctions métaboliques, dont certaines d'intérêt biotechnologique. La métagénomique a ainsi par exemple révélé la présence de la protéorhodopsine au sein de nombreuses lignées bactériennes, notamment du clade SAR86 (Béja et al., 2000). Ces photorécepteurs n'avaient jusqu'alors été identifiés qu'au sein des archées halophiles (Oesterhelt et Stoeckenius, 1971). Après caractérisation, ces protéines membranaires se sont avérées être des pompes à protons permettant la synthèse d'ATP sous l'action de la lumière (Béja et al., 2000). Cette nouvelle famille de rhodopsine, pour lesquels deux variants spectraux ont été identifiés, a permis de définir un nouveau type de phototrophie au sein des environnements marins (Béja et al., 2001). Les études métagénomiques ultérieures ont permis de révéler la forte abondance et la distribution globale des protéorhodopsines au sein de différentes lignées bactériennes et archéennes marines

(De La Torre et al., 2003; Sabehi et al., 2003; Sabehi et al., 2004; Venter et al., 2004; Rusch et al., 2007), ainsi que les processus évolutifs ayant permis cette large répartition phylogénétique (Frigaard et al., 2006; McCarren et DeLong, 2007). Notons que la protéorhodopsine est également présente chez l' $\alpha$ -protéobactérie *Candidatus Pelagibacter ubique* (clade SAR11; Giovannoni et al., 2005a). Le rôle de la protéorhodopsine reste peu documenté, mais fournirait un avantage métabolique en présence de lumière (Lami et al., 2009), en augmentant notamment la capacité de survie des cellules en milieu dépourvu en nutriments (DeLong et Béjà, 2010; Gómez-Consarnau et al., 2010).

Parrallèlement, de nouveaux types de photosynthèse aérobie anoxygénique (AAP) ont été identifiés par des approches de métagénomique (Béjà et al., 2002). Les bactéries AAP, utilisant de la bactériochlorophylle  $a$  pour capturer l'énergie lumineuse, présentent une large distribution au sein des océans (Kolber et al., 2000; Kolber et al., 2001). La majorité des gènes impliqués, notamment *bch* et *crt*, codant respectivement pour la bactériochlorophylle  $a$  et les voies de biosynthèse des caroténoïdes, ainsi que les gènes *puf*, codant pour le centre réactionnel, sont inclus dans une région génomique de 45 Kpb (Alberti et al., 1995). Ces gènes initialement découverts chez les  $\alpha$ -protéobactéries ont par la suite été mis en évidence au sein de la classe  $\beta$  et  $\gamma$ , ainsi que chez les Roseobacterales (Béjà et al., 2002; Allgaier et al., 2003; Oz et al., 2005; Cho et al., 2007; Fuchs et al., 2007).

La découverte de ces deux modes de phototrophie en milieu marin a complètement changé la vision du monde scientifique sur la nature et la diversité des systèmes utilisant la lumière comme source d'énergie.

### **III.3. L'entrée dans l'ère de la métagénomique ciblée**

#### **III.3.1 La métagénomique ciblée**

##### *III.3.1.1. Définition et principe*

Les études métagénomiques classiques de type shotgun fournissent un répertoire substantiel de gènes, mais sont fortement limitées techniquement pour la reconstitution de génomes complets, en dépit de grands progrès en bioinformatique (Blainey, 2013). Les génomes assemblés se limitent bien souvent aux taxons les plus abondants dans l'écosystème considéré (Gilbert et Dupont, 2011). De plus, ces génomes sont en général des séquences consensus provenant de multiples microorganismes (Luo et al., 2011a; Dupont et al., 2012; Iverson et al., 2012). Ces assemblages sont bien souvent insuffisants pour modéliser les réponses des communautés microbiennes ou concevoir le rôle de chaque entité biologique au sein des écosystèmes. En effet, la distinction entre les gènes spécifiques d'un seul organisme et ceux partagés par différents représentants est rendue difficile.

La nécessité d'étudier les gènes dans un contexte génomique complet apparaît primordiale. Se situant à l'interface entre la génomique (approche culturale) et la métagénomique (approche culture-indépendante), la métagénomique ciblée est une approche émergente. L'idée est de diminuer la complexité d'un échantillon à un nombre de cellules plus restreint, afin de séquencer spécifiquement les organismes d'intérêt. Longtemps impossible d'un point de vue technique, la génomique à partir d'une seule cellule (SCG pour "Single Cell Genomics") est désormais réalisable (Raghunathan et al., 2005; Zhang et al., 2006). Bien que la métagénomique ciblée n'englobe pas uniquement cette dernière méthodologie, un intérêt particulier à la SCG sera apporté au cours du paragraphe qui suit.

##### *III.3.1.2. Intérêts de la métagénomique ciblée*

En diminuant la complexité du jeu de données, les étapes d'assemblage de génomes issus de SCG sont grandement facilitées, contournant même le problème des génomes consensus. Cette méthodologie permet ainsi d'analyser la variabilité génétique à l'échelle de la cellule, en étudiant par exemple les régions hypervariables, les taux de mutations, ou les échanges génétiques (Pamp et al., 2012; Shapiro et al., 2012). Ces phénomènes, qui peuvent être sous l'influence du transfert horizontal de gènes, sont beaucoup plus fréquents qu'initialement présumés et ont joué un rôle fondamental dans l'évolution des procaryotes (Fraser et al., 2007; Vergin et al., 2007; Shapiro et al., 2012). En étudiant la diversité au sein d'organismes très proches, par exemple d'une même espèce, la SCG permet d'accroître les connaissances sur la notion de génome commun et de pan-génome (Tettelin et al., 2008).

Egalement confrontée à de fortes contraintes méthodologiques, la SCG n'apparaît pas comme une méthode de substitution de la métagénomique mais vient compléter ces études. Cette stratégie permet de relier sans ambiguïté les informations sur les fonctions métaboliques et l'affiliation phylogénétique des cellules. La SCG permet également d'obtenir des informations génomiques sur la biosphère rare ou peu documentée. L'augmentation du nombre de génomes séquencés grâce à de telles approches permet également de grandement améliorer les modèles d'annotation de gènes ou d'affiliation taxonomique actuellement utilisés dans toute approche métagénomique (Wu et al., 2009).

### III.3.2 Un champs en pleine mutation

La SCG a été rendue possible par un grand nombre d'évolutions technologiques. Néanmoins, ces techniques restent difficiles à mettre en œuvre et demande une préparation rigoureuse de l'échantillon. Indépendamment de l'organisme ciblé, l'approche SCG se compose (1) d'une étape d'isolement, (2) d'une lyse cellulaire permettant de libérer l'ADN qui (3) sera amplifié par amplification de génomes complets (WGA) avant (4) de fabriquer une banque en vue (5) du séquençage.

#### III.3.2.1. *Isoler les microorganismes d'intérêt*

Les techniques d'isolement cellulaire peuvent être divisées en deux catégories : la micromanipulation et la microencapsulation. Contrairement aux techniques d'encapsulation, la micromanipulation permet d'identifier la cellule d'intérêt avant de procéder à son isolement. Les techniques de micropipettage ont été les premières techniques utilisées pour l'isolement de bactéries ou d'archées (Raghunathan et al., 2005; Ishoey et al., 2006; Kvist et al., 2007; Hongoh et al., 2008; Woyke et al., 2010; Grindberg et al., 2011). Dans le même temps, des techniques de microfluidique ont été développées dans le laboratoire de Stephen Quake, mais en raison de la grande complexité de ces dernières, leur utilisation reste restreinte (Ottesen et al., 2006; Marcy et al., 2007). La microfluidique présente néanmoins l'avantage de diminuer les volumes dans lesquels les cellules sont isolées, diminuant ainsi les possibilités de contamination des échantillons pour la SCG (Marcy et al., 2007; Youssef et al., 2011; Leung et al., 2012). Récemment d'autres technologies très sophistiquées, et encore peu répandues, ont également été appliquées pour des applications SCG, utilisant des pinces optiques ou optoélectroniques. Ces deux méthodes permettent l'isolement dans des volumes extrêmement faibles, de l'ordre du femtolitre, de particules colloïdales (comme les cellules). Les pinces optiques utilisent l'énergie de photons générés par des lasers pour maintenir les cellules (Ashkin et Dziedzic, 1987; Ashkin et al., 1987) et ont été utilisées en combinaison avec de la microfluidique pour de la SCG à partir de bactéries (Blainey et al., 2011; Youssef et al., 2011; Marshall

et al., 2012; Pamp et al., 2012). Permettant une meilleure conservation des cellules, les pinces optoélectroniques n'ont pas été à ce jour utilisées pour des approches SCG. Le fonctionnement de ces méthodes est décrit dans une revue récente (Blainey, 2013).

**Tableau 4 : Méthodes d'isolement utilisées pour des approches génomique à partir d'une cellule (SCG).** Les pinces optoelectroniques n'ont pas été utilisées à ce jour en SCG. Modifié d'après Blainey (2013).

Technique	Technologie	Critères d'isolement	Volume échantillon	Débit (# cellules)	Vol. réactif (coût WGA)	Remarques	
Dilution manuelle	Pipettage aléatoire	Aucun	Microlitre	Dizaine	Microlitre (élevé)	Aléatoire	
Encapsulation	Puce microfluidique	Chambres microfabriquées	Fluorescence et observation microscopique possible	Nanolitre	Centaine	Nanolitre (faible)	Simple, rapide
	Cytométrie en flux	Microgouttes en phase aqueuse	Propriétés optiques, intensité de fluorescence	Nanolitre	Centaine	Nanolitre (faible)	Rapide, nécessité d'un grand nombre de cellules, expérience nécessaire
	Microgouttes en émulsion	Microfluidique passive	Fluorescence et observation microscopique possible	Picolitre	Milliers	Picolitre (faible)	Excellent rendement, mais non standardisée
Micromanipulation	Micropipettage	Pipettage précis	Fluorescence et observation microscopique possible	Nanolitre	Dizaine	Microlitre (élevé)	Requiert plateforme
	Microfluidique en flux	Microfluidique active	Fluorescence et observation microscopique possible	Picolitre	Dizaine	Nanolitre (faible)	Contrôle fu flux nécessaire
	Pincés optiques	Microscopie laser	Fluorescence et observation microscopique possible	Femtolitre	Dizaine	Microlitre (élevé)	Intégrable avec microfluidique
	Pincés optoélectroniques	Opto-électrocinétique	Fluorescence et observation microscopique possible	Femtolitre	Centaine - Milliers	Non appliqué	Intégrable avec microfluidique

Parmi les méthodes d'encapsulation, la cytométrie en flux s'est imposée comme la méthode de choix en SCG. Cette technologie à haut débit permet l'isolement des cellules en se basant sur un certain nombre de critères physiques ainsi que le couplage à différentes techniques de marquage. La majorité des études SCG menées à ce jour utilisent la cytométrie en flux (Stepanauskas et Sieracki, 2007; Rodrigue et al., 2009; Woyke et al., 2009; Swan et al., 2011; Yoon et al., 2011; Dupont et al., 2012). Cette technique permet l'encapsulation de cellules en phase liquide dans un volume d'échantillon de l'ordre du nanolitre, augmentant la probabilité de contamination par rapport aux approches décrites précédemment (Blainey, 2013). Une encapsulation en microgouttes d'huile par mélange ou en système microfluidique permet de diminuer ce volume (Thorsen et al., 2001). Principalement utilisée pour des applications génomiques par PCR à haut débit (Tewhey et al., 2009; Agresti et al., 2010; Zeng et al., 2010), la viscosité du mélange pose le problème de la délivrance du matériel génétique et de la compatibilité avec les réactifs conventionnels de WGA (Abate et al., 2010), ainsi que le ciblage spécifique du microorganisme d'intérêt (Link et al., 2006). Enfin, dans une des premières études SCG, Zhang et collaborateurs ont utilisé la technique de dilution manuelle afin d'isoler les cellules, mais avec une très faible efficacité (Zhang et al., 2006).

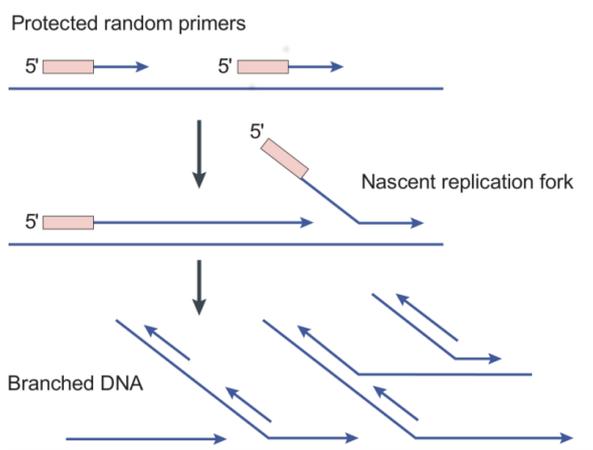
### III.3.2.2. *Techniques de lyse cellulaire*

La lyse cellulaire est une étape nécessaire permettant la libération du matériel génétique au sein du volume réactionnel. Le protocole de lyse doit à la fois permettre une lyse efficace et compatible avec les réactifs de WGA. Des procédés physiques (dénaturation thermique, choc thermique, sonication), l'utilisation d'enzymes hydrolytiques, de détergents ou de pH extrêmes peuvent être utilisés lors de cette étape (Marcy et al., 2007; Swan et al., 2011; Stepanauskas, 2012; Blainey, 2013). La lyse alcaline reste à ce jour la technique la plus utilisée (Raghunathan et al., 2005), mais son efficacité varie largement en fonction de l'organisme considéré et se situe typiquement en dessous 40% (Marcy et al., 2007; Swan et al., 2011).

### III.3.2.3. *Amplification de génomes complets : techniques et contraintes*

Depuis plus de 20 ans, plusieurs méthodes d'amplification de génomes complets (WGA) ont été développées. Ces dernières sont toutes basées sur l'utilisation d'une ADN polymérase à faible taux d'erreurs mais diffèrent par le type d'amplification (synthèse par déplacement de brins ou séparation thermique), le type d'amorces, le biais d'amplification, la taille des produits amplifiés, ainsi que la disponibilité de kits (voir revue de Blainey, 2013). Parmi ces différentes méthodes, la DOP-PCR utilise des oligonucléotides hybrides de 15 bases, partiellement dégénérés (Telenius et al., 1992). L'amplification se déroule en deux étapes: la première consistant à favoriser l'extension des amorces, la seconde favorisant la réplication. Une récente variante, nommée D-DOP-PCR (displacement-DOP-PCR, commercialisée sous le nom PicoPLEX) a été développée pour permettre la synthèse par déplacement de brin lors de la première étape. Malgré ces progrès, l'utilisation de la DOP-PCR en SCG reste marginale (Leung et al., 2012).

L'amplification par déplacement de brin (MDA) est communément utilisée pour les approches SCG. La MDA utilise des hexamères protégés en 3' afin d'éviter leur dégradation due à l'activité 3' – 5' exonuclease de l'ADN polymérase *phi29*, capable de synthèse par déplacement de brin en conditions isothermales (Fig. 48; Dean et al., 2001; Zhang et al., 2001). Cette enzyme possède une grande processivité et permet la synthèse de fragments de hauts poids moléculaire, jusqu'à 70 Kpb (Blanco et Salas, 1984; Blanco et al., 1989; Dean et al., 2001), ainsi qu'un biais de plusieurs ordres de magnitude plus faible que les autres techniques de WGA (Dean et al., 2002). La MDA permet d'obtenir des quantités d'ADN de l'ordre du microgramme à partir de quelques nanogrammes, et même d'une seule cellule. De nombreux kits sont proposés commercialement par GE Healthcare et Qiagen.



**Figure 48 : Principe de la réaction d'amplification par déplacement de brins utilisant l'ADN polymérase phi29.** D'après Lasken (2012).

L'amplification par MDA est de type exponentiel et aspécifique, imposant que toute trace d'ADN contaminant soit en concentration suffisamment faible dans l'échantillon pour ne pas interférer avec les analyses ultérieures. La suppression de cette contamination reste un défi majeur en SCG (Blainey, 2013). La présence d'ADN contaminant provient (1) de l'échantillon lui-même, (2) de l'environnement (laboratoire instrumentation, pratiques de laboratoire, etc.) et (3) des réactifs utilisés pour la WGA. La réduction de la contamination provenant des deux premiers points peut-être réalisée par des pratiques de laboratoire très rigoureuses, ainsi qu'une attention particulière portée à la propreté et à la décontamination des instruments utilisés (Stepanauskas et Sieracki, 2007). Des procédés comme l'autoclavage, l'utilisation de détergents, le traitement aux UVs, l'utilisation de salles blanches ou de hottes HEPA sont autant de solutions proposées (Stepanauskas et Sieracki, 2007; Woyke et al., 2011; Stepanauskas, 2012). L'étape d'isolement cellulaire est critique, puisqu'elle conditionne autant la qualité que le volume final d'échantillon. La microfluidique présente l'avantage de travailler avec de faibles volumes réactionnels et de diminuer la susceptibilité de l'échantillon à la contamination (Marcy et al., 2007). L'ADN exogène, provenant de cellules lysées, présent dans un échantillon est également éliminé efficacement en utilisant deux étapes de tris cellulaires successives par cytométrie en flux (Rodrigue et al., 2009).

Les fournisseurs de kits de WGA préconisent l'utilisation de 10 ng d'ADN. Cette préconisation n'est pas une limitation technique, la réaction étant possible à partir de quantités beaucoup plus faibles, mais permet de s'affranchir des contraintes posées par la présence d'ADN dans les réactifs, estimée à 1 femtogramme pour 50 µL réactionnels (Woyke et al., 2009; Blainey et Quake, 2011). La décontamination de ces réactifs par traitement UV apparait comme une possibilité efficace, simple à mettre en oeuvre (Woyke et al., 2011).

Les premières tentatives de reconstruction de génomes à partir d'une seule cellule ont mis en évidence un biais d'amplification et l'apparition de séquences chimériques durant la réaction de MDA (Zhang et al., 2006). Le biais d'amplification se traduit à la fois (1) au niveau de la couverture du génome et (2) entre les différents taxons lors de l'amplification d'un échantillon complexe. Ce biais est principalement dû à une hybridation stochastique des amorces lors des premières étapes de la MDA, conduisant à la sur-représentation de certaines régions génomiques (Dean et al., 2002; Hosono et al., 2003; Raghunathan et al., 2005; Zhang et al., 2006; Rodrigue et al., 2009; Woyke et al., 2009; Vaultot et al., 2012; Blainey, 2013). À partir d'un échantillon complexe, la structure des communautés peut ainsi être modifiée lors de la WGA, notamment lorsque que la quantité de matériel de départ est faible (Abulencia et al., 2006; Chen et al., 2008; Lepère et al., 2011). Des chimères provenant de réarrangements chromosomiques aspécifiques lors de l'étape d'amplification ont été rapportés (Zhang et al., 2006; Lasken et Stockwell, 2007; Chen et al., 2008). Ces chimères qui représentent entre 10% et 50% des produits synthétisés, sont des séquences hybrides formées par deux séquences qui n'étaient pas originellement liées (Zhang et al., 2006). Les progrès d'analyse *in silico* ont permis de s'affranchir partiellement des problèmes posés par le biais d'amplification pour la reconstruction de génomes (Rodrigue et al., 2009; Chitsaz et al., 2011; Swan et al., 2011; Bankevich et al., 2012). Un effort de séquençage supplémentaire pourrait constituer une des solutions possibles au problème de séquences chimériques (Rodrigue et al., 2009; Woyke et al., 2010).

### III.3.3 Apports de la métagénomique ciblée

#### III.3.3.1. *Obtenir de nouveaux génomes de référence*

La SCG a déjà permis d'obtenir un certain nombre de génomes partiels ou presque complets d'un certain nombre de microorganismes de diverses lignées taxonomiques. Le génome de deux nouvelles espèces de Crenarchaeota ont été obtenus à partir d'un échantillon de sol, après isolement en utilisant la technique du FISH (Kvist et al., 2007), ou de sédiments d'un environnement de faible salinité (Blainey et al., 2011). Un génome d'euryarchée provenant des environnements hypersalins a également été décrit (Ghai et al., 2011).

La technique a également été appliquée aux picoeucaryotes marins, permettant d'obtenir un génome partiel de picobiliphyte (Yoon et al., 2011), un phylum récemment décrit proche des cryptophytes (Not et al., 2007; Cuvelier et al., 2008) et récemment renommé Picozoa (Seenivasan et al., 2013). À défaut d'obtenir des génomes complets, deux autres études récentes ont utilisé la métagénomique ciblée sur des cellules picoeucaryotes, afin de mieux comprendre le rôle écologique des différents groupes qui constituent cette fraction de taille (Cuvelier et al., 2010; Vaultot et al., 2012).

Deux études ont permis de séquencer deux représentants d'un nouveau phylum bactérien, nommé TM7, pour lesquels aucun génome ni souche en culture n'étaient disponibles. La cytométrie en flux a été utilisée pour isoler les cellules individuelles d'un échantillon de sol, permettant d'obtenir un génome partiel (Podar et al., 2007). L'utilisation de la microfluidique a quant à elle permis d'obtenir le génome d'une bactérie affiliée au phylum OP11, largement représenté au sein des environnements terrestres et aquatiques (Youssef et al., 2011).

La connaissance du bactérioplancton marin a également été améliorée par cette approche avec la détermination des génomes de deux flavobactéries photohétérotrophes (Woyke et al., 2009), deux  $\lambda$ -protéobactéries (Chitsaz et al., 2011; Swan et al., 2011), une  $\gamma$ -protéobactérie (Swan et al., 2011) et le premier génome de *Beggiatoa* (Mußmann et al., 2007). La combinaison de la métagénomique et de la SCG a quant à elle permis de séquencer deux génomes de la  $\gamma$ -protéobactérie SAR86 (Dupont et al., 2012), un des clades les plus abondants dans l'environnement océanique, et un troisième génome de ce clade a récemment été obtenu par Rusch et al. (2013).

### III.3.3.2. *Interactions microbiennes in situ*

L'étude des interactions microbiennes, notamment des processus symbiotiques, reste bien souvent difficile à appréhender. La mise en culture des symbiontes pour des applications génomiques est très souvent impossible en l'absence de son hôte, compliquant les approches génomiques traditionnelles. Ce domaine est certainement celui où la SCG, en permettant de se focaliser sur les symbiontes, a démontré tout son potentiel.

Lors de l'isolement et du séquençage de cellule picobiliphytes, Yoon et collaborateurs ont ainsi également séquencé le génome d'un virus ainsi que deux génomes partiels de protéobactéries, probablement ingérées par ces cellules eucaryotes (Yoon et al., 2011). En utilisant la même approche pour le protiste *Paulinella chromatophora*, des contigs contenant de l'ADN cyanobactérien, provenant certainement d'une proie, ont été obtenus, ainsi que le génome partiel d'un cyanophage de *Synechococcus* (Bhattacharya et al., 2013).

Des interactions symbiotiques ou parasitiques ont également été identifiées entre divers protistes et bactéries (Martinez-Garcia et al., 2011a). Dans cette étude utilisant des SAGs (Single Amplified Genomes) et un criblage par PCR des gènes codant pour l'ARNr 16S et 18S, l'association entre *Candidatus Pelagibacter candidatus* et un protiste du groupe MAST-4, ainsi qu'entre une actinobactérie et un chrysophyte ont par exemple été détectés (Martinez-Garcia et al., 2011a). La SCG a aussi permis l'obtention de génomes du protiste *Candidatus Sulcia muelleri* DMIN, retrouvé en symbiose avec des insectes cicadelles (Woyke et al., 2010).

Enfin les interactions bactéries-virus peuvent à présent être étudiées, avec l'exemple des virus du coccolithophore *Emiliana huxleyi* (Martínez et al., 2011), ou des virus infectant les bactéries du tube digestif des termites (Tadmor et al., 2011).

### III.3.3.3. Diversité et fonction

De nombreuses études de SCG ont permis le couplage entre information phylogénétique et certains marqueurs métaboliques. Swan et collaborateurs ont ainsi mis pu attribuer un métabolisme chimiolithoautotrophe à des protéobactéries incultivées constituant une faction importante de la biomasse de l'océan mésopélagique (Swan et al., 2011). La découverte de tels métabolismes dans ces écosystèmes suggèrent un rôle important de ces organismes dans les cycles biogéochimiques globaux, notamment celui du carbone (Swan et al., 2011). L'absence d'ADN plastidial ou de gènes plastidiaux nucléaires a démontré que les picobiliphytes étaient en réalité hétérotrophes, expliquant ainsi leur position phylogénétique singulière par rapport aux autres lignées eucaryotes phytoplanctoniques originellement décrites (Not et al., 2007; Yoon et al., 2011).

La présence de gènes de protéorhodopsine a été confirmée au sein des différentes lignées marines bactériennes comme les flavobactéries et les  $\gamma$ -protéobactéries du clade SAR86 (Stepanauskas et Sieracki, 2007; Woyke et al., 2009; Dupont et al., 2012; Rusch et al., 2013). Une autre forme de rhodopsine a quant à elle été révélée au sein des génomes d'Actinobactéries, bactéries dominant les communautés bactériennes d'eau douce (Martinez-Garcia et al., 2011b; Garcia et al., 2013). Les  $\beta$ -proteobactéries du genre *Polynucleobacter* possèdent quant à elles les gènes impliqués dans la biosynthèse de bactériochlorophylle *a* (Martinez-Garcia et al., 2011b).

Les écotypes HLIII et HLIV de *Prochlorococcus* ne possèdent aucun représentant en culture et les connaissances sur les capacités métaboliques de ces cellules restent à ce jour limitées. Récemment, des gènes codant pour des sidérophores du fer ont été mis en évidence au sein de génomes très partiels de *Prochlorococcus* du clade HLIV, et pourraient expliquer le succès écologique de ces derniers dans les régions riches en nutriments mais pauvres en Chl car limitées en fer (zones HNLC; Rusch et al., 2010; Malmstrom et al., 2013). Notons par ailleurs qu'une région hypervariable contenant des gènes impliqués dans la modification de la paroi cellulaire, similaire à celles retrouvées chez les deux *Synechococcus* marins WH8102 et CC9605, ont été identifiés au sein du génome d'un représentant du sous-groupe III de SAR86 (Rusch et al., 2013). Ces gènes sont probablement impliqués dans les interactions avec les phages ou les brouteurs (Palenik et al., 2003; Palenik et al., 2006).

---

# OBJECTIFS

---



---

---

## OBJECTIFS

---

La compétition pour la lumière a conduit au développement d'une grande variété de pigments et protéines chromophorylées au sein de la radiation des *Synechococcus* marins. A ce jour, les études génomiques et physiologiques, notamment celles menées sur des souches en culture au sein de mon équipe d'accueil, ont permis de définir six types pigmentaires (Fig. 30). Cette diversité phénotypique, qui semble assez unique pour un groupe phylogénétiquement aussi restreint, pourrait expliquer la capacité des membres de ce genre à coloniser un grand nombre de niches lumineuses distinctes. *Synechococcus* est en effet retrouvé dans des environnements marins très divers, depuis les milieux eutrophes estuariens jusqu'à l'océan ouvert oligotrophe, où les conditions lumineuses varient largement, aussi bien d'un point de vue qualitatif que quantitatif. De plus, des études préliminaires menées *in situ*, utilisant des techniques basées sur l'étude des propriétés de fluorescence des cellules de *Synechococcus*, ont en effet suggéré que la pigmentation pourrait avoir un rôle sur la structure de ces communautés picoplanctoniques (Olson et al., 1988, 1990b; Wood et al., 1998). Néanmoins, contrairement à *Prochlorococcus*, les facteurs environnementaux affectant la structure des communautés de *Synechococcus* reste largement méconnus.

Les principaux objectifs de cette thèse ont été d'accroître les connaissances sur la diversité pigmentaire de *Synechococcus* afin de répondre aux questions suivantes :

- Quelle est l'étendue de la diversité pigmentaire présente dans le milieu naturel et quels sont les patrons de distribution de ces types pigmentaires ?
- Existe-t'il une diversité génétique et phénotypique au sein des souches en culture capables d'AC4 ?
- Quels sont les processus évolutifs associés à ce processus physiologique ?

Pour répondre à ces questions, différentes approches (génomique, métagénomique, physiologique) ont été utilisées au cours de cette thèse afin de d'étendre les travaux précédemment réalisés par l'équipe (Everroad et al., 2006; Six et al., 2007), à un nombre beaucoup plus grand de souches en culture ainsi qu'au milieu naturel.

Afin d'étudier la diversité et la distribution des différents types pigmentaires dans l'environnement marin, le premier objectif de cette thèse a été de développer une approche de métagénomique originale, combinant la cytométrie en flux, l'amplification de génomes complets ainsi que la fabrication de banques fosmidiques (voir **Chapitre II**). Elle nous a permis de cibler spécifiquement la région génomique PBS, où se situe la majorité des gènes impliqués dans la biosynthèse et la régulation des bras des PBS (voir §II.2.5). L'étude des gènes de cette région, dont la taille varie de 8.5 Kpb à 28.5 Kpb, permet en effet de déterminer le type pigmentaire des cellules sans aucune connaissance préalable de leurs propriétés spectrales et en s'affranchissant des méthodes culturales qui fournissent bien souvent une vision biaisée de la diversité existant dans le milieu naturel. Notre approche a été appliquée à différents échantillons provenant de divers environnements océaniques, dans le but (à terme) de séquencer environ 200 régions PBS.

Parmi les différents types pigmentaires de *Synechococcus*, certaines cellules sont capables de s'acclimater aux conditions fluctuantes de la qualité de lumière dans l'environnement marin par un processus appelé AC4. Ce phénomène a été beaucoup moins étudié que les processus d'AC de type II et III retrouvés chez les cyanobactéries d'eau douce et qui impliquent des mécanismes très différents (voir Chapitre I §II.3.2 et II.3.3). Au cours de cette thèse, j'ai participé à la purification et au séquençage complet de 25 nouvelles souches de *Synechococcus*, incluant notamment un certain nombre de souches capables d'AC4, dans le cadre des projets Génoscope METASYN et ANR PELICAN. Les régions PBS de ces différents génomes ont été annotées et des comparaisons génomiques, décrites dans le **chapitre III**, ont été menées afin d'identifier (1) les gènes impliqués dans ce processus physiologique et (2) de proposer des hypothèses quant aux processus évolutifs associés et (3) d'appréhender les conséquences d'un tel mécanisme sur la distribution des différentes lignées de *Synechococcus*. Enfin des caractérisations physiologiques de souches en cultures représentatives des différents géotypes ont été menées afin (4) d'évaluer la diversité phénotypique associée au processus d'AC4.

Les différents résultats obtenus au cours de cette thèse ainsi que les perspectives de recherche seront discutés dans le **chapitre IV**.

---

## CHAPITRE II

# ROLE DE LA DIVERSITÉ PIGMENTAIRE DANS L'ADAPTATION A LA NICHE CHEZ *SYNECHOCOCCUS*

---



---

---

# Article I : DIVERSITÉ, ABONDANCE ET DISTRIBUTION DES DIFFÉRENTS TYPES PIGMENTAIRES DANS L'ENVIRONNEMENT PAR UNE APPROCHE CIBLÉE DE METAGÉNOMIQUE

---

## I. CONTEXTE DU PROJET

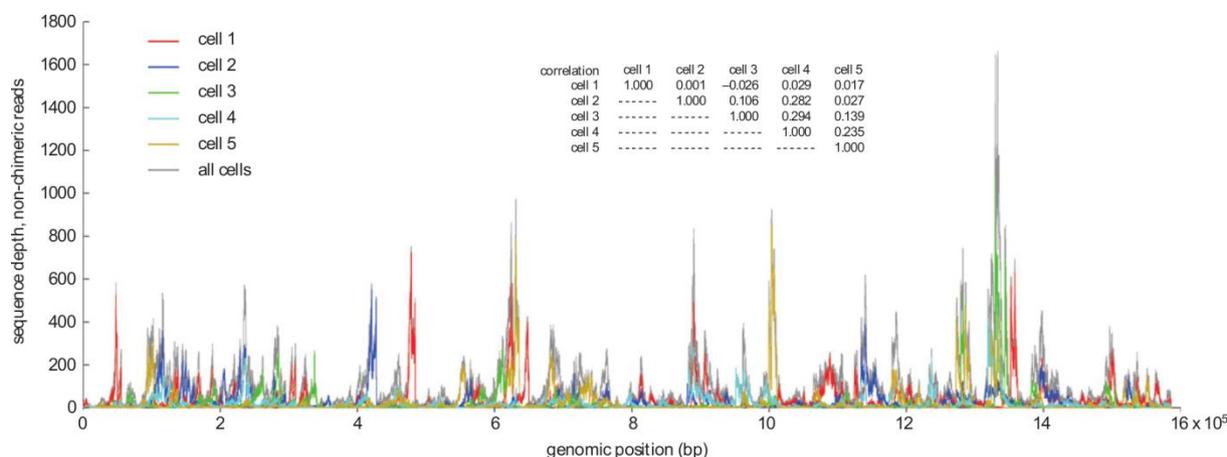
En dépit des nombreuses études de diversité menées sur *Synechococcus* jusqu'à présent, peu d'informations sont encore disponibles quant aux variables environnementales affectant la structure des communautés de *Synechococcus*. En effet, bien que les nutriments et la température semblent constituer des facteurs importants, notamment pour les clades dominants (I, II, III et IV; Zwirgmaier et al., 2008; Huang et al., 2012), ces paramètres n'expliquent cependant qu'une faible partie des variations observées dans la distribution de ces organismes. Une autre différence majeure entre les souches de *Synechococcus* qui pourrait influencer sur la structure des communautés *in situ* et expliquer l'ubiquité de ce genre cyanobactérien est sa grande diversité pigmentaire. *Synechococcus* est en effet retrouvé dans les environnements estuariens turbides aux eaux claires de l'océan ouvert. Les qualités spectrales de ces régimes aquatiques varient largement. Les longueurs d'onde de couleur orange/rouge dominant dans les régimes turbides alors que la couleur violette/bleue est prédominante dans les eaux claires, pauvres en nutriments. De manière intéressante, la pigmentation de ces souches apparaît totalement indépendante de la phylogénie obtenue à partir de l'ARNr 16S, ce qui pourrait en partie expliquer la difficulté de mise en relation entre la distribution des clades phylogénétiques et les paramètres physico-chimiques du milieu (Six et al., 2007b).

La distribution et l'abondance des différents types pigmentaires *in situ* de *Synechococcus* restent largement méconnus. La majorité des études menées s'attachant à étudier cette distribution en milieu marin ont utilisé des méthodes basées sur la détermination des propriétés optiques des cellules comme la cytométrie en flux (Olson et al., 1988; Olson et al., 1990b) ou la spectrofluorimétrie (Wood et al., 1998). Les conclusions de ces études proposent une domination des cellules riches en PUB dans les environnements oligotrophes alors que les cellules ne possédant pas ou peu ce pigment coloniseraient les milieux côtiers. Néanmoins ces études ne permettent pas

de distinguer la proportion relative de chacun des types pigmentaires composants ces communautés qui peuvent parfois être complexes.

Au cours de cette thèse, nous avons développé une approche de métagénomique originale ciblant spécifiquement la région PBS, rassemblant la majorité des gènes codant pour les protéines intervenant dans la régulation et la synthèse des projections radiaires des PBS, qui permet la détermination du type pigmentaire indépendamment de connaissances concernant les propriétés optiques des cellules. Une approche combinant des méthodologies sophistiquées comme la cytométrie en flux, la WGA et la fabrication de banques de fosmidés a été retenue. Les régions d'intérêt sont ensuite séquencées après criblage des différentes banques. L'utilisation combinée de l'ensemble de ces méthodologies, qui d'après nos connaissances n'a jamais été réalisée, peut induire un certain nombre de biais techniques qu'il est nécessaire de contrôler. Dans ce contexte, nous avons cherché à identifier et à contrôler au mieux ces biais technologiques intrinsèques, notamment les biais d'amplification survenant lors de la WGA fréquemment répertoriés dans la littérature (Fig. 49). Cette étude a permis l'écriture d'un premier article, dont je suis premier auteur, actuellement en préparation.

Cette approche a également été utilisée pour analyser la diversité pigmentaire au sein de régimes océaniques distincts. Dans cette optique, 12 banques fosmidiques ont été construites et criblées, mais par manque de temps, les résultats concernant le séquençage des fosmidés d'intérêt n'ont pu être obtenus avant la fin de cette thèse.



**Figure 49** : Biais d'amplification lors de la réaction d'amplification de génomes complets. Différentes réactions indépendantes de WGA ont été réalisées à partir de cellules de bactéries filamenteuses segmentées (SFB) et la couverture des différents génomes obtenus après assemblage *de novo* des séquences obtenues a été déterminée. D'après Blainey (2013).

## **II. MATÉRIELS ET MÉTHODES**

Un certain nombre d'optimisations ont été réalisées dans le cadre du développement de cette approche métagénomique. Au cours de ce paragraphe, les protocoles des étapes critiques seront développés. Les techniques couramment utilisées en laboratoire ne sont quant à elles décrites que dans la partie correspondante de l'article.

### **II.1. Stratégie d'échantillonnage et tri cellulaire**

#### **II.1.1 Collecte et concentration des échantillons environnementaux**

Différents échantillons naturels ont été prélevés au cours de différentes campagnes océanographiques et utilisés au cours de cette thèse afin d'étudier la distribution et l'abondance de ces types pigmentaires *in situ*. Dans un premier temps, les résultats concernant le premier article présenté dans ce chapitre ont été obtenus à partir d'un échantillonnage effectué au site ASTAN, situé au large de Roscoff (48° 46' 40 N, 3° 56' 15 W). Cet échantillon a été utilisé pour la majorité des mises au point techniques ainsi que les validations de notre approche méthodologique.

En complément de ce premier article, un échantillonnage à plus grande échelle a été réalisé. Ces échantillons proviennent de campagnes océanographiques distinctes et de régions océaniques présentant des conditions hydrologiques et trophiques distinctes. Ces échantillons ont principalement été utilisés afin d'augmenter les connaissances sur la diversité pigmentaire. Les trois stations longues (A, B, C) de la campagne océanographique BOUM (Biogeochemistry from the Oligotrophic to the Ultra-oligotrophic Mediterranean), qui s'est déroulée entre Juin et Juillet 2008 à bord de l'Atalante, ont été utilisées. Des informations complémentaires concernant la diversité des picocyanobactéries au cours de ce transect sont présentées en Annexe A. Des travaux complémentaires à cette campagne sont également disponibles dans un numéro spécial de la revue Biogeosciences à l'adresse [http://www.biogeosciences-discuss.net/special\\_issue63.html](http://www.biogeosciences-discuss.net/special_issue63.html). Des sites côtiers du bassin Méditerranéen ont également été échantillonnés au cours de l'été 2009 dans le cadre du programme d'échantillonnage SOMLIT (Service d'Observation du Milieu LITtoral; Goberville et al., 2010). Les données physico-chimiques correspondant à ces prélèvements sont disponibles à l'adresse suivante <http://somalit.epoc.u-bordeaux1.fr/fr/>. Deux autres stations de mer du Nord, échantillonnées par Dominique Marie au cours de la sortie 811 à bord du CEFAS Endeavour ont aussi été utilisées. Enfin, des échantillons de l'Océan Atlantique Nord, prélevés par Isabelle Mary au cours d'une campagne à bord du RSS Discovery (14 Juillet 2011 – 4 Août 2011) ont également été analysés.

Des échantillons d'eau de mer ont été prélevés pour la détermination des concentrations de picophytoplancton par cytométrie en flux. Ces derniers ont été fixés au glutaraldéhyde à une

concentration finale de 0.25% (Grade II, Sigmaaldrich, Saint-Louis, MO, USA), placés au noir pendant 10 min, puis congelés rapidement dans l'azote liquide avant conservation à -80°C jusqu'à analyse.

Pour le tri cellulaire des cellules du genre *Synechococcus*, les échantillons naturels ont été concentrés après préfiltration sur 10 et 3 µm en utilisant des filtres Nucleopore en polycarbonate (Millipore). Le filtrat a ensuite été concentré grâce à des unités de filtration Cell Trap de porosité 0.2 µm (MEM-TEQ Venture, Brighton, UK) ou par filtration tangentielle en utilisant des systèmes de porosité 100 kDa (TFF pour "Tangential Flow Filtration"; Sartorius Stedim Biotech). Du Pluronic F-68 (Sigma Aldrich), un surfactant permettant une meilleure cryopréservation des cellules, a été ajouté à l'échantillon à une concentration finale de 0,1%. Enfin, l'échantillon a été plongé dans l'azote liquide avant d'être stocké à -80°C.



*Un dispositif Cell-Trap permet de concentrer 2 à 10 L d'eau de mer dans un volume final de 1 à 3 mL alors que la TFF permet la concentration de 1 à 8 L d'eau de mer pour un volume final de 20 à 25 mL. La filtration tangentielle permet une meilleure préservation de l'intégrité cellulaire et présente des rendements beaucoup plus élevés que les dispositifs Cell-Trap.*

### II.1.2 Détermination des abondances picophytoplanctoniques par cytométrie en flux

La technique de cytométrie en flux consiste à faire passer des cellules, une à une devant un faisceau laser (excitation à 488 nm). Plusieurs paramètres sont analysés pour chaque cellule : des signaux de diffusion proportionnels à la taille (FSC : Forward Scatter / SSC : Side Scatter) et des signaux de fluorescence. Les différentes populations picoplanctoniques sont ainsi discriminées selon leur taille (SSC) et leur composition pigmentaire. En effet, les cellules de *Synechococcus* sont excitées à 488 nm, et la fluorescence orange émise par la PE et la fluorescence rouge de la Chl *a* sont enregistrés séparément (Marie et al., 1999). A cette même longueur d'onde d'excitation, les antennes photosynthétiques de *Prochlorococcus* émettent uniquement de la fluorescence rouge. Le marquage de l'ADN des cellules par un colorant approprié induit une fluorescence verte, qui permet d'obtenir des informations sur la quantité de matériel génétique présente dans chaque cellule. Ce marquage permet aussi de visualiser les bactéries hétérotrophes et de les discriminer des cyanobactéries, sur la base de leur différents rapports de fluorescence rouge/verte.

Les abondances respectives des différentes populations présentes dans l'échantillon naturel ont été mesurées grâce au cytomètre en flux FacsCANTO (Becton Dickinson Biosciences, SanJose, CA, USA), équipé d'un laser d'excitation à 488 nm, soit en mesurant la fluorescence naturelle des cellules, soit après marquage des cellules au SYBR-Green (Molecular Probes). Ce fluorochrome permet ainsi d'évaluer la concentration en bactéries hétérotrophes dans un échantillon donné et d'en déduire le

taux de contamination des cultures de *Synechococcus*. La détermination de la concentration en bactéries hétérotrophes ne peut s'effectuer que sur des échantillons fixés après marquage au SYBR-Green (Life technologies, Invitrogen) pendant 15 min au noir, à une concentration finale de 1X (Marie et al., 1999). En fonction de la concentration cellulaire, certains échantillons ont été dilués dans du milieu de culture PCRS11 stérile.

Les signaux de fluorescence et de diffraction sont normalisés en utilisant des billes fluorescentes de 0.95  $\mu\text{m}$  (Polysciences Warrington, PA, USA) en ajustant le gain de voltage avant chaque analyse. Les signaux de fluorescence rouge (FL3, un proxy de la concentration en Chl *a* et APC) sont détectés à travers un filtre passe-haut de 650nm, la fluorescence orange (FL2, proxy de la PE) et la fluorescence verte (FL1, contenu en ADN marqué au SYBR-Green) sont détectés en utilisant des filtres passe-bande de 530/30 nm (FL1) et 585/42 nm (FL2). Les résultats ont été analysés en utilisant le logiciel Cytowin v4.31 (Vaulot, 1989).

### **II.1.3 Tri de cellules du genre *Synechococcus* à partir d'échantillons naturels**

#### *II.1.3.1. Nettoyage du cytomètre et préparation d'expérience*

Les échantillons naturels ont été triés en utilisant un trieur FacsARIA (Becton Dickinson Biosciences, SanJose, CA, USA). Afin d'éviter la contamination des échantillons par toute trace d'ADN non spécifique, de grandes précautions ont été prises quant à la préparation et au nettoyage de l'appareil. Le cytomètre a d'abord été nettoyé par une succession de solution de lavage Helmanex 2% (Hellmanex® II, Hellma GmbH & Co. KG, Germany) dans de l'eau chaude, durant une période minimale de 2 h. L'appareil a ensuite été rincé sur la nuit en utilisant de l'eau milliQ. Les capillaires d'injection de l'échantillon ont été autoclavés et la chambre d'injection nettoyée en utilisant la solution de lavage. La chambre d'injection et les capillaires ont été nettoyés 5 min entre chaque tri (Cleaning concentrate, Biorad), puis rincés à l'eau milliQ pendant 10 min.

Du tampon PBS (Phosphate Buffer Saline; 137 mM NaCl, 2.7 mM KCl, 10 mM  $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ , 2 mM  $\text{KH}_2\text{PO}_4$ , pH 7.5) a été utilisé comme liquide de gaine. Le tampon a été préparé par dissolution de sels calcinés (4 h à 450°C) dans des bouteilles de polycarbonate dédiées. Le tampon a ensuite été stérilisé aux UV durant 2 h sous une hotte microbiologique, puis autoclavé 20 min à 121°C. Le tampon a été filtré extemporanément en utilisant des unités Stericup (Millipore, Molsheim, France) et transférés dans le réservoir de liquide de gaine, également préalablement stérilisé aux UV.

L'ensemble des tubes utilisés pour la cytométrie et la récupération des cellules ont été autoclavés et stérilisés aux rayons UV avant utilisation.



*Au cours de cette thèse d'autres tampons ont été utilisés comme liquide de gaine et leur compatibilité avec l'amplification de génomes complets (WGA) a été testée. Aucune amplification n'a pu être obtenue en utilisant du NaCl 0.1% ou du PCRS11 (milieu de culture utilisé pour la culture de Synechococcus, probablement en raison de la forte teneur en sels de ces tampons). L'utilisation de TE (Tris/EDTA) est possible mais reste moins efficace, en raison de la dégradation rapide de certaines cellules de Synechococcus observée par cytométrie en flux dans ce tampon.*

### II.1.3.2. Tri cellulaire

Les cellules du genre *Synechococcus* ont été triées à partir d'échantillons naturels concentrés. Afin d'obtenir des échantillons triés avec une grande pureté, trois tris successifs ont été réalisés avec un trieur en mode "Purity", en veillant à ne pas dépasser plus de 1000 évènements/sec. Le tri cellulaire a été effectué à une pression de 70 psi, une fréquence de 90 000 Hz et en utilisant une buse de 70  $\mu\text{m}$ . Ces paramètres déterminent la concentration des cellules après tri qui, dans ces conditions, était de 500 cellules. $\mu\text{L}^{-1}$ . Le premier tri a été effectué en se basant uniquement sur la fluorescence naturelle des cellules en utilisant le mode « purity » (Yield Mask = 32, Purity Mask = 32, Phase Mask = 0) et en recueillant les signaux de fluorescence rouge (Chl) et de taille (SSC). Les deuxième et troisième tris ont été réalisés après marquage de l'échantillon au SYTO-9 (Invitrogen S-34854) et SYTO-13 (Invitrogen S-7575), à une concentration finale de 5  $\mu\text{M}$ . Ces deux molécules, qui sont des marqueurs de l'ADN ayant la propriété de pénétrer dans les cellules dont la membrane plasmique est intacte, permettent donc de détecter les bactéries hétérotrophes dépourvues de pigments. Après 15 min de marquage, les cellules sont triées en utilisant le mode « purity » ou préférentiellement le mode « single-cell » (Yield Mask = 0, Purity Mask = 32, Phase Mask = 16). Le tri en mode "single-cell" ne peut-être effectué que si un grand nombre de cellules est disponible dans l'échantillon de départ. La contamination des échantillons triés est évaluée après chaque tri par l'analyse, à l'aide du FACSCanto, d'un échantillon marqué au SYBR-Green (voir ci-dessus). Après tri, les cellules sont brièvement centrifugées et stockées à 4°C avant de procéder à une réaction de WGA ou une extraction d'ADN.

## II.2. Amplification de génomes complets à partir d'échantillons triés

Les réactions de WGA ont été réalisées en utilisant le kit GenomiPhi V2 (GE Healthcare, Waukesha, WI). Les tampons de lyse (400 mM KOH, 100 mM DTT, 10 mM EDTA) et de neutralisation (400 mM HCl, 600 mM Tris-HCl, pH 7.5) ont été filtrés sur 0.2  $\mu\text{m}$  et stérilisés aux UV pendant 30 min juste avant WGA. L'ensemble des échantillons a été traité suivant les recommandations du fournisseur en utilisant une lyse chimique qui permet en évitant de fractionner l'ADN de favoriser l'amplification de produits de grande taille par rapport à d'une lyse thermique. Brièvement, 5  $\mu\text{L}$  de cellules triées (correspondant à environ 2500 cellules dans nos conditions de tri) ont été lysées en

ajoutant 1  $\mu\text{L}$  de tampon de lyse et incubés 10 min sur glace. Les échantillons ont ensuite été neutralisés avec 1  $\mu\text{L}$  de tampon de neutralisation. Les réactions ont été réalisées dans un volume final de 24  $\mu\text{L}$  après ajout du tampon de resuspension (« sample buffer », 7  $\mu\text{L}$ ), de tampon de réaction (9  $\mu\text{L}$ ) et de l'enzyme phi29 (1  $\mu\text{L}$ ). Les mélanges réactionnels ont été incubés 2 h 30 à 30°C avant d'inactiver l'enzyme à 65°C durant 5 min en utilisant un GeneAmp PCR system 9700 (Applied Biosystems™). Les produits ont ensuite été conservés à -20°C jusqu'à analyse.



*Ces réactions sont très sensibles à toute trace de contamination. De grandes précautions ont été prises lors de ces expériences. L'ensemble des manipulations ont été réalisées sous une hotte PCR HEPA équipé de néons UV, rigoureusement nettoyée et stérilisée avant chaque usage. Un lot de pipettes a également été dédié à cet usage.*



*Un comparatif sur les biais d'amplification a été mené sur les kits Genomiphi v2, ainsi que sur le kit Repli-G mini kit (Qiagen). Le biais d'amplification concernant les différents génotypes de Synechococcus, ainsi que le biais d'amplification entre Synechococcus et les bactéries hétérotrophes a alors été évalué. Ce tests nous ont permis de mettre en évidence que le biais d'amplification du kit Repli-G mini kit était plus important que celui rencontré en utilisant le kit GenomiPhi v2, notamment lorsque les réactions d'amplification étaient réalisées à partir d'un faible nombre de cellules. Néanmoins ces tests ont aussi mis en évidence que les fragments amplifiés obtenus à partir du kit Repli-G étaient de plus grande taille que ceux obtenus avec GenomiPhi V2, et ce en utilisant dans les deux cas une lyse chimique. D'autre part si le rendement obtenu avec GenomiPhi se situe autour de 500 ng d'ADN final par réaction, celui-ci était 5 fois supérieur en utilisant le kit Repli-G.*

### **II.3. Fabrication de bibliothèques de fosmides**

#### **II.3.1 Traitement enzymatique à la Nuclease S1**

La WGA utilisant l'enzyme phi29 conduit à la génération de structures branchées, où les différents brins d'ADN sont interconnectés les uns aux autres. Afin de fabriquer des bibliothèques de fosmides à partir de produits WGA, il est nécessaire de procéder à la suppression de ces jonctions entre les différentes molécules d'ADN branchées.

Les bibliothèques de fosmides ont été fabriquées à partir d'un volume maximum de 160  $\mu\text{L}$  de produits WGA (correspondant généralement au pool de 7 ou 8 réactions indépendantes). Le traitement enzymatique a été effectué en ajoutant 1000 U de Nuclease S1 (Promega), dans un volume réactionnel final de 200  $\mu\text{L}$ . La solution a été incubée 1 h à 37°C puis stoppée 10 min à 70°C en ajoutant un volume d'EDTA 0.5 M pH 8.0. L'ADN a ensuite été purifié en ajoutant du réactif et du colorant Sureclean (Bioline France, Paris, France) et la réaction a été incubée au moins 30 min à température ambiante. Le surnageant a été éliminé après centrifugation 10 min à 16 000 g et le culot a été lavé deux fois par ajout d'éthanol 70% froid. Après élimination des traces résiduelles d'éthanol par séchage au Speed-Vac, l'ADN a été réhydraté dans 40  $\mu\text{L}$  d'Eau Baxter stérile, sous agitation orbitale à 500 rpm et à 4°C pendant une nuit.

### II.3.2 Sélection des fragments d'intérêt par PFGE

Les fragments d'intérêt, de taille supérieure à 25 Kpb, ont été sélectionnés par PFGE (CHEF DR-II BioRad, Hercules, CA) sur gel d'agarose Low-Melting Point 1% (Invitrogen) en tampon TBE 0.5X, après 18 h de migration à 200 V, avec des paramètres de pulses initiaux et finaux de 0.5 s et 1.5 s respectivement. Les bandes externes, contenant du marqueur de taille GeneRuler™ High Range DNA Ladder (Fermentas, Villebon sur Yvette, France), ont été utilisées pour déterminer la région du gel à exciser après marquage au bromure d'éthidium. Durant cette étape, uniquement les bandes externes ont été exposées aux rayons UV afin de ne pas diminuer l'efficacité de clonage. Après fusion de l'agarose par incubation à 70°C durant 15 min, et transfert à 45°C durant 10 min, celui-ci a été digéré durant 5 h à 45°C en ajoutant un volume de 1 µL de GELase (Epicentre) pour 100 mg d'agarose. La réaction a été stoppée en inactivant l'enzyme 10 min à 70°C. Afin de précipiter les éventuelles traces d'agarose non digéré, la solution a été placée sur glace au moins 15 min, puis centrifugée à 7500 g, 4°C durant 20 min. Le surnageant, contenant l'ADN, a ensuite été purifié en utilisant des Amicon Ultra-2 (porosité 30 kDa, Millipore, Molsheim, France), préalablement rincés avec de l'eau Baxter stérile, par centrifugation à 3000 g, 4°C durant 20 min. Cette étape a été répétée 3 fois, et le temps de la dernière centrifugation a été ajusté afin d'obtenir un volume final de 48 µL. La quantité d'ADN a ensuite été évaluée au Nanodrop ND-100 (ThermoScientific), et la qualité a été vérifiée sur gel d'agarose 1%.



*Cette étape conditionne directement la qualité des banques qui seront obtenues. Il est important d'éviter toute exposition aux UV des produits de WGA lors de l'excision des fragments ainsi que de respecter les étapes de purification.*

### II.3.3 Traitement enzymatique

Le traitement enzymatique de réparation des extrémités des brins d'ADN a été réalisé suivant les recommandations du fournisseur (Epicentre). Une légère modification a cependant été apportée en ajoutant 40 U d'ADN polymérase I (New England Biolabs, Evry, France) au mix enzymatique fourni par le fournisseur, et la réaction a été incubée à 25° durant 3 h. L'ADN a ensuite été purifié en utilisant des Amicon Ultra 0.5 (porosité 30 kDa, Millipore, Molsheim, France), à 7000 g, 4°C durant 10 min. Cette étape a été répétée, et le temps de centrifugation a été ajusté afin d'obtenir un volume final le plus faible possible, généralement autour de 30 µL. La quantité d'ADN a ensuite été évaluée au Nanodrop ND-100 (ThermoScientific), et la qualité a été vérifiée sur gel d'agarose 1%.

### II.3.4 Ligation, empaquetage et titrage

La réaction de ligation a été réalisée en utilisant le kit CopyControl Fosmid Library Production Kit (EPICENTRE) suivant les recommandations du fournisseur., i.e. en incubant approximativement 300 ng d'ADN et le vecteur pCC2FOS™ pendant 16 h à 10°C, puis en inactivant l'enzyme 10 min à 70°C. L'empaquetage a été réalisé suivant les recommandations du fournisseur, et la solution a ensuite été diluée dans 100 µL de tampon Phage Dilution Buffer. La solution de phages a ensuite été diluée au 1/10<sup>ème</sup> et 1/100<sup>ème</sup> et 10 µL de chacune de ces dilutions ont été utilisés pour infecter une culture d'*Escherichia coli* EPI300-T1R, en milieu LB + MgSO<sub>4</sub> 10 mM + Maltose 0.2%, dont la DO est comprise entre 0.8 et 1.0. Après transfection 1 h à 37°C sous agitation orbitale (80 rpm), la culture transfectée a été étalée sur gélose LB + 12.5 µg.mL<sup>-1</sup> de chloramphenicol. Le titre a été déterminé à partir de 3 réplicats par dilution.

En fonction des résultats de titrage, la totalité des phages ont été utilisés pour transfecter une culture d'*E. coli* afin d'obtenir un titre final d'environ 3000 cfu.mL<sup>-1</sup>. Après transfection, du glycérol a été ajouté à une concentration finale de 20%, et la culture transfectée a été conservée à -80°C après congélation dans de l'azote liquide. Avant congélation, 3 aliquots de 50 µL de culture ont été étalés sur gélose LB + Cm afin de déterminer le titre final.



*L'étape critique de la réaction de transfection est le respect de la DO de la culture d'E. coli.*

### II.3.5 Détermination de la taille des inserts

Afin d'obtenir des quantités d'ADN suffisantes, les clones d'intérêt ont été induits à forte copie, selon les recommandations du fournisseur, en utilisant la solution d'induction fournie. Les fosmides ont été extraits en utilisant le kit NucleoSpin Plasmid (Macherey-Nagel, Hoerdt, France) et quantifiés au Nanodrop. La taille des inserts a été déterminée par PFGE après digestion par *NotI* selon les recommandations du fournisseur.

## III. ARTICLE

## Use of targeted metagenomic approach to study in situ diversity of pigment types in marine *Synechococcus*

Florian Humily<sup>1,2</sup>, Dominique Marie<sup>1,2</sup>, Gregory Farrant<sup>1,2</sup>, Frédéric Partensky<sup>1,2</sup>, Sophie Mazard<sup>3,4</sup>,  
Morgan Perennou<sup>5</sup>, Karine Labadie<sup>6</sup>, Jean-Marc Aury<sup>6</sup>, Patrick Wincker<sup>6</sup>, Audrey Nicolas<sup>1,2</sup>, David J.  
Scanlan<sup>3</sup> and Laurence Garczarek<sup>1,2\*</sup>

<sup>1</sup> UPMC-Université Paris VI, Station Biologique, CS 90074, 29688 Roscoff cedex, France

<sup>2</sup> CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment, Oceanic Plankton group,  
Marine Phototrophic Prokaryotes team, Place Georges Teissier, CS 90074, 29688 Roscoff cedex,  
France

<sup>3</sup> School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.

<sup>4</sup> Department of Chemistry and Biomolecular Science, Macquarie University, North Ryde, NSW 2109,  
Australia

<sup>5</sup> CNRS, FR2424, Service Information et Génomique, Station Biologique, 29688 Roscoff cedex, France

<sup>6</sup> Commissariat à l'Énergie Atomique, Institut de Génomique/Génomoscope, 91000 Evry, France

### IN PREPARATION

\*For correspondence. E-mail [laurence.garczarek@sb-roscoff.fr](mailto:laurence.garczarek@sb-roscoff.fr); Tel. (+33) 2 9829 2538; Fax (+33) 2 9829 2324.

Keywords : marine cyanobacteria, phycobilisomes, whole genome amplification (WGA), flow cytometry cell sorting, fosmid library

In preparation for FEMS Microbiology Ecology

**ABSTRACT**

Marine *Synechococcus* cells, the second most abundant oxygenic phototroph on Earth, display a wide pigment diversity. However, the distribution and diversity of these various pigment types in oceanic regimes remains largely misunderstood. In this study, we have combined flow cytometry cell sorting, whole-genome amplification (WGA) and fosmid library to specifically target a genomic region encoding phycobilisome (PBS) rods. This multi-step metagenomic approach induces a number of methodological biases, notably due to the WGA step, which can affect the quality and representativeness of sequences with regard to the initial environmental sample. To detect and better estimate the effect of these biases, we performed a characterization of the community composition at each step of the procedure using terminal restriction fragment length polymorphism using as genetic markers, the *petB* gene for *Synechococcus* and the 16S rRNA for the whole bacteria community. This approach allowed us to select reliable WGA products to construct a fosmid library from a natural sample collected at a coastal site off Roscoff (France). Sequencing of 25 fosmids containing the targeted region allowed us to assemble partial PBS regions with similar gene content and organization to cultured isolates. Altogether, this study provides novel insights about *Synechococcus* community structure and pigment types diversity at a representative coastal station of the English channel.

## INTRODUCTION

An open question in the field of microbiology is whether cultured isolates are valid representatives of microbial diversity present in the environment. During the last decade, the development of culture-independent approaches, associated with high throughput sequencing, has revolutionized studies in microbial ecology (Gilbert and Dupont, 2011). Environmental genomics (also called metagenomics) indeed allows scientists to bypass the need for isolation and cultivation steps, which are tedious and often unsuitable for a large number of microorganisms, notably in the marine environment (Venter et al., 2004; Rusch et al., 2007; Biers et al., 2009). The outcome and numerous recent improvements of new sequencing technologies, coupled with reduced costs, have provided access to extensive gene content information of the different members of natural communities and even sometimes allowed to reconstruct entire metabolic pathways or whole genomes (Tyson et al., 2004; Tringe and Rubin, 2005; Martín et al., 2006; Wooley et al., 2010). However, metagenomics often fails to link diversity information of specific microbes to their specific ecological functions, especially in complex environments, in which only the most abundant biosphere is generally considered (Pedrós-Alió, 2012).

To characterize the genetic repertoire of a particular group, reducing the sample complexity appears to be an appropriate strategy. Analysis of sub-communities exhibiting lower diversity, such as specific taxa or populations (Kalyuzhnaya et al., 2008) or even single cells (for review see Stepanauskas, 2012), is now possible by an approach termed “targeted enrichment” (Hallam et al., 2006). The latter consists in focusing on a particular cell population, based on its specific phenotypic or spectral characteristics. In marine ecology, fluorescence-activated cell sorting by flow cytometry has become the most popular isolation method, allowing high-throughput cell separation based on various parameters such as light scatter and induced or natural fluorescence (Sekar et al., 2004; Stepanauskas and Sieracki, 2007; Woyke et al., 2009; Yoon et al., 2011). After enrichment, sequencing effort can also be directed towards specific genes or metabolic pathways (i.e. targeted metagenomics), selected using different screening methods (for a review see Suenaga, 2012). However, these methodologies often imply to work with a small cell number, and thus low DNA amounts, which is generally insufficient for current sequencing procedures, although the recent advent of single-molecule sequencing technologies (third-generation) opens new opportunities to solve this problem (Schadt et al., 2010). This limitation is currently circumvented using whole-genome amplification (WGA), enabling to obtain enough DNA for sequencing from a very low amount of DNA or cells (Abulencia et al., 2006; Podar et al., 2007; Chen et al., 2008; Lepere et al., 2011) or even single cells (Zhang et al., 2006; Rodrigue et al., 2009). Multiple displacement amplification (MDA) emerged as the reference technique, due to the high processivity and high

strand displacement activity of the phi29 enzyme (Dean et al., 2001). Despite significant advances in recent years, the most challenging problem of this approach remains amplification biases, which can severely alter our perception of the community structure by differentially amplifying certain taxa with regard to others, as evidenced for soil bacteria (Abulencia et al., 2006). Furthermore, WGA-induced chimeras may also affect results from whole genome sequencing, *de novo* assembly and DNA libraries preparation (Lasken and Stockwell, 2007; Chen et al., 2008).

*Synechococcus* is the second most abundant single-celled cyanobacterium dominating marine ecosystems (Scanlan et al., 2009) and plays a key role in global carbon cycle (Li, 1994; Richardson and Jackson, 2007; Buitenhuis et al., 2012). This oxygenic photoautotroph, which is found in the upper layer across large environmental gradients from coastal waters to open-ocean, exhibits a high genetic diversity (Zwirgmaier et al., 2008; Huang et al., 2012; Mazard et al., 2012). Among numerous clades unveiled using various genetic markers, clades I to IV constitute the most abundant ones, with clades I and IV prevailing in nutrients-rich coastal waters, clade II in subtropical/tropical waters and clade III in oligotrophic regimes (Zwirgmaier et al., 2008; Mella-Flores et al., 2011; Mazard et al., 2012). A leading hypothesis to explain the dominance of *Synechococcus* across large oceanic regimes is the considerable pigment diversity found within this genus (Six et al., 2007). As the majority of cyanobacteria, *Synechococcus* uses a pigment-protein macrocomplex, called the phycobilisome (PBS), composed of phycobiliproteins, as its main light-harvesting system. Three main pigment types were defined based on the major phycobiliprotein found in PBS rods : type 1 contains only phycocyanin (PC), type 2 PC and phycoerythrin I (PEI) and type 3 PC, PEI and PEII (Six et al., 2007). The latter type has been subdivided into 5 sub-types based on the ratio of the two phycobilins, phycourobilin (PUB,  $A_{\max} = 495 \text{ nm}$ ) or phycoerythrobilin (PEB,  $A_{\max} = 545 \text{ nm}$ ), linked to these phycobiliproteins (Humily et al., submitted). This ratio, generally assessed by the  $EX_{495:545\text{nm}}$  proxy (i.e. the fluorescence excitation ratio of these two chromophores), can be low (3a), medium (3b), high (3c) or variable (3d-e). The latter sub-type corresponds to cells performing type IV chromatic acclimation (CA4) that are able to match their pigmentation with incident light (Palenik, 2001; Everroad and Wood, 2006; Humily et al., submitted). Furthermore, these phenotypes were recently complemented by adding a suffix (A or B) to indicate the occurrence of one of the two CA4 island types, CA4-A or CA4-B, in their genomes. These different pigment types also differ in their content in genes involved in the synthesis and regulation of the PBS rods, which are gathered into a specific genomic region ranging in size between 9 and 30 kbp. Although this pigment diversity likely allows these organisms to colonize the large variety of light environments existing in the marine ecosystem, from green turbid coastal to blue oligotrophic waters (Olson et al., 1988; Olson et al., 1990; Partensky et al., 1999), the influence of pigmentation on the niche partitioning of *Synechococcus* spp. remains poorly understood.

In this study, we have developed a targeted metagenomics approach to specifically clone and sequence PBS genomic regions, the gene content of which is sufficient to predict cell pigmentation from natural *Synechococcus* populations, without prior knowledge of their spectral properties. This method, which consists in a unique combination of cell sorting by flow cytometry, WGA as well as fosmid library construction is applicable to the study of genetic diversity of any large DNA fragments from uncultured *Synechococcus* cells.

## MATERIALS AND METHODS

### *Sample collection and cell sorting*

A seawater surface sample (8L) was collected at the Service d'Observation en Milieu Littoral (SOMLIT)-Astan station off Roscoff (Brittany, France), a representative site of the western English channel, using a Niskin bottle mounted on a CTD frame (Table 1). The sample was pre-filtered through 5 µm polycarbonate filters (Millipore, Molsheim, France) to remove large planktonic cells and particles, then concentrated by tangential filtration on a 100 000 MWCO membrane as previously described (Marie et al., 2010). Picophytoplanktonic cell concentrations were determined by flow cytometry using a FACS Canto (Becton Dickinson, San Jose, CA) as detailed previously (Marie et al., 1999).

**TABLE 1.** Physico-chemical parameters and phytoplankton abundance at the SOMLIT-Astan site at the time of sampling (May 29, 2012 at 11:35 am).

Parameter	Value
Latitude	48° 46' 40 N
Longitude	3° 56' 15 W
Sample depth (m)	1
Water column depth (m)	60
Temperature (°C)	12.55
Salinity (psu)	35.24
Chl <i>a</i> (µg.L <sup>-1</sup> )	0.60
NH <sub>4</sub> (µM)	1.40
NO <sub>3</sub> (µM)	1.10
NO <sub>2</sub> (µM)	0.17
PO <sub>4</sub> (µM)	0.22
<i>Synechococcus</i> (cells.ml <sup>-1</sup> )	841
Photosynthetic eukaryotes (cells.ml <sup>-1</sup> )	3441

*Synechococcus* cells were sorted using a FACS Aria flow cytometer (Becton Dickinson, San Jose, CA). In order to prevent sample contamination, the flow cytometer was cleaned by a succession of 2% bleach solution (Hellmanex® II, Hellma GmbH & Co. KG, Germany) in warm water, during at least 2 h and overnight rinsing with milliQ water, as previously described (Stepanauskas and Sieracki, 2007). Phosphate Buffer Saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub> · 2 H<sub>2</sub>O, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5) was used as sheath fluid, with a pressure of 70 p.s.i., for cell sorting. The phosphate buffer was prepared by dissolving combusted (4 h at 450°C) salts in milliQ water. It was then UV-treated for 2 h, autoclaved at 121°C for 20 min and filtered on 0.2 µm before transfer into a UV-treated tank. Prior to sorting, a 0.2 µm pore size Sterivex® filter (Millipore, Molsheim, France) was connected on the sheath fluid line injection, just before the chamber.

To minimize contamination by heterotrophic cells, a three-step sorting procedure of *Synechococcus* cells was used. Cells were first sorted based on their natural autofluorescence using

the “purity” mode and the signal was triggered on the red fluorescence from chlorophyll and side scatter. Then, heterotrophic bacteria and picoeukaryotes were screened out after DNA staining for 15 min using a combination of two nucleic acid stains, SYTO-9 and SYTO-13, at a final concentration of 5  $\mu\text{M}$  each (Invitrogen, Carlsbad, CA; del Giorgio et al., 1996; Lebaron et al., 1998; Gasol et al., 1999). Between each sorted samples, the sample injection and chamber lines were cleaned for 10 min using an alkaline detergent (BioRad, Hercules, CA) and rinsed for at least 10 min with autoclaved milliQ-water. Sorted cells were collected in UV-treated Eppendorf tubes, quickly centrifuged, and stored at 4°C until WGA or DNA extraction.

#### *Whole genome amplification*

WGA reactions were carried out under an UV3 HEPA PCR Workstation (UVP, Cambridge, UK) using Genomiphi v2 kit (GE Healthcare, Waukesha, WI). Lysis (400 mM KOH, 100 mM DTT, 10 mM EDTA) and neutralization buffers (400 mM HCl, 600 mM Tris-HCl, pH 7.5) were filtrated on 0.2  $\mu\text{m}$  and UV-treat 30 min just before WGA. All samples were denatured using a chemical procedure that we found to be more suitable than thermal denaturation to obtain large size amplification products, necessary for fosmid library construction (Fig. S1). Approximately 2500 *Synechococcus* sorted cells (corresponding to 5  $\mu\text{L}$  in our sorting conditions) were lysed by adding 1  $\mu\text{L}$  of lysis buffer, incubated for 10 min on ice before addition of 1  $\mu\text{L}$  neutralization buffer. WGA reactions were carried out in 24  $\mu\text{L}$  final volume by adding sample buffer (7  $\mu\text{L}$ ), reaction buffer (9  $\mu\text{L}$ ) and phi29 enzyme (1  $\mu\text{L}$ ), then incubated for 2.5 h at 30°C before inactivating the enzyme by incubation for 5 min at 65 °C. Blank controls with sterile water and sheath fluid collected at the exit of the chamber were taken for each experiment. For PCR reactions, WGA products were diluted 10-fold in sterile MilliQ water. WGA products and dilutions were stored at -20°C until processing.

#### *PCR amplification and T-RFLP screening*

Terminal-restriction fragment length polymorphism (T-RFLP) was used to assess the purity of sorting and amplification biases. While the *petB* gene, coding for the cytochrome  $b_6$  subunit of the cytochrome  $b_6/f$  complex, allowed us to evaluate the genetic diversity of *Synechococcus* (Mazard et al., 2012), the small subunit (16S) rRNA gene was used to check contamination by heterotrophic bacteria. Genomic DNA was extracted from concentrated natural samples (~ 40 000 *Synechococcus* cells) and sorted samples (~ 50 000- 100 000 cells) using a modified protocol of the DNA Blood and Tissue kit, as previously described (Qiagen, Courtaboeuf, France; Balzano et al., 2012). All PCR reactions for T-RFLP screening were performed in 25  $\mu\text{L}$  volume containing 1.25 U GoTaq *polymerase*

(Promega, Madison, WI), 1X buffer provided with the polymerase and 1.5 mM MgCl<sub>2</sub>. Primers and PCR conditions are listed in Table 2. Forward primers 16S\_27F and petBF were 5'-labeled with 6-carboxyfluorescein (6-FAM, Eurogentec, Seraing, Belgium) for fragments detection by T-RFLP. All amplifications were performed using a GeneAmp PCR system 9700 (Applied Biosystems, Carlsbad, California) and the program consisted of an initial denaturation step of 5 min at 94°C, followed by 30 cycles (30 s at 94°C, 30 sec at 55°C and 1 min at 72°C) and a final extension step of 10 min at 72°C (Table 2). In order to assign terminal-restriction fragments (T-RFs) of marine *Synechococcus* 16S rRNA and *petB*, T-RF databases were generated using sequences from natural populations and cultured isolates (Mazard et al., 2012) that were virtually digested *in silico* using various restriction enzymes (Tables S1 and S2). While for 16S rRNA, we used *MspI* as previously described (Mazard et al., 2011), for *petB*, this database allowed us to identify *HpaII/HinfI* as the most appropriate set of restriction enzymes since it allowed us to obtain up to eight T-RFs (Table S2). Both amplicon types were digested for 3 h at 37°C, using manufacturer recommendations (New England Biolabs, Evry, France).

**TABLE 2. List of primers used in this study**

Primer name	Sequence (5' to 3')	Targeted gene/operon	Use	Product	Amplicon length <sup>a</sup>	Tm (°C)	dNTPs (μM)	Primers conc. (nM)	# cycles	References
16S_27F <sup>b</sup> 16S_1492R	AGAGTTTGATCMTGGCTCAG TACGGYTACCTGTACGACTT	16S rRNA	T-RFLP	16S rRNA Eubacteria	1448	55	100	400	30	Field et al. (1998) Lane (1991)
petBF <sup>b</sup> petBR	TACGACTGGTTCCAGGAACG GAAGTGCATGAGCATGAA	<i>petB</i>	T-RFLP/Pyrosequencing	b <sub>6</sub> subunit of cytochrome <i>b6/f</i>	597	55	250	1 μM	30 28-30 <sup>c</sup>	Mazard et al. (2012) Mazard et al. (2012)
mpeB_220F mpeA_263R	TACACCAACCGCAARATGGC ACRAAGTCRCGCTTGCACTT	<i>mpeBA</i>	Screening (hybridization/PCR)	PEII α- and β- subunits	626	55	300	500	30	This study This study
SynB1F SynA3R	ATGCTCGACGCATTCTCCMG CGGTGGTGAYGACGGAYTTCAT	<i>cpeBA</i>	Pyrosequencing	PEI α- and β- subunits	635	55	250	500	30-35 <sup>c</sup>	Everroad et al. (2006) Everroad et al. (2006)
<i>c/rpcB</i> _244F SynpcA-Rev	TGCCTGCGGCATGGAGATC ATCTGGGTGGTGAGGG	<i>r/cpcBA</i>	Pyrosequencing Screening (hybridization/PCR)	PC α- and β- subunits	523	50	250	500	30-40 <sup>c</sup> 30	This study Haverkamp et al. (2008)
<i>cpeC</i> _187F <i>cpeC</i> _832R	TTCAAGCTYGGTGARATCAG TGAAYTGCTCNGAGAGYTTG	<i>cpeC</i>	Screening (hybridization/PCR)	PE associated linker	645	55	300	500	35	This study This study
<i>cpeU</i> _224F <i>cpeU</i> _452R	GATTTTGGTGGGARAGYAA ACAAACCARCAKCGYTCDAT	<i>cpeU</i>	Screening (PCR)	Putative phycobilin lyase	228	55	30	500	30	This study This study

<sup>a</sup> Amplicon length was determined on *Synechococcus* sp. WH8102 (GenBank accession number BX548020).

<sup>b</sup> 16S rRNA and *petB* forward primers used for T-RFLP were 5'-labeled with 6-carboxyfluorescein (6-FAM)

<sup>c</sup> Number of cycles was adjusted to obtain a faint band on agarose gel for pyrosequencing

After inactivation of the enzyme at 80°C for 15 min, terminal-restriction fragments (T-RF) were diluted in Hi-DiTM Formamide (Applied Biosystems, Foster City, CA, USA) and separated in a 3130 *xl* Genetic Analyzer (Applied Biosystems, Carlsbad, California). Data were analyzed using Genemapper (Applied Biosystems, Carlsbad, California) and only T-RF between 85 and 500 bp were included in the analysis. T-RFLP peaks were compared to the T-RF databases (see above and Table S1 and S2).

### *Fosmid library construction*

Twenty WGA products, checked by T-RFLP screening to display no significant diversity bias with regard to the initial sample, were pooled for fosmid library construction. As previously suggested, DNA branching was reduced by cutting junctions using an enzymatic treatment with 1000 U of Nuclease S1 (Promega, Madison, WI) during 1 h (Zhang et al., 2006; Neufeld et al., 2008). DNA was then purified by precipitation using SureClean reagent (Bioline France, Paris, France), washed using ethanol 70%, and rehydrated in sterile milliQ water. The metagenomic library was constructed using CopyControl™ HTP Fosmid Library Production Kit with pCC2FOS™ Vector (Epicentre, Madison, USA). For the end-repaired procedure, slight modifications to the standard protocol were made by adding 40 U of DNA polymerase I (New England Biolabs, Evry, France) to the enzymatic mix provided by the manufacturer and reaction was incubated at 25°C for 3 h. The end-repaired DNA was separated in 1% low melting point agarose gel (Invitrogen, Carlsbad, CA) in Tris-Borate-EDTA buffer 0.5X, at 200 V with linear pulse time of 0.5-1.5 s for 20 h on a PFGE Chef DRII (BioRad, Hercules, CA). Fragments between 25 and 55 kbp were cut out of the gel after staining the DNA ladder with ethidium bromide. DNA was purified after agarose digestion for 5 h at 45°C in the presence of 10 U.g<sup>-1</sup> GELase (Epicentre), and concentrated on 100 kDa Amicon devices (Millipore, Molsheim, France). DNA (~ 300 ng) was added to a ligation reaction with the pCC2FOS™ Vector and packaged according to the manufacturer protocol. Once titrated, the entire phages suspension was used to infect an *Escherichia coli* EPI300-T1R culture in exponential phase. The infected metagenomic library was frozen in liquid nitrogen and stored at -80°C in glycerol (20% final concentration) until screening. After thawing, the desired number of clones was plated on LB Agar containing 12.5 µg.mL<sup>-1</sup> chloramphenicol for screening procedure.

### *Fosmid library screening*

Colony picking of 55 296 clones and library screening using high-density colony filter were performed at the "Centre National de Ressources Génomiques Végétales" (Institut National de Recherche Agronomique, Toulouse, France), as described elsewhere (Gonthier et al., 2010). In order to increase the number of selected PBS region-containing fosmids and the coverage of this region, hybridizations were performed using probes targeting three loci located at different positions in this region: the *cpeC* gene coding for a PEI linker polypeptide, the *mpeBA* operon coding for the PEII  $\alpha$  and  $\beta$  subunits and *rpcBA/cpcBA* operon encoding the PC  $\alpha$  and  $\beta$  subunits (see Fig. 4 in Six et al., 2007). Probes were synthesized by PCR using the pool of WGA products of the Astan station and degenerate primers that were designed with Primaclade (Gadberry et al., 2005) and manually refined

based on alignments of these three loci retrieved from the *Synechococcus* genomes available (Table 2). PCR products were purified on agarose gel using the Qiaquick Gel Purification kit (Qiagen, Courtaboeuf, France) and quantified by Quanti-iT™ PicoGreen dsDNA assay (Invitrogen, Carlsbad, CA) according to manufacturer's instructions using a Tecan microplate reader (Tecan, Männedorf, Switzerland). Amplicons were labeled with [ $\alpha$ -<sup>32</sup>P]dCTP by random priming using Ready-To-Go DNA Labelling Beads and purified through Illustra™ ProbeQuant™ G-50 Micro Columns (GE Healthcare, Waukesha, WI). Hybridization of the filters was performed with pools of the three probes as described (Sambrook and Russel, 2001). Filters were scanned and analyzed using a Storm 860 PhosphorImager (GE Healthcare, Waukesha, WI).

#### *Fosmid selection and sequencing*

To select fosmids to be sequenced, positives clones selected by hybridization were secondarily screened by colony PCR targeting the three above mentioned loci as well as another gene of the PBS region, *cpeU*. After induction to high copy number according manufacturer's protocol, fosmid DNA was extracted using NucleoSpin Plasmid kit (Macherey-Nagel, Hoerd, France) and quantified with a ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA). End-sequencing of PCR positive clones was then performed according to the Epicentre manual, in order to check for the theoretical size and position of each insert within the PBS region. The real insert size was then determined on 25 selected clones by PFGE after fosmid DNA digestion by *NotI* as described above, before sequencing. These clones were pooled before generating a paired-end library, which was the sequenced using the Illumina technology (2 x 151 bp reads) with at least 10 bp overlap. Reads corresponding to the vector and the *E. coli* host were eliminated through mapping using the Geneious Workbench (Biomatters, Auckland, New-Zealand), then an assembly was realized using the CLC AssemblyCell software (CLCBio, Prismet, Denmark).

#### *454 pyrosequencing*

To complement the T-RFLP analyses, *petB*, *r/cpcBA* and *cpeBA* loci were sequenced by 454 pyrosequencing to assess the diversity of *Synechococcus* samples at each step of our metagenomic approach: TFF, after cell sorting and after WGA. Amplifications were conducted with "fusion" primers constituted of a 454 adaptor, a TCAG key used for amplicon sequencing and a multiplex identifier (MID) tag for sample identification from a mixed pool, according to the manufacturer instructions.

While *cpeBA* operon was obtained using the primer set SynB1F - SynA3R (Everroad and Wood, 2006), we amplified the *r/cpcBA* operon using the reverse primer designed by Haverkamp et al. (2008) but re-designed the forward primer using Primaclade (Gadberry et al., 2005) on a manually

refined multiple sequence alignment containing *Synechococcus* sequences from marine, brackish and freshwater environment. This primer set was then tested on a selection of 14 marine *Synechococcus* strains representative of the different pigment types (Six et al. 2007; Humily et al., submitted). PCR reactions were performed in triplicates, in 50 µL final volume following conditions described in Table 2, using the minimum number of cycles to obtain a thin band on agarose gel to minimize PCR biases. Amplicons were pooled and purified using the NucleoSpin® Extract II kit (Macherey-Nagel, Hoerd, France) to obtain a final amount of at least 2 µg. All samples were then pooled in equimolar amounts for 454 pyrosequencing on a Roche GS-FLX system (454 Life Sciences from Roche, Brandford, CT, USA) using a emPCR amplicon kit Lib-L according to the manufacturer's protocol at the Biogenouest® Genomics platform (Rennes, France).

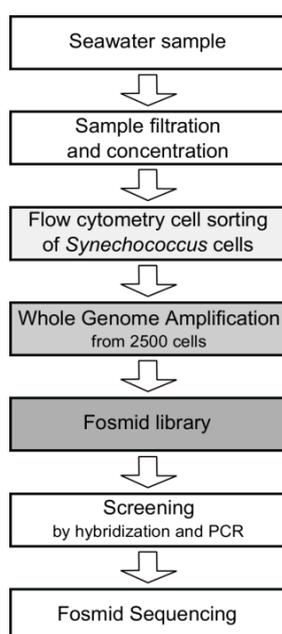
#### *Sequence treatment and phylogenetic analyses*

Sequences that were considered of low quality, i.e. when the key was absent, amplicon length <250 pb or >600pb or undetermined bases represented more than 7% of the sequence, were removed from the analysis. Sequences were denoised using Acacia (v 1.52; Bragg et al., 2012) after trimming to 450 bp for *petB* and to 430 bp for the other loci. Intergenic sequences within *r/cpcB* and *r/cpcA* operons were then removed before clustering using MOTHUR (Schloss et al., 2009). To define OTUs, a cut-off value of 94% nt sequence similarity was used for *petB*, as previously described (Mazard et al., 2012), and 95% for the other loci. OTU defined by singletons sequences were removed before calculation of rarefaction curves, diversity indices and coverage value using MOTHUR. Sequences were then used to generate multiple alignment using MAFFT (G-INS-I option, v 6.953; Katoh et al., 2005).

Taxonomic assignment of each OTU was made by phylogenetic analyses using the same database used for T-RFs identification for *petB* and 42 protein sequences retrieved from marine *Synechococcus/Cyanobium* genomes for *r/cpcBA* and *cpeBA* loci, corresponding to 333 and 340 aligned positions, respectively. Phylogenetic trees with references sequences were performed by Maximum likelihood using PhyML (v3.0; Guindon et al., 2009), a LG+G model (Le and Gascuel, 2008) and a BIONJ starting tree and visualized using FigTree v 1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>). Robustness of inferred topology was supported by bootstraps resampling (1000). Pyrotag sequences and a few additional partial sequences retrieved from Genbank were then aligned with the reference alignments and added to consensus trees using the ADD\_BY\_PARSIMONY algorithm implemented in ARB (Ludwig et al., 2004). Unique environmental gene sequences obtained in this study were deposited in GenBank database under the following accession numbers: XXXXXX-XXXXXX for *petB* amplicons, XXXXXX-XXXXXX for *r/cpcBA* amplicons and XXXXXX-XXXXXX for *cpeBA* amplicons.

## RESULTS AND DISCUSSION

Major breakthroughs have been achieved in recent years in the understanding of the genetic and functional diversity of marine microbial communities (Gilbert and Dupont, 2011). These progresses have been allowed by the advent of a number of novel technologies allowing to access at first the genetic diversity of natural populations. These include PCR which allowed the discovery of SAR11, the most abundant group of heterotrophic bacteria in the ocean (Giovannoni et al., 1990). Its combination with flow cytometry then allowed to target a specific abundant population, such as *Prochlorococcus* (Urbach et al., 1998). However, the access to functional diversity on the level of the global community required to access either fragment of larger size with for instance the BAC (Béjà et al., 2000) or a very deep sequencing coverage (Venter et al., 2004), that is now considerably facilitated by the advent of NGS (Rusch et al., 2007). With regard to all these approaches, the major advance allowed by WGA was to give access both genetic and functional diversity and to target specific organisms by increasing the DNA amount by several orders of magnitude, from femto or picogram quantities of DNA (Dean et al., 2001). The combination of cell sorting and WGA allowed for instance to study the genetic diversity of several phytoplankton groups (Cuvelier et al., 2010; Lepere et al., 2011; Vaultot et al., 2012; Malmstrom et al., 2013), including *Synechococcus* (Palenik et al., 2009; Mazard et al., 2011). It was also successfully applied to sequence the UCYN-A genome starting from only 5,000 sorted cells (Tripp et al., 2010). Nevertheless, starting from more than one cell most often result in obtaining consensus genomes, that can only be avoided by working at the single cell level (SAGs), but that nowadays only allows to amplify and sequence partial genomes with no control on the sequenced genomic regions (Zhang et al., 2006; Rodrigue et al., 2009; Malmstrom et al., 2013).

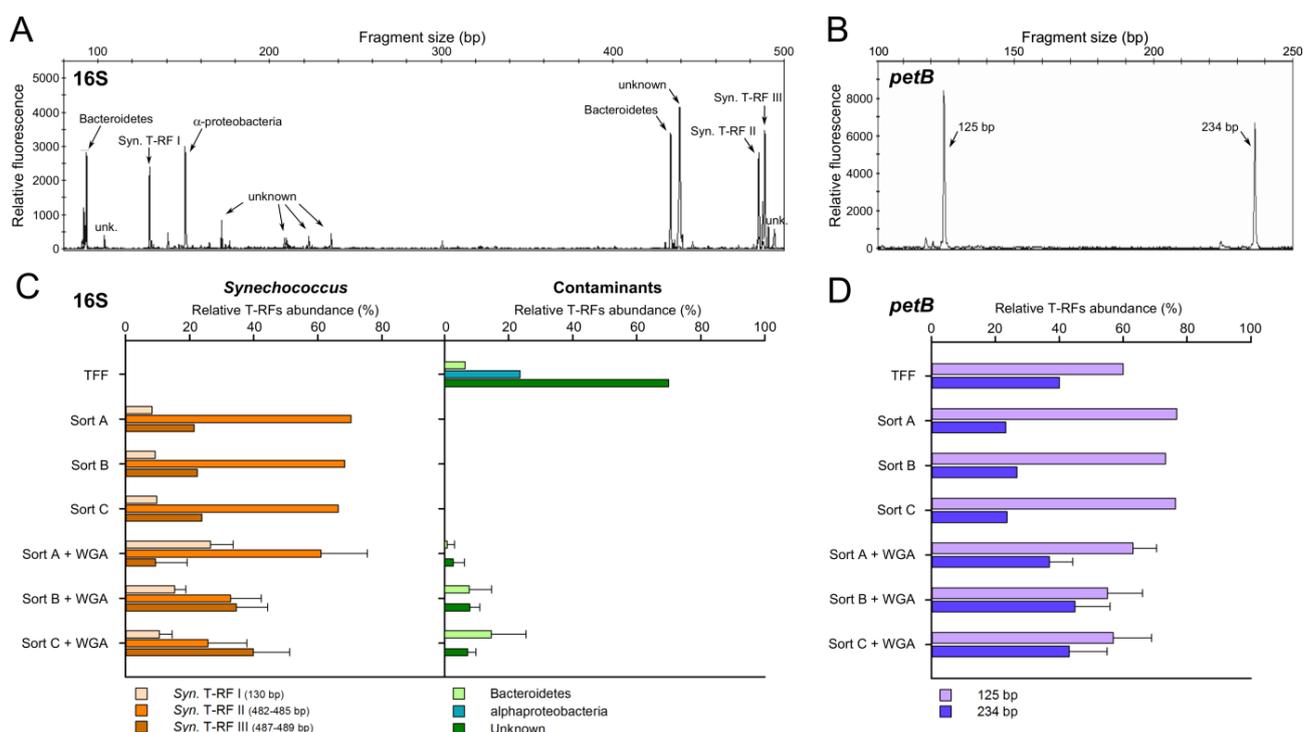


**FIG. 1.** Schematic representation of the targeted metagenomic pipeline used for PBS genomic regions analyses.

Here, we developed a metagenomic pipeline, combining flow cytometry cell sorting, WGA and fosmid library construction, which allows not only to specifically target *Synechococcus* cells but also specific regions of its genome (Fig. 1). Since this approach is not devoid of the biases inherent to each method, we evaluated their effects at each step of the procedure by following both the genetic composition of the *Synechococcus* community and the relative abundance of contaminant heterotrophic bacteria by T-RFLP analyses and/or pyrosequencing. While T-RFLP is a highly reproducible and low cost technique allowing to screen rapidly a high number of samples, pyrosequencing give access to the true genetic diversity of the samples.

*Efficiency of the combination of cell sorting and WGA to obtain large DNA amounts from natural Synechococcus populations*

T-RFLP targeting 16S rRNA was used to assess the contamination level by non-*Synechococcus* bacterial DNA at different steps of the procedure: i) the initial sample concentrated by TFF, ii) the flow sorted samples (3 technical replicates, named A to C, each constituted of 3 consecutive sorts, the first one based on natural fluorescence and two after DNA staining) and iii) the 48 samples obtained after WGA of the final sort of each technical replicate (Figs. 2A, 2C and S3). Peaks due to individual T-RF were identified based on an *in silico* analysis of a large 16S rRNA *Synechococcus* database that allowed us to predict T-RF at 129, 491 and 492 bp as *Synechococcus*, so by extension all the other T-RF observed in environmental samples were attributed to heterotrophic contaminants (Table S1). Compared to the above predictions, three T-RF were actually observed in the Astan sample that can be attributed to *Synechococcus* : *Syn.* T-RF I at 130 bp, *Syn.* T-RF II at 482-485 bp and *Syn.* T-RF III at 487-489 bp (Fig. 2A), and this was confirmed by T-RFLP analyses of cultured and sorted *Synechococcus* cells (see below). Although somewhat large (up to 7 bp), such a drift between predicted and measured T-RF sizes has already been reported and was shown to increase with T-RF size (Kaplan and Kitts, 2003; Bukovská et al., 2010). We also noticed that *Syn.* T-RF II and III were duplicated, which may be due to polymerase errors during PCR and/or during WGA. Additionally, the flow cytometry set up used in this study (Exc 488 nm) did not allow to detect PC-rich *Synechococcus* cells (pigment type 1), often found in coastal or low salinity waters (Haverkamp et al., 2008). However, a complementary flow cytometric analysis of the initial Astan samples with dual laser of 488 nm and 635 nm (exciting PC) allowed to check that only PE-containing *Synechococcus* cells were present in the sample (data not shown). Furthermore, no *Prochlorococcus* cells were detected since this genus is systematically absent from the English channel waters (Sournia and Birrien, 1995; N. Simon, pers. comm).



**FIG. 2.** T-RFLP analysis of the diversity of *Synechococcus* and contaminant bacteria at station Astan as assessed at different stages of the metagenomic pipeline by targeting the 16S rRNA (A, C) and *petB* (B, D) genes. (A) Composite profile of 16S rRNA T-RF. (B) T-RFLP profile targeting *petB* for the initial sample concentrated by TFF. (C, D) Histograms of relative T-RF abundance of the *Synechococcus* and/or bacterial community composition based on T-RFLP targeting 16S rRNA (C) or *petB* (D) genes. Mean and standard deviations were calculated on the 47 WGA reactions (out of 48) for which *Synechococcus* T-RF were detected.

While the concentration of *Synechococcus* in the initial sea water sample was fairly low ( $\sim 840 \text{ cells.ml}^{-1}$ , Table 1), virtually no residual contaminants could be detected after cell sorting in all three replicates, neither by flow cytometry analysis ( $< 2\%$ ; Fig. S2), nor by T-RFLP targeting 16S rRNA (Fig. 2C). This indicates that the three-step sorting procedure was highly reproducible, since the three technical replicates gave similar results, and also very efficient to remove all types of contaminants. The first sorting step, based on natural fluorescence and side scatter, efficiently eliminated photosynthetic picoeukaryotes and large non-living particles. The two other sorting steps, performed after DNA staining using a combination of SYTO-9 and SYTO-13, allowed us to completely remove heterotrophic bacteria. Altogether, these 3 steps also reduced free DNA, which can be present at high concentrations in environmental samples particularly after tangential filtration that can lead to cellular lysis (Rodrigue et al., 2009), to insignificant levels.

After the WGA step, some T-RF peaks assigned to contaminants in the initial sample re-appeared and in most cases, their relative abundance represented 0 and 46 % of the total bacterial T-RF (Figs. 2C and S3), though in one out of the 48 WGA products, contaminants represented all of the T-RF present in the sample. As suggested in a previous study (Swan et al., 2011), a discrepancy in cell lysis efficiency between *Synechococcus* and their bacterial counterparts might explain this differential

amplification. Preliminary tests performed on cultured isolates indeed showed that thermal lysis allows to obtain less contaminated profiles than chemical lysis (data not shown), but we did not retain the latter option, which generates smaller WGA products (Fig. S1). This differential amplification may also suggest that heterotrophic bacterial DNA could be preferentially amplified over *Synechococcus* DNA during the WGA step. It is also worth noting that the relative proportion of contaminants was higher in WGA reactions performed on the second and third sorted replicates (sorts B and C), which chronologically were performed after sort A (Fig 2C). This may result from a specific decrease of *Synechococcus* cell integrity and/or DNA degradation in the TFF concentrated sample that was kept on ice during several hours until cell sorting. Such modifications of the global community structure as a result of WGA have already been reported in the literature and were attributed to either amplification of free contaminant DNA (Raghunathan et al., 2005; Kvist et al., 2007; Woyke et al., 2010) or differential amplification between taxa (Abulencia et al., 2006; Chen et al., 2008; Lepere et al., 2011; Blainey, 2013). The latter effect has been suggested to result from difference in GC%, presence of repeated regions (Pinard et al., 2006), different chromosome conformation (Schoenfeld et al., 2010), chromosomal breaks, relative cell abundance or DNA amount of the various taxa present in the sample (Chen et al., 2008; Arakaki et al., 2010) and/or stochastic effects during initial priming and amplification (Rodrigue et al., 2009). In the present study, screening of the WGA products using T-RFLP targeting 16S rRNA allowed us to circumvent this problem by selecting the samples with the lowest contaminant load.

#### *Effect of cell sorting and WGA on the Synechococcus cell communities*

The abovementioned differential amplification of taxa during the WGA step may also affect the relative abundance of the different clades within the *Synechococcus* population. To estimate this potential bias, we developed a T-RFLP approach targeting the *petB* gene, which is specific of photosynthetic organism and provides a much finer taxonomic resolution of the *Synechococcus* genus than 16S rRNA (Mazard et al., 2012). It must be noted that a given T-RF may correspond to different (sub-)clades and that reciprocally some strains belonging to the same (sub-)clade may have different T-RF (Table S2), indicating that this approach does not perfectly reflect the composition of the *Synechococcus* population at the (sub-)clade level. Still, variations in the T-RF pattern observed between the different stages of the metagenomic pipeline can be used for rapid screening of changes in the *Synechococcus* community composition between samples.

The set of restriction enzymes *HpaII/HinfII* used for the T-RFLP analyses of the Astan samples provided two T-RFs (125 and 234 bp), which were among the eight T-RF predicted from *in silico* analysis of a database including sequences from both *Synechococcus* isolates and environmental

samples (Fig. 2B and 2D; Table S2). These T-RF sizes include strains belonging to the four main *Synechococcus* clades dominating in oceanic regimes (I-IV; Zwirgmaier et al., 2008), with T-RF 125bp encompassing strains from clades I, III and IV and T-RF 234bp, strains II, IV, VIII, XVI and sub-cluster 5.3-I (Table S2). Although a moderate change was observed in the relative proportion of the two T-RF in all 3 replicates after cell sorting, the 125 bp T-RF remained dominant in the sorted samples (Fig. 2D). The WGA also seemingly allowed to preserve the diversity of the initial sample since both T-RF we found in similar proportion after this step, despite some variability in the community composition between WGA products. In combination with 16S rRNA T-RFLP, these *petB* profiles helped us to select twenty WGA products not only with a low contamination level by heterotrophic bacteria but also representative of the *Synechococcus* community present at the Astan site, among which the equivalent of seven products were used to construct the fosmid library.

#### *Assessment of the content and quality of the fosmid library from the Astan station*

The fosmid library allowed us to obtain a very good coverage of genetic material from the Astan station, since it contained more than 500 000 clones, among which 50 000 were randomly screened by hybridization using a pool of three PBS probes targeting the *c/rpcBA*, *cpeBA* and *mpeBA* loci, then by PCR targeting the same genes as well as *cpeU*. This two-step screening was efficient to retrieve clones containing at least a part of the PBS region, since the great majority (94%) contained at least one of the four genes screened by PCR (Table 3). Although fosmid end-sequencing and mapping of these sequences onto reference *Synechococcus* genomes revealed the presence of some chimeric sequences, these were not retained for further sequencing. Selection of the 25 clones to be sequenced also took into account the predicted coverage of the PBS region as well as the largest possible genetic diversity as assessed by BLASTN analyses. After sequencing and read assembly, the great majority of contigs obtained were indeed included into the PBS gene region and sometimes encompassed genes of the conserved adjacent regions (Fig. 3). Nevertheless, read assembly only allowed us to obtain partial PBS gene regions, probably due to the small size of the paired-end library (~ 350 bp) and the high level of multiplexing (all 25 fosmids were pooled before library construction; Fig. 3).



**FIG. 3.** Alignment of the contigs retrieved from the Astan station against PBS regions from the most closely related marine *Synechococcus* reference genomes. Grey shaded areas above each reference genomic region indicate the coverage obtained by mapping all environmental reads onto each reference PBS region using BlastN. Yellow and green shaded areas indicate specific gene signatures as described in the main text. Only the contigs with a size larger than 1700 bp and non-redundant are shown on the figure.

**TABLE 3. Characteristics of SOMLIT-Astan fosmid library**

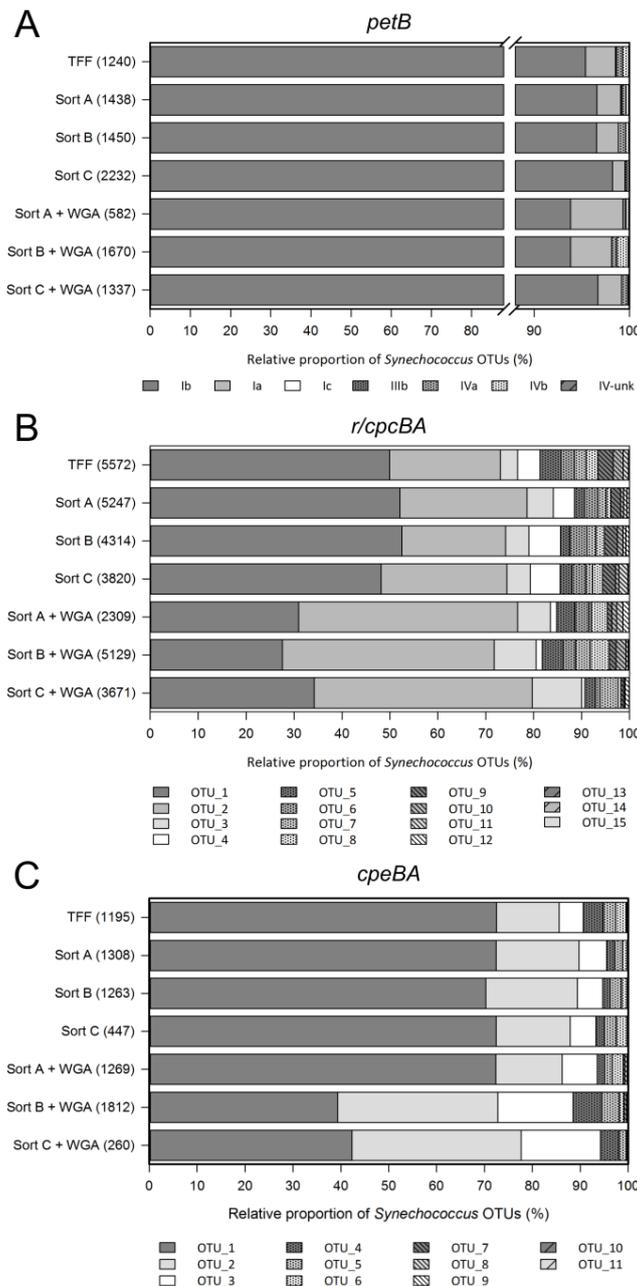
Step	Parameter	Value
<i>Fosmid library</i>	Total number of clones	506,296
	Number of screened clones	55,296
<i>Screening by hybridization</i>	Number of positive clones	768
<i>Screening by PCR</i>	Number of positive clones	268
	# with at least 1 gene	94%
	# with at least 2 genes	58%
	# with at least 3 genes	19%
	# with 4 genes	4%
<i>Validations</i>	Number of clones screened by end-sequencing	80
	Sequences affiliated to <i>Synechococcus</i>	93%
	Average insert size (bp) ( $n=25$ )	35,386
<i>Sequencing/Assembly</i>	Number of fosmid sequenced	25
	Contigs > 1 kbp	69
	Mean coverage depth	5,136 x

*Analysis of the Synechococcus community structure and pigment type at the Astan station off Roscoff*

The Astan site is representative of permanently mixed waters of the Western Channel as no stratification occurs during the year (Sournia and Birrien, 1995). It is characterized by mesotrophic waters with moderate nitrates and phosphates levels, and is weakly influenced by the Penzé estuary located near Roscoff (Table 1). Picophytoplankton is dominated year-round by picoeukaryotes, while the photosynthetic prokaryotes fraction is exclusively represented by *Synechococcus*, which can reach abundances exceeding  $10^4$  cells.mL<sup>-1</sup> (Simon N., pers. comm.). Sampling was performed in May from surface water, after the annual early spring bloom of *Synechococcus* occurring in March due to water mixing.

Genetic diversity of the *Synechococcus* population was analyzed using pyrosequencing targeting the *petB* gene. After cleaning and denoising, at least 582 *petB* amplicons were obtained for each sample (Fig. 4), which proved sufficient to cover the whole *Synechococcus* diversity, as attested by a Good's coverage index close or equal to 1 (Table 4). Additionally, rarefaction curves were performed at 94 % identity, a cut off previously used by Mazard et al. (2012) and shown to be suitable for assignment of sequences at the sub-clade level (Fig. S4). This allowed us to identify 9 OTU altogether among the entire dataset, but only 4 were represented by more than 4 reads, suggesting that these rare OTU might correspond to sequencing errors and that the curves were in fact saturated in all samples. It is also worth noting that based on several diversity estimators, including the  $S_{\text{Chao1}}$  and Shannon-Weaver diversity indices (Table 4; Shannon and Weaver, 1949; Chao et al., 2005), no significant modifications of the community structure occurred along the metagenomic pipeline (Fig. 4A and Table 4). The *Synechococcus* population at the Astan station at the sampling

date appears largely dominated by clade I, with sub-clade Ib representing more than 96 % of the total reads, clade Ia up to 3% while the remaining community was composed of members of sub-clades Ic and IVa (Fig. 4A). These results are consistent with observations made in the Western Atlantic ocean (Ahlgren and Rocap, 2012) or cold mesotrophic waters of North Atlantic were representatives of clade I were always more abundant than clade IV (Huang et al., 2012; Mazard et al., 2012). However, this contrasts with abundances profiles found in California coastal waters (Tai and Palenik, 2009), or in Arctic, South Pacific and Eastern Atlantic were clade IV cells typically dominate (Zwirgmaier et al., 2008).



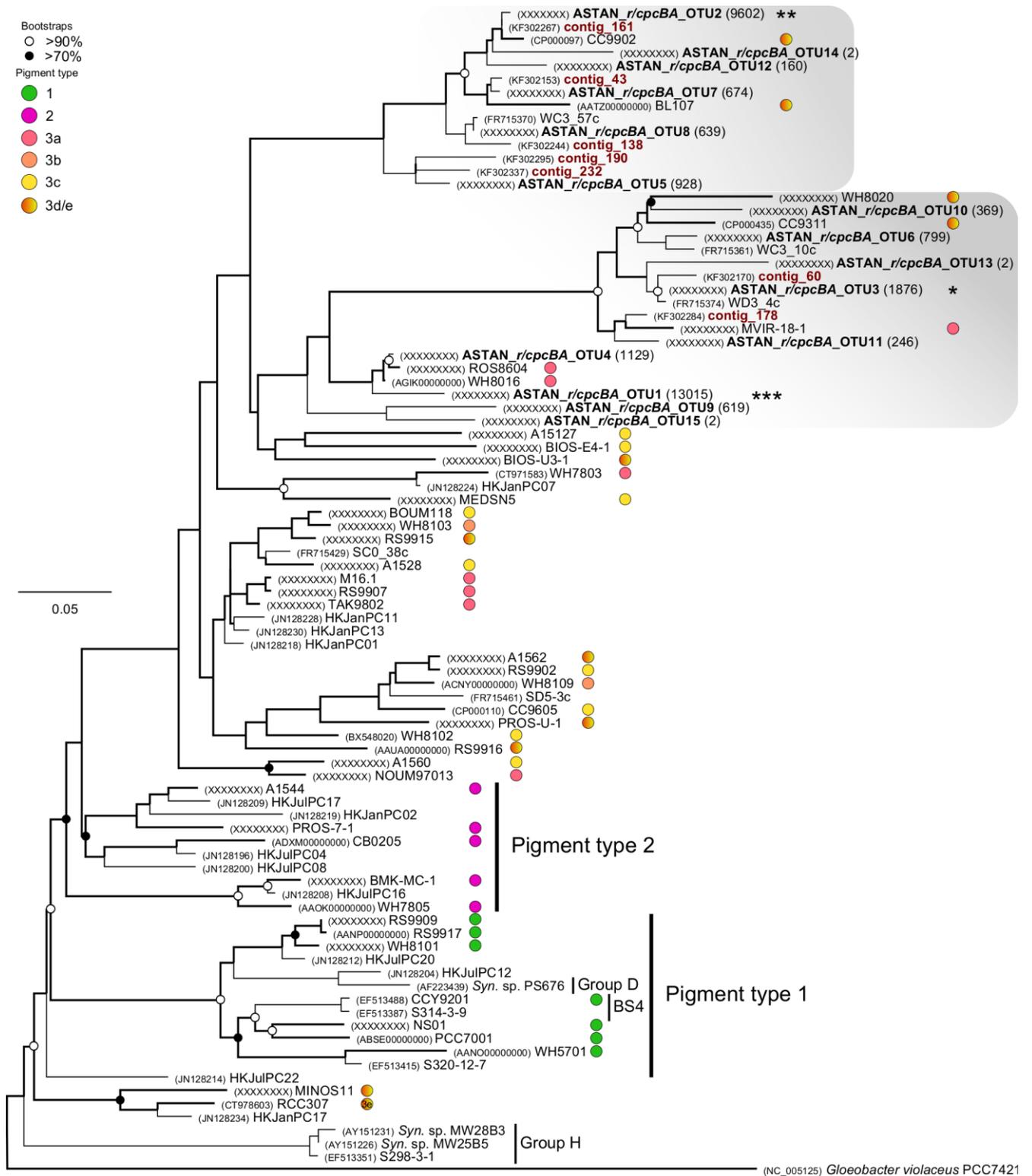
**FIG. 4.** *Synechococcus* community structure determined at each step of the metagenomic pipeline by pyrosequencing targeting *petB* (A), *r/cpcBA* (B) and *cpeBA* (C) loci. OTU were calculated using 94% and 95 % similarity cut-off for *petB* and the two operons, respectively.

**TABLE 4. Description and analyses of the pyrosequenced samples analyzed during this study.**

Gene/ Operon	Sample	# raw reads	# cleaned reads	% removed reads	# OTUs	Good's coverage	S <sub>Chao-1</sub>	Shannon index
<i>petB</i>	TFF	2119	1240	41	6	0.99	6	0.23
	Sort A	2116	1438	32	7	1.00	7.5	0.18
	Sort B	2335	1450	38	6	1.00	4	0.18
	Sort C	3545	2232	37	6	1.00	6	0.16
	Sort A + WGA	1061	582	45	5	1.00	5	0.2
	Sort B + WGA	3388	1670	51	7	0.99	6	0.28
	Sort C + WGA	2457	1337	46	5	1.00	4	0.16
<i>r/cpcBA</i>	TFF	6494	5572	14	13	0.99	14	1.55
	Sort A	6361	5247	18	12	1	11	1.39
	Sort B	5259	4314	18	13	1	13	1.5
	Sort C	4704	3820	19	13	0.99	12	1.53
	Sort A + WGA	2694	2309	14	12	1	11	1.51
	Sort B + WGA	6456	5129	21	13	0.99	26	1.52
	Sort C + WGA	4890	3671	25	13	0.99	12	1.24
<i>cpeBA</i>	TFF	6494	1195	82	8	1.00	9	0.98
	Sort A	6361	1308	79	7	1.00	7	0.89
	Sort B	5259	1263	76	8	0.99	7	0.85
	Sort C	4704	447	90	6	1.00	7	0.94
	Sort A + WGA	2694	1269	53	9	1.00	7	1.01
	Sort B + WGA	6456	1812	72	9	1.00	10	1.41
	Sort C + WGA	4890	260	95	6	0.99	6	1.24

Pyrosequencing of the *r/cpcBA* and *cpeBA* operons (encoding  $\alpha$  and  $\beta$  subunits of PC and PE I, respectively) was used to assess the pigment type diversity at the Astan station in complement to the analysis of PBS-containing fosmid. It must be noted that phylogeny made with these genes are not congruent with those based on core genes (including *petB*) and this has been related to lateral transfer that would have occurred during the evolution of these antenna genes (Six et al., 2007; Everroad and Wood, 2012). As for *petB*, Good's index indicated that the minimum number of sequences obtained for each operon, 3671 for *r/cpcBA* and 260 for *cpeBA* (Table 4), was sufficient to correctly cover the pigment diversity in all samples analyzed. Rarefaction curve allowed us to identify 15 *r/cpcBA* OTU and 11 *cpeBA* OTU among the entire dataset but only 12 and 8 respectively, were represented by more than 4 reads, indicating that these samples are also well covered (Fig. S4). For both markers, the cell sorting step apparently did not alter the *Synechococcus* community composition in contrast to the WGA step, which led to a notable modification of the relative proportion of the different OTU in most samples (Fig. 4B-C). This seemingly higher variability with regard to *petB* might be due to the larger number of OTU discriminated by these two markers, allowing to better highlight the differential taxa amplification caused by WGA (Fig. 4B-C). Yet, it is important to note that all dominant OTU are still present in WGA samples.

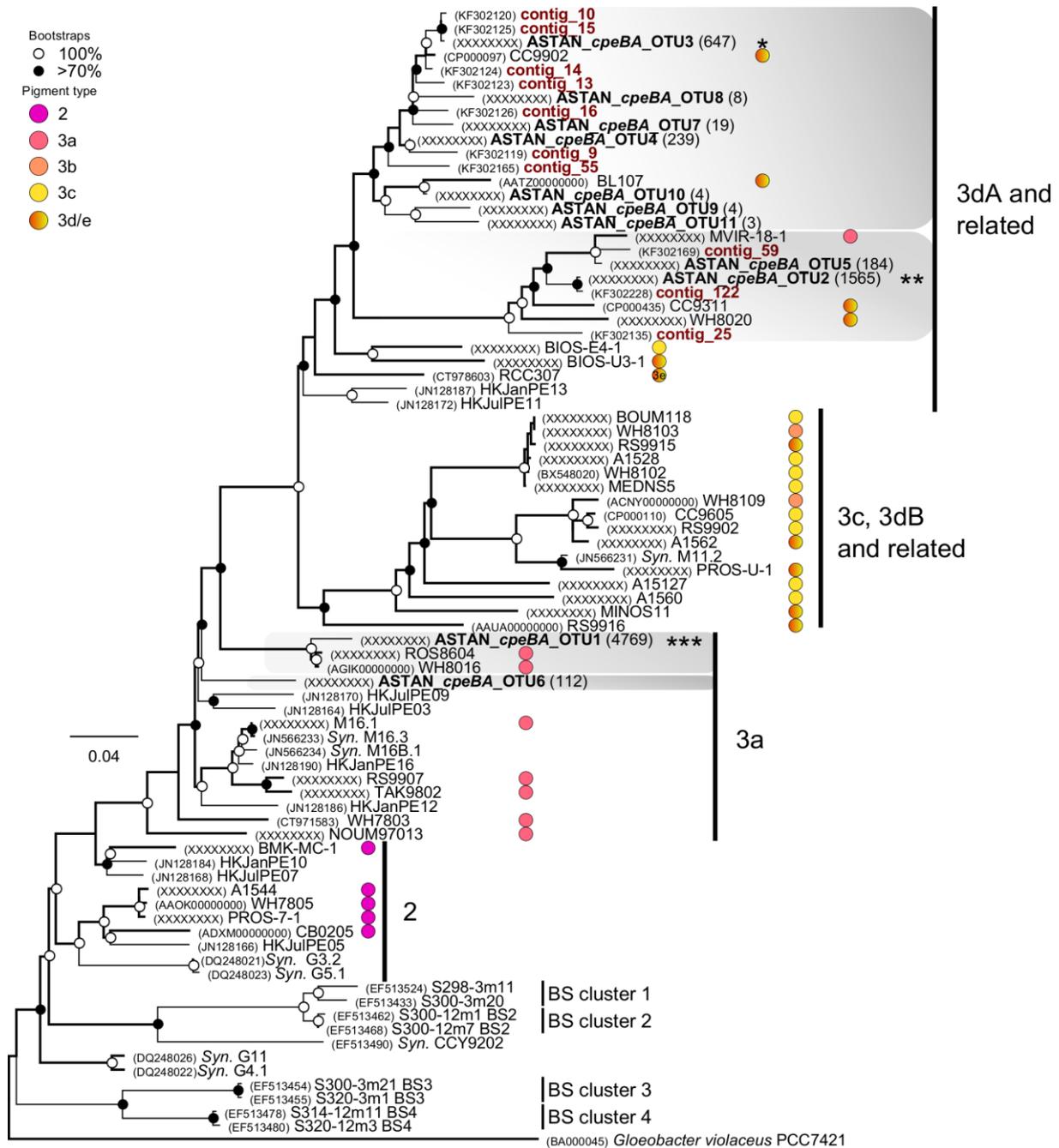
We then performed phylogenetic analyses of these sequences together with laboratory strains of known pigmentation to try and assign to specific pigment types OTU obtained by pyrosequencing and contigs retrieved from fosmid sequencing (Figs. 5 and 6).



**FIG. 5. Phylogenetic tree of *r/cpcBA* sequences from reference strains and environmental samples.** The phylogenetic tree was built using Maximum likelihood based on 340 aligned amino acid, 42 reference strains and the LG+G model. Environmental sequences (n=22) retrieved from public databases and from this study were added by ARB\_Parsimony to the initial consensus tree shown in bold lines. Bold and red letters correspond to sequences obtained by pyrosequencing and fosmid library sequencing, respectively. Asterisks indicate the three most abundant OTU in clone libraries sequenced by pyrosequencing. *Gloeobacter violaceus* PCC7421 was used as an outgroup. Only bootstrap values > 70 % are shown. Clades and/or pigment types were named following the nomenclature reported in previous studies (Crosbie et al., 2003; Haverkamp et al., 2008; Humily et al., submitted).

The *r/cpcBA* tree resolved well pigment types 1 or 2 (sensu Six et al., 2007), since strains exhibiting these pigmentations formed two well-defined clusters. In contrast, this marker was unable to distinguish between the different sub-types of pigment type 3 since some clusters encompassed 3a to 3d strains and all 4 pigment sub-types could be found in different clusters. However, none of the *r/cpcBA* sequences obtained at the Astan station, either by pyrosequencing or fosmid libraries, clustered with pigment types 1 or 2, suggesting that they were no representatives of these pigment types in the samples. The *cpeBA* tree brings complementary information since it allows to split members of the pigment type 3 into distinct clusters. The first one gathered pigment types 3dA strains, i.e. strains that are able to perform type IV chromatic acclimation (CA4), or related strains that possess a CA4-A genomic island involved in this process but display a different pigment phenotype (Humily et al., submitted). The second one encompassed pigment type 3 c, i.e. high PUB strains, and 3dB, i.e. CA4 strains possessing a CA4-B genomic island. At last, strains branching at the basis of these two clusters all belong to pigment type 3a, i.e. low PUB strains. Thus, the *cpeBA* phylogeny allowed us to putatively assign the most abundant OTU at station Astan (*cpeBA*\_OTU1) to type 3a and the second and third to type 3dA (or related strains), these abundant OTU representing 63, 20 and 8 % of the total *cpeBA* reads, respectively. Surprisingly, *cpeBA* sequences retrieved from fosmid clones only corresponded to pigment type 3dA and none to pigment 3a, which most likely have been counterselected during our screening procedure.

Our metagenomic approach also allowed us to obtain qualitative data on the whole PBS genomic regions of natural *Synechococcus* population at the Astan station. Mapping of all unassembled reads onto each reference PBS region showed that there was a good coverage of the region, especially for the chromatically acclimating strains, BL107 and CC9902 (Fig. 4) in agreement with the *cpeBA* phylogeny. Assembly of these reads using the CLC assembler provided partial environmental PBS regions that were aligned to the genome of the most closely related *Synechococcus* strain (Fig. 4). All environmental contigs, some of which covering up to 12 genes, showed a gene content and organization very similar to at least one of the reference strains (Six et al., 2007; Humily et al., 2013). As expected from *cpeBA* phylogenetic analyses, all the contigs were related to pigment type 3dA or related strains (i.e. 3aA; Fig. 6). Accordingly, the fusion gene unk7+8, which is a signature gene of typical pigment type 3a strains, such as WH8016, was never found in our dataset. It is also worth noting that two contigs (16 and 55) have a similar gene content that all CA4-A strains previously described (Humily et al., submitted), but lacks the uncharacterized phycobilin lyase gene *mpeU*. This gene organization was only found in the recently sequenced *Synechococcus* sp. MVIR-18-1, which despite having all other PBS genes present in CA4-A strains was unexpectedly found to display a pigment 3a phenotype (Humily et al., submitted).



**FIG. 6. Phylogenetic tree of *cpeBA* sequences from reference strains and environmental samples.** The tree was built by Bayesian Inference analysis based on 333 aligned amino acid from 42 marine *Synechococcus* isolates. Environmental sequences (n=21) retrieved in public databases and from this study were added by ARB\_Parsimony to the initial consensus tree shown in bold lines. Bold and red letters correspond to sequences obtained by pyrosequencing and fosmid library sequencing, respectively. Asterisks indicate the three most abundant OTU in clone libraries sequenced by pyrosequencing. *Gloeobacter violaceus* PCC7421 was used as an outgroup. Only bootstrap values > 70 % are shown. Clades an/or pigment types are named following the nomenclature reported in previous studies (Haverkamp et al., 2008; Humily et al., submitted).

## CONCLUSIONS AND FUTURE APPLICATIONS

In this paper, we described an original metagenomic pipeline that we used to explore the diversity of *Synechococcus* pigment types but that could be easily applied to any other selected region of the genome (Fig. 1). This approach is particularly well suited for such a model organism because of its abundance in the field and its natural autofluorescence which makes it easy to enumerate and sort by flow cytometry. Although it exhibits the biases inherent to each technology employed, the use of control molecular methods allowed us to get round most of these biases: T-RFLP to bypass the bias of differential amplification of taxa included in the initial environmental and fosmid end-sequencing to avoid chimeras, both effects being mainly due to the WGA step. This strategy allowed us to build library qualitatively (though not quantitatively) representative of the genetic and functional diversity of the initial sample, as attested by pyrosequencing of *petB*, *r/cpcBA* and *cpeBA* loci. Although phylogenetic analyses of the resulting sequences showed that there was a co-dominance of pigment type 3a and 3d in the Astan sample, which was maintained along the whole metagenomic pipeline, only fosmids of the latter pigment type were apparently selected for sequencing. This suggests that primers used to screen this library may have selected out fosmids containing type 3a PBS regions. However, the sequencing of a higher number of fosmids and/or the use of a mix of probes more representative of the community should allow us to circumvent easily this problem.

The dominance of pigment type 3a at the Astan station is congruent with previous observations using dual beam flow cytometry or spectrofluorimetry, which have shown the prevalence of low-PUB cells in coastal environments (Olson et al., 1990; Lantoiné and Neveux, 1997; Wood et al., 1998). The latter cells were shown to almost always co-occur with high-PUB cells in these environments (Olson et al., 1990; Katano et al., 2007). Results from the present study suggest that these coastal high-PUB strains could actually be 3d (CA4) cells, and not as previously assumed pigment type 3c, i.e. cells with a fixed high-PUB pigmentation, since those were not observed in our sequence libraries. By extension, it is possible that pigment 3c could rather be restricted to oceanic waters, where their high-PUB content can allow efficient capture of the blue photons prevailing in these environments (Olson et al., 1990; Kirk, 1994).

To our knowledge, this is first study that shows the presence of CA4 strains in the natural environment. Indeed, due to their variable PUB to PEB ratio, they cannot be easily discriminated from pigment types with fixed pigment ratio (mainly 3a and 3c) by fluorescence based approaches. Although a single environmental sample was analyzed here, the application of this molecular approach to a variety of trophic environments and depths should allow to decipher the distribution and ecological importance of CA4 cells and more globally of all *Synechococcus* pigment types.

*It is worth noting that data obtained from sequencing did not allow us to reconstruct entire fosmid inserts sequences. Some optimizations will be needed concerning preparation of library before sequencing, or maybe by using a lower level of multiplexing. We also planned to optimize assembly step by using other softwares. This will allow us to provide more reliable conclusions about PBS gene organization and various phylogenetic analyses. This work was still not done at the time of thesis writing.*

## ACKNOWLEDGEMENTS

This work was supported by the collaborative program METASYN with the Genoscope, the "Agence Nationale de la Recherche" Microbial Genomics Program (PELICAN, ANR-09-GENM-030) and the European Union's Seventh Framework Programs, MicroB3 and MaCumba (grant agreements 287589 and 311975, respectively). We thank Fabienne Rigaud-Jalabert for sample collection at the Astan station and "Service d'Observation en Milieu Littoral, INSU-CNRS, Station Biologique de Roscoff" for provided data. We are most grateful to the Biogenouest® Genomics core facility for its technical support, notably Gwenn Tanguy, for genotyping operations, Alexandra Dheilly, Delphine Naquin and Oscar Lima for pyrosequencing. H  l  ne Berg  s, Arnaud Bellec and Jo  lle Fourment from the CNRGV Platform are acknowledged for fosmid library screening. We are indebted to Dominique Boeuf for help with phylogenetic analyses.

## REFERENCES

- Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W.Q., Chang, S.H., Chang, H.W., Watson, D., Brodie, E.L., Hazen, T.C., and Keller, M.** (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* **72**: 3291-3301.
- Ahlgren, N.A., and Rocap, G.** (2012) Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* **3**: doi: 10.3389/fmicb.2012.00213.
- Arakaki, A., Shibusawa, M., Hosokawa, M., and Matsunaga, T.** (2010) Preparation of genomic DNA from a single species of uncultured magnetotactic bacterium by multiple-displacement amplification. *Appl Environ Microbiol* **76**: 1480-1485.
- Balzano, S., Marie, D., Gourvil, P., and Vault, D.** (2012) Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *ISME J* **6**: 1480-1498.
- B  j  , O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., and Jovanovich, S.B.** (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516-529.
- Biers, E.J., Sun, S.L., and Howard, E.C.** (2009) Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Appl Environ Microbiol* **75**: 2221-2229.

- Blainey, P.C.** (2013) The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol Rev* **37**: 407-427.
- Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G.W.** (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nature Methods* **9**: 425-426.
- Buitenhuis, E.T., Li, W.K.W., Vaulot, D., Lomas, M.W., Landry, M., Partensky, F., Karl, D.M., Ulloa, O., Jacquet, S., Lantoiné, F., Chavez, F., Macias, D., Gosselin, M., and McManus, G.B.** (2012) Bacterial biomass distribution in the global ocean. *Earth System Science Data Discussions* **5**: 301-315.
- Bukovská, P., Jelínková, M., Hřelová, H., Sýkorová, Z., and Gryndler, M.** (2010) Terminal restriction fragment length measurement errors are affected mainly by fragment length, G + C nucleotide content and secondary structure melting point. *J Microbiol Meth* **82**: 223-228.
- Chao, A., Chazdon, R.L., Colwell, R.K., and Shen, T.J.** (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* **8**: 148-159.
- Chen, Y., Dumont, M.G., Neufeld, J.D., Bodrossy, L., Stralis-Pavese, N., McNamara, N.P., Ostle, N., Briones, M.J.I., and Murrell, J.C.** (2008) Revealing the uncultivated majority: combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated *Methylocystis* in acidic peatlands. *Environ Microbiol* **10**: 2609-2622.
- Crosbie, N.D., Pöckl, M., and Weisse, T.** (2003) Dispersal and phylogenetic diversity of non-marine picocyanobacteria, inferred from 16S rRNA gene and *cpcBA*-intergenic spacer sequence analyses. *Appl Environ Microbiol* **69**: 5716-5721.
- Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messié, M., Tringe, S.G., Woyke, T., Welsh, R.M., Ishoey, T., and Lee, J.-H.** (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci U S A* **107**: 14679-14684.
- Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S.** (2001) Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11**: 1095-1099.
- del Giorgio, P.A., Bird, D.F., Prairie, Y.T., and Planas, D.** (1996) Flow cytometric determination of bacterial abundance in lake plankton with the green nucleic acid stain SYTO-13. *Limnol Oceanogr* **41**: 783-789.
- Everroad, C.R., and Wood, M.A.** (2006) Comparative molecular evolution of newly discovered picocyanobacterial strains reveals a phylogenetically informative variable region a  $\beta$ -phycoerythrin. *J Phycol* **42**: 1300-1311.
- Everroad, C.R., and Wood, M.A.** (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* **64**: 381-392.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P.** (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**: 237-240.
- Gadberry, M.D., Malcomber, S.T., Doust, A.N., and Kellogg, E.A.** (2005) Primaclade - a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* **21**: 1263-1264.
- Gasol, J.M., Zweifel, U.L., Peters, F., Fuhrman, J.A., and Hagstrom, A.** (1999) Significance of size and nucleic acid content heterogeneity as measured by flow cytometry in natural planktonic bacteria. *Appl Environ Microbiol* **65**: 4475-4483.
- Gilbert, J.A., and Dupont, C.L.** (2011) Microbial metagenomics: beyond the genome. *Ann Rev Mar Sci* **3**: 347-371.
- Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G.** (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60-63.
- Gonthier, L., Bellec, A., Blassiau, C., Prat, E., Helmstetter, N., Rambaud, C., Huss, B., Hendriks, T., Bergès, H., and Quillet, M.-C.** (2010) Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). *BMC Research Notes* **3**: 225.
- Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V., and Gascuel, O.** (2009) PhyML: Fast and accurate phylogeny reconstruction by Maximum Likelihood. *Infect Genet Evol* **9**: 384-385.

- Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., and DeLong, E.F. (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* **4**: e95.
- Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., and Stal, L.J. (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* **10**: 174-188.
- Huang, S., Wilhelm, S.W., Rodger Harvey, H., Taylor, K., Jiao, N., and Chen, F. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6**: 285-297.
- Humily, F., Partensky, F., Six, C., Farrant, G., Ratin, M., Marie, D., Kehoe, D.M., and Garczarek, L. (2013) A gene island with dual evolutionary origin is involved in chromatic acclimation in marine *Synechococcus*. *Environ Microbiol* Submitted.
- Kalyuzhnaya, M.G., Lapidus, A., Ivanova, N., Copeland, A.C., McHardy, A.C., Szeto, E., Salamov, A., Grigoriev, I.V., Suci, D., and Levine, S.R. (2008) High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature Biotech* **26**: 1029-1034.
- Kaplan, C.W., and Kitts, C.L. (2003) Variation between observed and true terminal restriction fragment length is dependent on true TRF length and purine content. *J Microbiol Meth* **54**: 121-125.
- Katano, T., Kaneda, A., Kanzaki, N., Obayashi, Y., Morimoto, A., Onitsuka, G., Yasuda, H., Mizutani, S., Kon, Y., and Hata, K. (2007) Distribution of prokaryotic picophytoplankton from Seto Inland Sea to the Kuroshio region, with special reference to 'Kyuchō' events. *Aquat Microb Ecol* **46**: 191-201.
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucl Acids Res* **33**: 511-518.
- Kirk, J.T.O. (1994) Light and photosynthesis in aquatic ecosystems. Cambridge, UK: Cambridge University Press.
- Kvist, T., Ahring, B.K., Lasken, R.S., and Westermann, P. (2007) Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl Microbiol Biotechnol* **74**: 926-935.
- Lane, D. (1991) 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*. Stackebrandt, E., and Goodfellow, M. (eds). New-York: John Wiley & Sons, pp. 115-175.
- Lantoine, F., and Neveux, J. (1997) Spatial and seasonal variations in abundance and spectral characteristics of phycoerythrins in the tropical northeastern Atlantic Ocean. *Deep-Sea Res Pt I* **44**: 223-246.
- Lasken, R.S., and Stockwell, T.B. (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol* **7**: 19.
- Le, S.Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**: 1307-1320.
- Lebaron, P., Parthuisot, N., and Catala, P. (1998) Comparison of blue nucleic acid dyes for flow cytometric enumeration of Bacteria in aquatic systems. *Appl Environ Microbiol* **64**: 1725-1730.
- Lepere, C., Demura, M., Kawachi, M., Romac, S., Probert, I., and Vault, D. (2011) Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiol Ecol* **76**: 513-523.
- Li, W. (1994) Primary productivity of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169-175.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., Lai, T., Steppi, S., Jobb, G., and Förster, W. (2004) ARB: a software environment for sequence data. *Nucl Acids Res* **32**: 1363-1371.
- Malmstrom, R.R., Rodrigue, S., Huang, K.H., Kelly, L., Kern, S.E., Thompson, A., Roggensack, S., Berube, P.M., Henn, M.R., and Chisholm, S.W. (2013) Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* **7**: 184-198.
- Marie, D., Shi, X.L., Rigaut-Jalabert, F., and Vault, D. (2010) Use of flow cytometric sorting to better assess the diversity of small photosynthetic eukaryotes in the English Channel. *FEMS Microbiol Ecol* **72**: 165-178.

- Marie, D., Brussaard, C., Partensky, F., Vaultot, D., and Wiley, J. (1999) Flow cytometric analysis of phytoplankton, bacteria and viruses. *Current Protocols in Cytometry Supplement 10*: 11.11.11-11.11.15.
- Martín, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., Yeates, C., He, S., Salamov, A.A., and Szeto, E. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotech 24*: 1263-1269.
- Mazard, S., Ostrowski, M., Garczarek, L., and Scanlan, D.J. (2011) A targeted metagenomic approach to determine the "Population Genome" of marine *Synechococcus*. In *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*. de Bruijn, F.J. (ed), pp. 301-307.
- Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J. (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol 14*: 372-386.
- Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahe, F., Bariat, L., Courties, C., Marie, D., Ras, J., Mauriac, R., Jeanthon, C., Bendif, E.M., Ostrowski, M., Scanlan, D.J., and Garczarek, L. (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences 8*: 2785-2804.
- Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C. (2008) Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol 10*: 1526-1535.
- Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V. (1988) Analysis of *Synechococcus* pigment types in the sea using single and dual beam flow-cytometry. *Deep-Sea Res 35*: 425-440.
- Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V. (1990) Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr 35*: 45-58.
- Palenik, B. (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol 67*: 991-994.
- Palenik, B., Ren, Q., Tai, V., and Paulsen, I.T. (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol 11*: 349-359.
- Partensky, F., Blanchot, G., and Vaultot, D. (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters : a review. *Bulletin de l'Institut océanographique 19 (Marine Cyanobacteria)*: 457-475.
- Pedrés-Alió, C. (2012) The rare bacterial biosphere. *Ann Rev Mar Sci 4*: 449-466.
- Pinard, R., de Winter, A., Sarkis, G.J., Gerstein, M.B., Tartaro, K.R., Plant, R.N., Egholm, M., Rothberg, J.M., and Leamon, J.H. (2006) Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics 7*: DOI: 10.1186/1471-2164-1187-1216.
- Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L., and Keller, M. (2007) Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl Environ Microbiol 73*: 3205-3214.
- Raghunathan, A., Ferguson, H.R., Bornarth, C.J., Song, W.M., Driscoll, M., and Lasken, R.S. (2005) Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol 71*: 3342-3347.
- Richardson, T.L., and Jackson, G.A. (2007) Small phytoplankton and carbon export from the surface ocean. *Science 315*: 838-840.
- Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R., and Chisholm, S.W. (2009) Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS One 4*: e686.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., and Remington, K. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol 5*: e77.
- Sambrook, J., and Russel, D.W. (2001) Hybridization of bacterial DNA on filters. In *Molecular Cloning : A laboratory Manual*. New-York: Cold Spring Harbor Laboratory Press, pp. 138-142.
- Schadt, E.E., Turner, S., and Kasarskis, A. (2010) A window into third-generation sequencing. *Hum Mol Genet 19*: R227-R240.

- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.
- Schoenfeld, T., Liles, M., Wommack, K.E., Polson, S.W., Godiska, R., and Mead, D. (2010) Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol* **18**: 20-29.
- Sekar, R., Fuchs, B.M., Amann, R., and Pernthaler, J. (2004) Flow sorting of marine bacterioplankton after fluorescence in situ hybridization. *Appl Environ Microbiol* **70**: 6210-6219.
- Shannon, C.E., and Weaver, W. (1949) The mathematical theory of communication. Urbana, IL: University of Illinois Press.
- Six, C., Thomas, J.C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., Scanlan, D.J., and Partensky, F. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.
- Sournia, A., and Birrien, J.-L. (1995) La série océanographique côtière de Roscoff (Manche occidentale) de 1985 à 1992. *Cahiers de Biologie Marine* **36**: 1-8.
- Stepanauskas, R. (2012) Single cell genomics: an individual look at microbes. *Curr Opin Biotechnol* **15**: 613-620.
- Stepanauskas, R., and Sieracki, M.E. (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci U S A* **104**: 9052-9057.
- Suenaga, H. (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol* **14**: 13-22.
- Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N.J., Masland, E.D.P., and Gomez, M.L. (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296-1300.
- Tai, V., and Palenik, B. (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* **3**: 903-915.
- Tringe, S.G., and Rubin, E.M. (2005) Metagenomics: DNA sequencing of environmental samples. *Nature Rev Gen* **6**: 805-814.
- Tripp, H.J., Bench, S.R., Turk, K.A., Foster, R.A., Desany, B.A., Niazi, F., Affourtit, J.P., and Zehr, J.P. (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464**: 90-94.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.
- Urbach, E., Scanlan, D.J., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J Mol Evol* **46**: 188-201.
- Vaulot, D., Lepère, C., Toulza, E., De la Iglesia, R., Poulain, J., Gaboyer, F., Moreau, H., Vandepoele, K., Ulloa, O., Gavory, F. and Piganeau, G. (2012) Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. *PLoS One* **7**: e39648.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D.Y., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., and Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.
- Wood, A.M., Phinney, D.A., and Yentsch, C.S. (1998) Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol-Prog Ser* **162**: 25-31.
- Wooley, J.C., Godzik, A., and Friedberg, I. (2010) A primer on metagenomics. *PLoS Comput Biol* **6**: e1000667.

- Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D.Y., McCutcheon, J.P., McDonald, B.R., Moran, N.A., Bristow, J., and Cheng, J.F.** (2010) One bacterial cell, one complete genome. *PLoS One* **5**: e10314.
- Woyke, T., Xie, G., Copeland, A., Gonzalez, J.M., Han, C., Kiss, H., Saw, J.H., Senin, P., Yang, C., Chatterji, S., Cheng, J.F., Eisen, J.A., Sieracki, M.E., and Stepanauskas, R.** (2009) Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: 10.
- Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., Yang, E.C., Duffy, S., and Bhattacharya, D.** (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**: 714-717.
- Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M.** (2006) Sequencing genomes from single cells by polymerase cloning. *Nature Biotech* **24**: 680-686.
- Zwirgmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vault, D., Not, F., Massana, R., Ulloa, O., and Scanlan, D.J.** (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147-161.

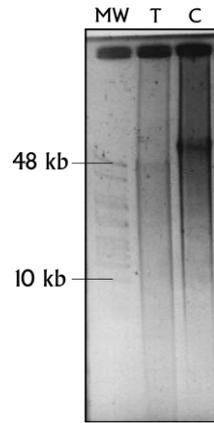
**SUPPLEMENTARY INFORMATION**

**TABLE S1.** Characteristics of the 16S rRNA T-RF found at different steps of the metagenomic pipeline

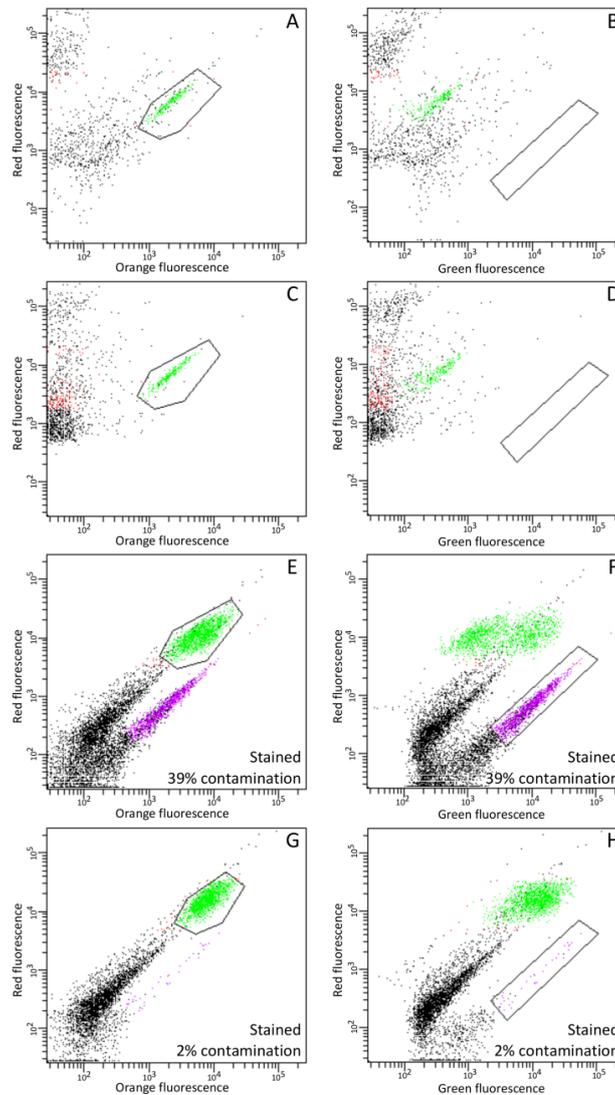
Gene	Restriction enzyme	<i>in silico</i> T-RFs (bp)	T-RF (bp)	Phylogenetic assignment	Closest species	
16S rRNA	<i>MspI</i>	90-91	89-93	<i>Bacteroidetes/Cytophaga</i>	<i>Microscilla</i> sp., <i>Marinoscillum furvescens</i>	
			104	n.d.		
			123	n.d.		
		129	130	cyanobacteria	<i>Synechococcus</i> ALMO3 (Ib)	
			150	151		$\alpha$ -proteobacteria
				172		n.d.
				210		n.d.
				221		n.d.
		435	434	<i>Bacteroidetes/shingobacteria</i>	<i>Mucilaginibacter rigui</i>	
			439	n.d.		
			461	456		$\gamma$ -proteobacteria
		482		n.d.		
		491-492	485-487	cyanobacteria	<i>Synechococcus</i>	
			491-492	487-489		cyanobacteria
		491		n.d.		
		494		n.d.		

**TABLE S2.** Characteristics of the *petB* T-RF either generated in silico using a database of cultured and environmental *Synechococcus* sequences, or experimentally detected at the Astan station.

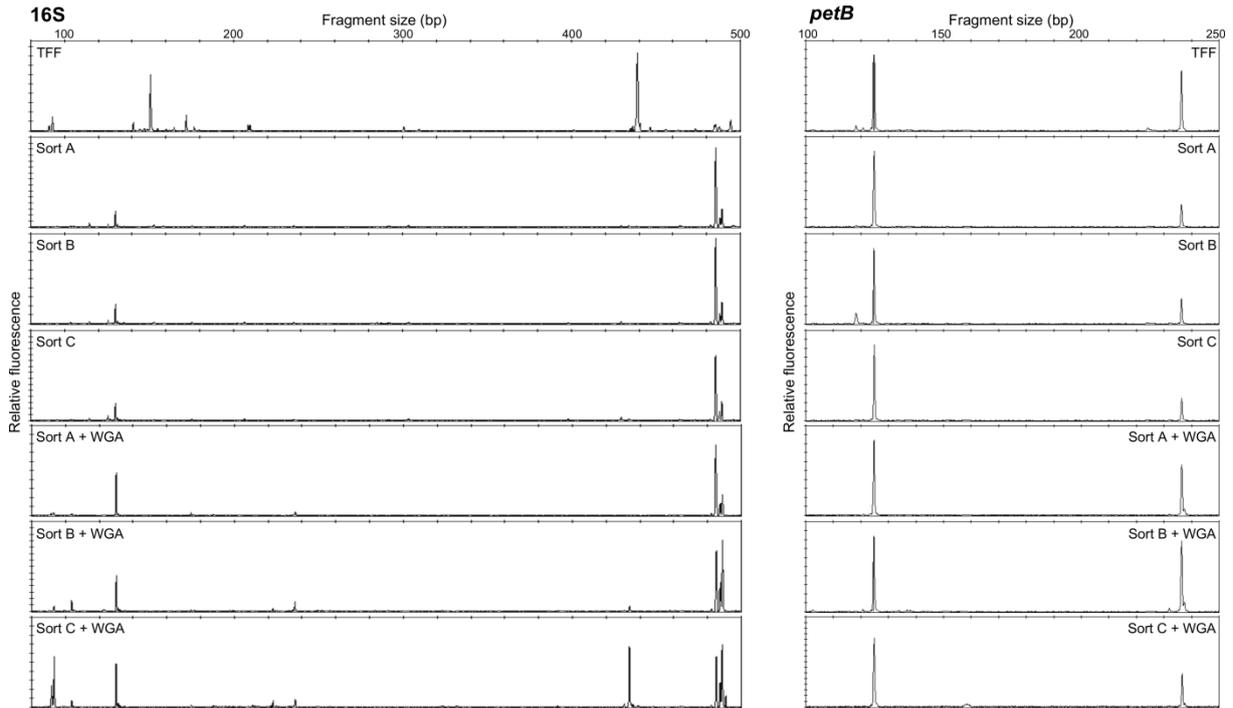
Gene	Restriction enzymes used	<i>In silico</i> T-RF (bp)	Experimental T-RF (bp)	<i>Synechococcus</i> isolates	Clade		
<i>petB</i>	<i>HinfI</i> , <i>HpaII</i>	101	99-101	BIOS-S15-1/MITS9220	CRD1a		
				Biosope_45 C4Y/BIOS-H3-2	CRD1b		
				PROSOPE_107/CC9311/WH8020	Ia		
				ALMO3/SYN20/WH8016	Ib		
				Biosope_141-D	Ic		
				A1511/RS9905/WH8102/WH8103	IIIa		
				A1528	IIIb		
				BL107	IVa		
				A15-37/RS9919	IIa		
				PROS-U-2	IIh		
		137	n.d.	A15-144	WPC1		
				CB0101	CB4		
				CB0205	CB5		
				RS9901/RS9916/RS9921	X		
				WH5701	Sub 5.2		
				231	231-236	RS9911	IIa
						A15-146/A5-19	IIb
						CC9605	IIc
						PROS-3-1	IIe
						A15-72	IIf
		CC9902	IVa				
		WH8101	VIII				
		PROS-7-2/PROS-3-2/M21B.3	XVI				
		RCC307	Sub 5.3-I				
		279	279-283	A15-43/A15-46/A15-48A1/WH7803	V		
				308	307-308	WH8109/RS9904	IIa
		WH7805/BL8	VIa				
		PROSOPE_130	VIc				
		RS9906/RS9917	VII				
		NS01/PCC 7001 (Cyanobium)	Cyanobium				
		A15-147	IIh				
		A15-60/A15-74	VII				
		MINOS11	Sub 5.3-I				



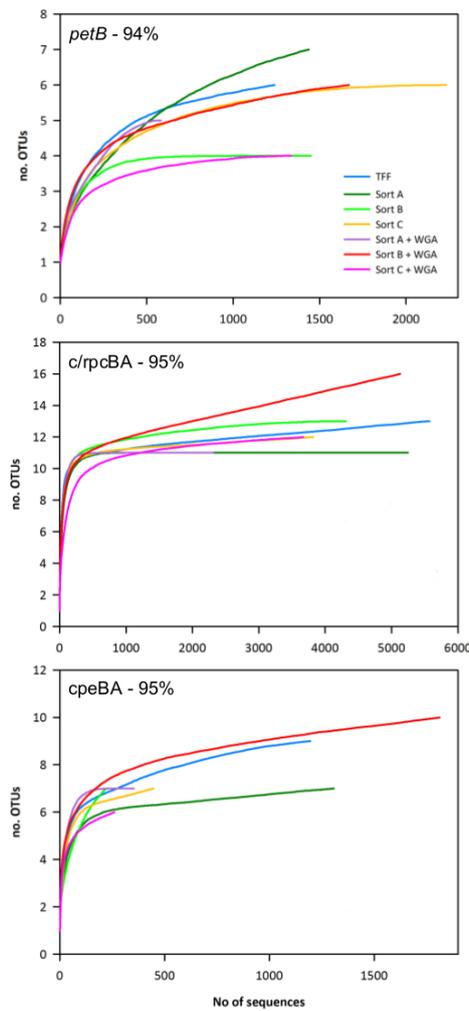
**FIG. S1.** Pulse Field Gel Electrophoresis analyses of WGA products from a sorted sample from the Astan station, amplified using Genomiphi v2 kit and denaturated either thermally (T, cells were boiled 3 min at 95°C) or chemically (C, alkaline denaturation and neutralization). MW (Molecular weight marker) : GeneRuler™ High Range DNA Ladder (Fermentas, Villebon sur Yvette, France).



**FIG. S2.** Flow cytometry analysis of the picophytoplanktonic community from the Astan station. (A,B) Initial unconcentrated sample (seawater), (C, D) after tangential flow filtration (TFF), (E, F) after 1 step of cell sorting, (G, H) after 3 steps of cell sorting. The red, orange and green fluorescence signals are proxy for chlorophyll, phycoerythrin and DNA content, respectively. The population indicated in green corresponds to *Synechococcus* cells, while contaminants are shown in purple.



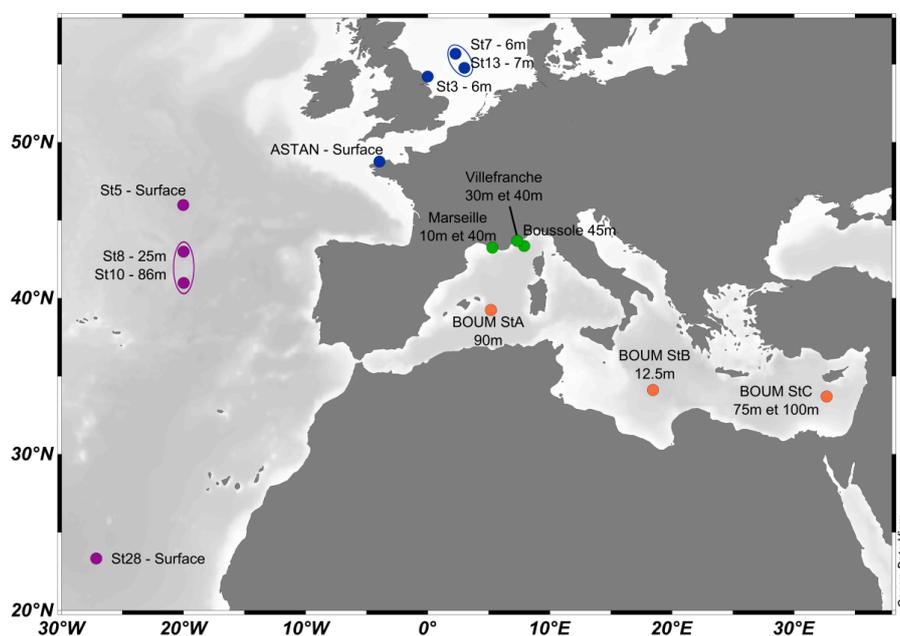
**FIG. S3.** Representative T-RFLP profiles at each step of the metagenomic pipeline.



**FIG. S4.** Rarefaction curves of the number of observe OTU for each sample at 94% an 95 % similarity cut-off for the *petB*, *c/rpcBA* and *cpeBA* loci, respectively.

#### IV. DONNÉES SUPPLÉMENTAIRES A CETTE ÉTUDE

Cette approche de métagénomique ciblée a été utilisée pour la fabrication d'un certain nombre de bibliothèques de fosmidés à partir d'échantillons environnementaux sélectionnés pour les différences de régime trophique et correspondant à 4 régions océaniques distinctes (Tableau 5) : la Mer du Nord, l'Océan Atlantique Nord ainsi que les eaux oligotrophes et côtières de Méditerranée (Fig. 50). Tout comme pour la station Astan, ces bibliothèques ont été criblées par hybridation au CNRGV et par PCR au laboratoire. Le choix des fosmidés d'intérêt, qui seront envoyés prochainement au Génomscope, est actuellement en cours.



**Figure 50** : Localisation géographique des différentes stations échantillonnées. Les échantillonnages provenant de la campagne océanographique BOUM sont indiqués en orange, les sites côtiers de Méditerranée en vert, ceux du site Astan et de la campagne CEFAS (Sortie 811) en bleu et enfin ceux d'une campagne réalisée en Atlantique Nord en violet.

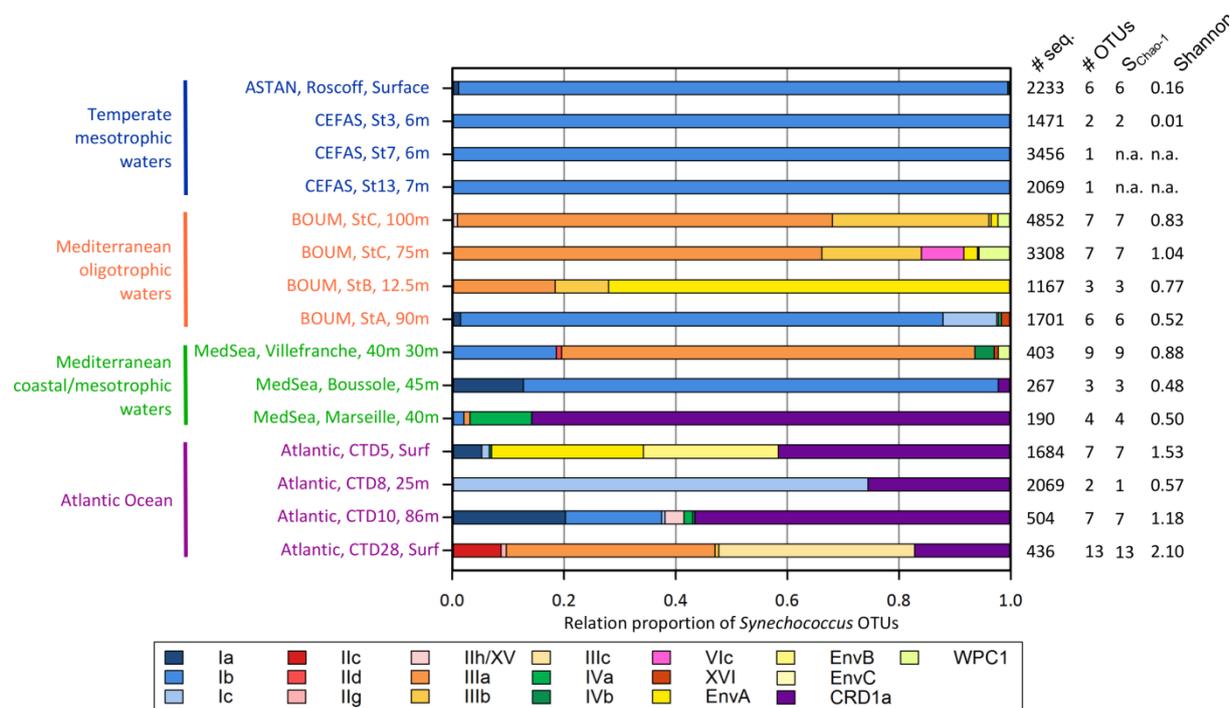
Une approche par pyroséquençage ciblant le gène *petB* a été retenue dans le but d'établir des corrélations entre la diversité pigmentaire qui sera déterminée dans ces échantillons par séquençage des régions PBS et la diversité génétique des *Synechococcus* marins composant ces échantillons. Le traitement a été réalisé comme décrit précédemment. L'ensemble des séquences obtenues appartient à des clades ou sous-clades déjà décrits dans la littérature (Fig. 51). De manière similaire à l'échantillon SOMLIT-Astan prélevé au large de Roscoff, la dominance du sous-clade Ib a été confirmée en Mer du Nord.

**Tableau 5 : Paramètres physico-chimiques et abondances des cellules phytoplanctoniques dans les différents échantillons environnementaux analysés**

Cruise	Sample	Region	Station Lat (Dec N)	Station Lon (Dec E)	Date	Time (GMT)	Sample depth (m)	Temperature (°C)	Salinity (psu)	DCM (m)	<i>Synechococcus</i> cells mL <sup>-1</sup>	<i>Prochlorococcus</i> cells mL <sup>-1</sup>	Picoeukaryotes 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 C	Mediterranean 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
BOUM	Station C	Mediterranean Sea	33.7125	32.6527	29-Jun-08	12:45	75	17.914	39.41	108	4.48E+03	-	5.26E+02
BOUM	Station B	Mediterranean 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 A	Mediterranean 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
-	Villefranche (pointB)	Mediterranean Sea	43.6833	7.3167	02-Jun-09	09:00	40	15.68	37.85	-	3.38E+04	4.33E+03	1.32E+03
-	Villefranche (pointB)	Mediterranean Sea	43.6833	7.3167	02-Jun-09	09:00	30	17.326	37.75	-	3.32E+04	-	1.61E+03
-	Nice (Boussole)	Mediterranean Sea	43.3667	7.9000	16-Jun-09	n.d.	45	n.d.	n.d.	-	1.80E+04	-	3.95E+03
-	Marseille (Frioul)	Mediterranean Sea	43.2417	5.2917	04-Jun-09	07:00	40	15.17	37.8513	-	3.49E+04	1.87E+03	2.09E+03
-	Marseille (Frioul)	Mediterranean 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	n.d.	n.d.	n.d.
Atlantic transect	Station 10	Atlantic Ocean	41.0009	-20.0086	24-Jul-11	16:30	86	14.37	35.92	55	n.d.	n.d.	n.d.
Atlantic transect	Station 28	Atlantic Ocean	23.3423	-27.1358	31-Jul-11	12:40	Surf	24.13	37	113	n.d.	n.d.	n.d.

Il est intéressant de noter que dans tous ces échantillons, aucune séquence appartenant au clade IV n'a été obtenue, alors que la codominance de ces clades a été rapportée de manière quasi-systématique, notamment dans les eaux froides de l'Atlantique Nord (Zwirgmaier et al., 2008; Mazard et al., 2012a). Une étude précédente menée au site SOMLIT-Astan avait également rapporté la présence du clade IV dans la Manche, mais en plus forte proportion que dans notre étude (~ 20%; Mazard, 2008), suggérant l'existence des variations saisonnières des proportions relatives de ces clades, comme cela a été rapporté dans les eaux côtières du Pacifique (Tai et Palenik, 2009).

La prévalence du clade III, incluant les sous-clades IIIa et IIIb, dans les eaux oligotrophes du bassin Est Méditerranéen confirme les résultats que nous avons obtenu en utilisant d'autres approches moléculaires, telles que les Dot-Blot ou bibliothèques de clones ciblant la région intergénique située entre l'ARNr 16S et 23S (voir Mella et al., 2012 en Annexe A). Le clade I domine en profondeur au niveau de la station A, présentant un régime déjà assez oligotrophe.



**Figure 51** : Abondance relative des différents clades de *Synechococcus*, en se basant sur l'analyse du gène *petB* à une résolution taxonomique de 94% d'identité. Le nombre de séquences utilisées pour la détermination des indices de diversité sont montrées à droite.

La composition en clades des communautés retrouvée au niveau de l'océan Atlantique Nord est par contre un peu plus surprenante. En effet, des clades pour lesquels aucun représentant en culture n'est actuellement disponible, dénommés Env, ont été rapportés. Le clade EnvA est retrouvé en forte proportion en surface au niveau de la station B (campagne BOUM) et de la station 5 (Atlantique). La présence de ce clade au niveau de la zone de transition entre les eaux mésotrophes

et oligotrophes, où il représentait environ 1/3 de la communauté totale de *Synechococcus*, avait déjà été rapportée (Mazard et al., 2012a). De plus, des données non publiées de l'équipe de David Scanlan montrent le clade EnvA domine spécifiquement en surface dans les régions tempérées, suggérant que les représentants de ce clade pourraient être adaptés aux fortes intensités lumineuses (Fig. 51).

Nos résultats soulignent une large distribution du clade CRD1. Ce clade, initialement décrit au niveau des remontées d'eaux froides du Costa Rica, très riches en nutriments (Saito et al., 2005), a ici été retrouvé dans tous les échantillons de l'océan Atlantique Nord analysés au cours de cette étude. De manière intéressante, ce clade domine également dans les environnements ultra-oligotrophes de l'Océan Pacifique Sud (Ostrowski, M. & Scanlan, D., comm. pers.) et sa présence avait également été rapportée dans l'Océan Atlantique (Mazard et al., 2012a). Ce profil de répartition suggère que ce clade pourrait être ubiquiste et ainsi posséder une grande versatilité métabolique (Fig. 52). Par ailleurs, l'aspécificité des sondes ciblant les clades V/VI/VII utilisées pour des analyses Dot-Blot ou la mauvaise résolution taxonomique entre les clades CRD1 et VII pourraient être à l'origine de la vision biaisée de la biogéographie décrite dans les études précédentes (Zwirgmaier et al., 2007; Zwirgmaier et al., 2008; Mella-Flores et al., 2011; Huang et al., 2012).

De très nombreuses études se sont attachées à la description de la diversité génétique de *Synechococcus*, en utilisant un grand nombre de marqueurs génétiques (voir Chapitre I). Au cours de cette thèse, nous avons également utilisé le gène marqueur l'ITS 16S-23S, et mis en évidence de nouveaux clades au niveau du sub-cluster 5.1 (MS1 et MS2) et du sub-cluster 5.3 (MS3 et MS4; Annexe A). Parallèlement à cette étude, Huang et collaborateurs ont également décrit de nouveaux clades en utilisant également l'ITS 16S-23S (Huang et al., 2012). Les analyses phylogénétiques que j'ai réalisé, regroupant des représentants de l'ensemble des clades actuellement décrits dans la littérature en utilisant ce marqueur génétique (Chapitre I Figs. 13 et 14) montrent par ailleurs que le sub-cluster 5.3 pourrait présenter une diversité génétique beaucoup plus importante qu'on ne le pensait initialement. Le gène marqueur *petB* offre l'avantage d'une plus grande facilité d'analyse et permet de mettre en évidence des patrons de distribution bien distincts vis-à-vis des autres marqueurs précédemment utilisés (ARNr16S, ITS, rpoC1, etc.; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008; Huang et al., 2012; Mazard et al., 2012a). La multiplication du nombre de gènes marqueurs (voir Tableau 1) et des environnements étudiés pose le problème de la comparaison des résultats obtenus dans ces différentes études. Dans ce contexte, il apparaît nécessaire de mener une étude qui évaluerait la pertinence des résultats en comparant ces différents gènes marqueurs.





---

## CHAPITRE III

# DIVERSITÉ PIGMENTAIRE DES SOUCHES EN CULTURE DE *SYNECHOCOCCUS*

---



---

---

# APPORTS DE LA GÉNOMIQUE COMPARATIVE A L'ÉTUDE DE LA PIGMENTATION CHEZ *SYNECHOCOCCUS*

---

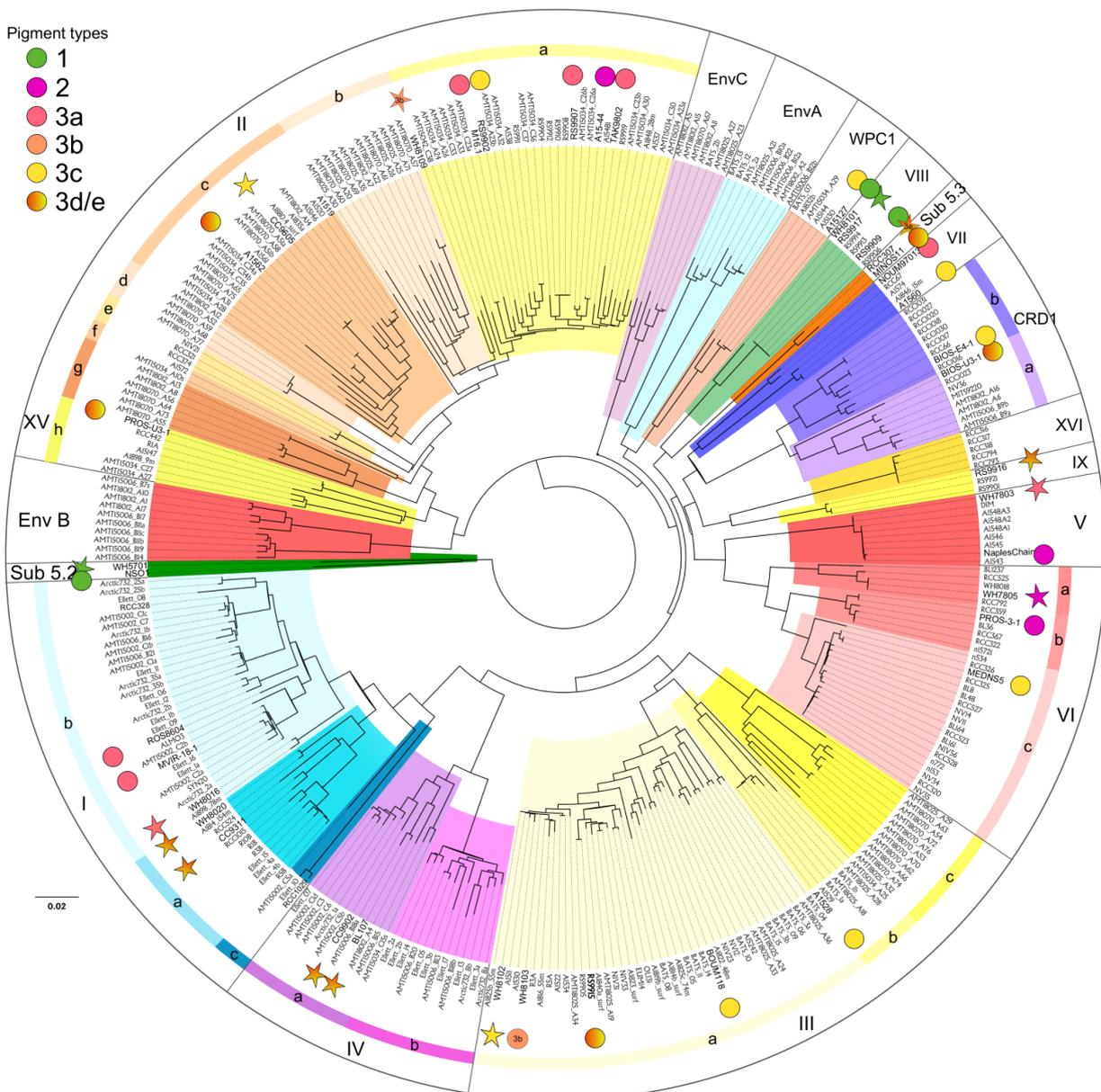
## I. CONTEXTE DU PROJET

Les comparaisons génomiques effectuées sur les 11 premiers génomes de *Synechococcus* ainsi que sur les 12 génomes de *Prochlorococcus*, disponibles en 2009, ont permis la détermination du nombre de gènes communs aux représentants de ces deux genres, de gènes spécifiques de chacun de ces groupes, mais également la mise en évidence de très nombreux gènes uniques retrouvés au sein des différents génomes, souvent regroupées au sein de régions hypervariables appelées îlots génomiques (voir §III.1.3; Kettler et al., 2007; Dufresne et al., 2008; Scanlan et al., 2009). Les études de génomique comparative ont également permis de mieux comprendre certains processus métaboliques comme les gènes impliqués dans l'acquisition des nutriments (azote, phosphore, métaux traces, etc.; Ostrowski et al., 2010), dans la capture de l'énergie lumineuse (Six et al., 2007b) ou encore dans les processus de régulation génique (Scanlan et al., 2009). Pour *Prochlorococcus*, la génomique comparative a de plus permis de proposer des scénarios évolutifs auxquels ce genre aurait été soumis (Partensky et Garczarek, 2010; Larsson et al., 2011; Luo et al., 2011b). La plus grande diversité génétique retrouvée au sein du genre *Synechococcus* rend plus compliqué d'établir de telles hypothèses concernant l'évolution de ce genre, notamment en raison du fait qu'avant cette thèse seul un génome de *Synechococcus* par clade avait été séquencé, à l'exception du clade IV (Scanlan et al., 2009). Cette limitation n'a notamment pas permis d'identifier des gènes spécifiques à chacune des lignées, qui auraient permettraient d'établir les bases moléculaires de la notion d'écotype chez *Synechococcus*.

Afin de mieux comprendre les spécificités moléculaires inhérentes aux génomes de *Synechococcus* et les scénarios évolutifs associés, 25 nouveaux génomes de *Synechococcus* ont été séquencés au cours de cette thèse dans le cadre des projets METASYN (en collaboration avec le Génoscope) et PELICAN (ANR Génomique microbienne). Dans ce contexte, j'ai largement participé, avec Morgane Ratin et Christophe Six, au clonage de plus d'une soixantaine de souches de *Synechococcus*. Ces cultures clonales sont désormais disponibles dans la Collection de Cultures de Roscoff. J'ai également effectué les extractions d'ADN sur les 25 souches retenues pour le

séquençage (Tableau 6), après la mise au point d'un protocole permettant l'extraction d'ADN à partir de grands volumes de culture.

Les différents génomes de *Synechococcus* séquencés au cours de cette thèse ont été choisis de manière à mieux représenter la diversité génétique au sein de cette radiation marine, en incluant notamment un nombre de clades et surtout de sous-clades bien plus important que pour les 11 premiers génomes ayant servi à la majorité des comparaisons génomiques publiées à ce jour par mon équipe d'accueil (Fig. 53 et Tableau 6). Le type pigmentaire et l'origine géographique des souches ont également été pris en compte.



**Figure 53** : Arbre phylogénétique en Neighbor-Joining basé sur l'analyse des séquences *petB* obtenues à partir d'échantillons environnementaux et d'isolats en culture. Les souches de *Synechococcus* dont le génome est entièrement séquencé sont montrées en gras. Les souches nouvellement séquencées dans le cadre du projet METASYN sont indiquées par des étoiles (n=14), les souches plus anciennes par des ronds. La couleur des symboles fait référence au type pigmentaire des souches, indiqué en haut à gauche du graphique par un code, tel que défini dans Humily et al., soumis (article II). Les souches CB0101, CB0205 et PCC 7001 du sub-cluster 5.2 ne figurent pas sur cet arbre. *Synechococcus* sp. WH5701 a été utilisé comme groupe externe.

Au cours de ce chapitre, les différents protocoles mis au point sont présentés. Les informations relatives à chacune des souches séquencées sont présentées dans le Tableau 6. Ces nouveaux génomes seront utilisés pour de nombreuses comparaisons génomiques. Au cours de cette thèse une attention particulière a été portée aux gènes de la région PBS et une description de ces dernières régions au sein des génomes des nouvelles souches est proposée.

## **II. MATÉRIELS ET MÉTHODES**

### **II.1. Préparation de l'ADN génomique**

#### **II.1.1 Culture cellulaire**

Comme mentionné plus haut, les souches de *Synechococcus* utilisées dans les différentes études réalisées au cours de ma thèse font partie de la Collection de Cultures de Roscoff (RCC). Les différentes informations (site et date d'isolement, type pigmentaire, affiliation phylogénétique, etc.) concernant chaque souche sont accessibles à l'adresse <http://www.sb-roscoff.fr/Phyto/RCC>. Les différentes souches ont été cultivées en flacons de culture plastique contenant du milieu PCRS11 (Rippka et al., 2000a). L'eau de mer est reconstituée à partir de Red Sea Salts (Houston, Texas, USA) en utilisant de l'eau distillée. Le milieu qui, de base, contient 400  $\mu\text{M}$  de  $(\text{NH}_4)_2\text{SO}_4$  est supplémenté par 1 mM de  $\text{NaNO}_3$ . Il permet d'obtenir des rendements maximaux de l'ordre de 1 à  $5.10^8$  cellules de *Synechococcus* par mL. En entretien de routine à la RCC, les souches sont cultivées à 20 - 22°C avec une intensité lumineuse de 10 à 20  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  et repiquées toutes les 3 semaines.

#### **II.1.2 Purification des souches de *Synechococcus***

Les souches de *Synechococcus* ont été clonées par 3 passages successifs sur boîte de Pétri. Une solution d'Agarose Low Melting (LMP, Invitrogen) à 0,35% en milieu PCRS11 a été autoclavée, puis refroidie jusqu'à une température de 28-29°C. Différentes quantités de cellules (correspondant à des volumes de culture compris entre 0,05  $\mu\text{L}$  à 150  $\mu\text{L}$ ) ont été ajoutées à 35 mL de cette gélose molle avant mélange et coulage dans des boîtes de Pétri. Après solidification, les boîtes de Pétri ont été parafilmées afin d'éviter la dessiccation. Les boîtes ont été incubées dans les conditions décrites précédemment jusqu'à l'apparition des premières colonies (généralement de 2 à 4 semaines). Différentes colonies de *Synechococcus* ont ensuite été prélevées à la pipette Pasteur et différents flacons contenant 10 mL de milieu PCRS11 ont été inoculés. Le développement d'une colonie en milieu liquide prend de 7 à 15 jours. Le taux de contamination des cultures par des bactéries hétérotrophes a été vérifiée par cytométrie en flux (voir Chapitre II § II.1.2).



*D'autres techniques d'isolement couramment utilisées en microbiologie ont été testées au cours de cette thèse sans donner de résultats convaincants pour les *Synechococcus marinus*.*

**Tableau 6 : Souches de *Synechococcus* séquencées à ce jour et caractéristiques principales.**

Name	RCC #	Sub-cluster	Clade/Subclade	Pigment type <sup>a</sup>	Geographic Origin	Habitat	Lat. [°North]	Long. [°East]	Depth (m)	Isolation Date	Sequencing center	Genome status	Accession number	Genome size (bp)	Genes	Clusters <sup>b</sup>	GC %	
<b>Cyanobium</b>																		
NS01	2436	5.2		1	Atlantic Ocean, North Sea	-	41.67	2.80	-	-	Genoscope	WGS		2 727 770	2,901	2,775	67.3	
PCC7001	-	5.2		1	Atlantic Ocean, Long Island sound	coastal	?	?		14-May-05	JCVI	WGS	ABSE00000000	2 832 412	2,770	2,651	68.7	
WH5701	1084	5.2		1	Atlantic Ocean, Long Island sound	-	29.47	34.92	-	1-Jan-57	JCV	WGS	AANO00000000	3 051 556	3,372	3,003	66	
<b><i>Synechococcus</i></b>																		
A15-127	2378	5.1	WPC1	3c	Atlantic Ocean	-	-31.12	-3.92	45	22-Oct-04	Genoscope	WGS		2 537 768	2,791	2,687	60.9	
A15-28	2556	5.1	IIIb	3c	Atlantic Ocean, North Gyre	-	31.25	-20.74	15	25-Sep-04	Genoscope	WGS		2 336 026	2,691	2,605	60.2	
A15-44	2527	5.1	IIa	2	Atlantic Ocean, Moroccan upwelling	-	21.68	-17.83	20	1-Oct-04	Genoscope	WGS		2 586 506	3,120	2,977	52.7	
A15-60	2554	5.1	VII	3c	Atlantic Ocean	-	17.62	-20.97	10	4-Oct-04	Genoscope	WGS		2 537 019	3,097	3,001	59.1	
A15-62	2374	5.1	IIc	3dB	Atlantic Ocean	-	17.62	-20.97	15	4-Oct-04	Genoscope	WGS		2 277 102	2,779	2,701	61.3	
BIOS-E4-1	2534	5.1	CRD1b	3cA	Pacific Ocean, South-East	oligotrophic	-31.87	-91.44	40	28-Nov-04	Genoscope	WGS		3 296 041	4,293	4,007	61.3	
BIOS-U3-1	2533	5.1	CRD1b	3dA	Pacific Ocean, Chile upwelling	-	-33.98	-73.39	5	6-Dec-04	Genoscope	WGS		2 693 428	3,312	3,185	53.2	
BL107	515	5.1	IVa	3dA	Mediterranean Sea, Blanes bay	coastal	41.72	3.57	1800	-	JCVI	WGS	AATZ00000000	2 285 035	2,529	2,440	55	
BMK-MC-1	2438	5.1	V	2	Mediterranean Sea, Naples bay	coastal	40.81	14.25	23	13-Oct-09	Genoscope	WGS		2 619 408	3,056	2,922	58.9	
BOUM118	2379	5.1	IIIa	3c	Mediterranean Sea	oligotrophic	33.63	32.65	5	1-Oct-08	Genoscope	WGS		2 314 805	2,828	2,712	61.9	
CB0101	-	5.2	CB4	1	Atlantic Ocean, Chesapeake Bay	coastal	37.46	-76.14	-	-	TIGR	WGS	ADXL00000000	2 691 045	3,010	2,849	64.2	
CB0205	-	5.2	CB5	2	Atlantic Ocean, Chesapeake Bay	coastal	37.46	-76.14	-	-	TIGR	WGS	ADXM00000000	2 431 158	2,719	2,618	52	
CC9311	1086	5.1	Ia	3dA	Pacific Ocean, Californian current	coastal	32.00	-124.52	95	1-Jan-93	JGI	Complete	CP000435	2 606 749	2,931	2,785	52.5	
CC9605	753	5.1	IIc	3c	Pacific Ocean, Californian current	oligotrophic	30.42	-123.98	-	1-Jan-96	JGI	Complete	CP000110	2 510 660	2,709	2,555	59.2	
CC9902	2673	5.1	IVa	3d	Pacific Ocean, Californian current	-	32.90	-117.25	--	1-Jan-99	JGI	Complete	CP000097	2 234 829	2,387	2,324	54.2	
M16.1	791	5.1	IIa	3a	Atlantic Ocean, Mexico Gulf	-	27.70	-91.30	275	9-Feb-04	Genoscope	WGS		2 109 296	2,468	2,412	60.3	
MEDNS5	2368	5.1	Vlc	3c	Mediterranean Sea	-	41.00	6.02	80	1-Jul-93	Genoscope	WGS		2 422 988	2,701	2,614	61.8	
MINOS11	2319	5.3	5.3-I	3dA	Mediterranean Sea	pelagic	34.00	18.02	20	19-Jun-96	Genoscope	WGS		2 290 218	2,784	2,700	46.8	
MVIR-18-1	2385	5.1	Ib	3aA	North Sea	-	61.00	1.98	25	23-Jul-07	Genoscope	WGS		2 433 324	3,052	2,971	52.8	
NOUM97013	2433	5.1	VII	3a	Equatorial Pacific Ocean	pelagic	-22.33	166.33	0	19-Nov-96	Genoscope	WGS		2 541 016	2,896	2,812	57.4	
PROS-7-1	2381	5.1	Vib	2	Mediterranean Sea	pelagic	37.40	15.62	5	26-Sep-99	Genoscope	WGS		2 552 284	2,852	2,761	58.2	
PROS-U-1	2369	5.1	IIh (XV)	3dB	Atlantic Ocean, Moroccan upwelling	pelagic	30.13	-10.07	5	12-Sep-99	Genoscope	WGS		2 559 798	3,072	2,992	59.3	
RCC307	307	5.3	5.3-I	3e	Mediterranean Sea	oligotrophic	39.17	6.17	15	26-May-96	Genoscope	Complete	CT978603	2 224 915	2,559	2,477	60.8	
ROS8604	2380	5.1	Ib	3a	Atlantic Ocean, Brittany coast	coastal	48.72	-3.98	0	24-Nov-86	Genoscope	WGS		2 842 043	3,707	3,537	55.8	
RS9902	2376	5.1	IIa	3c	Red Sea, Aqaba gulf	pelagic	29.47	34.94	1	29-Mar-99	Genoscope	WGS		2 452 432	2,955	2,846	60.7	
RS9907	2382	5.1	IIa	3a	Red Sea, Aqaba gulf	pelagic	29.47	34.92	10	23-Aug-99	Genoscope	WGS		2 559 783	2,898	2,793	59.2	
RS9909	2383	5.1	VIII	1	Red Sea, Aqaba gulf	pelagic	29.47	34.92	10	7-Sep-99	Genoscope	WGS		2 572 492	2,755	2,605	62.7	
RS9915	2553	5.1	IIIa	3d	Red Sea, Aqaba gulf	pelagic	29.47	34.94	10	18-Oct-99	Genoscope	WGS		2 403 460	2,834	2,705	59.6	
RS9916	555	5.1	IX	3dA	Red Sea, Aqaba gulf	pelagic	29.47	34.94	10	22-Nov-99	JCVI	WGS	AAUA00000000	2 664 874	2,994	2,828	59.8	
RS9917	556	5.1	VIII	1	Red Sea, Aqaba gulf	pelagic	29.47	34.92	10	22-Nov-99	JCVI	WGS	AANP00000000	2 584 919	2,795	2,612	64.8	
TAK9802	2528	5.1	IIa	3a	Pacific Ocean, Takapoto Atoll	coastal	-14.50	-145.33	7	6-Feb-98	Genoscope	WGS		2 209 584	2,651	2,557	61.4	
WH7803	28	5.1	V	3a	Atlantic Ocean, Sargasso Sea	pelagic	33.73	-67.48	25	3-Jul-78	Genoscope	Complete	CT971583	2 366 981	2,587	2,493	60.2	
WH7805	1085	5.1	VIa	2	Atlantic Ocean, Sargasso Sea	-	33.73	-67.50	-	30-Jun-78	JCVI	WGS	AAOK00000000	2 627 147	2,904	2,765	57.5	
WH8016	2535	5.1	Ib	3a	Atlantic Ocean, East coast USA	pelagic	41.52	-70.67	-	1-Jun-80	JGI	WGS	AGIK00000000	2 695 643	2,992	2,873	54.1	
WH8020	2437	5.1	Ia	3dA	Sargasso Sea	-	38.68	-69.34	50	26-Jun-80		Draft		2 290 798	3,140	3,034	53.3	
WH8101	2555	5.1	VIII	1	Atlantic Ocean, Woods hole Bay	-	41.52	-70.67	-	-	Genoscope	WGS		2 600 817	2,860	2,722	63.8	
WH8102	539	5.1	IIIa	3c	Atlantic Ocean, Caribbean Sea	pelagic	22.48	-65.62	15	15-Mar-81	JGI	Complete	BX548020	2 434 429	2,808	2,627	59.4	
WH8103	2366	5.1	IIIa	3bB	Atlantic Ocean	-	28.50	-67.40	-	-	Genoscope	WGS		2 398 638	2,803	2,683	60.3	
WH8109	2033	5.1	IIa	3bB	Atlantic Ocean, Sargasso Sea	pelagic	39.48	-70.49	-	1-Jun-81	JCVI	WGS		2 118 903	2,577	2,493	60.1	

<sup>a</sup> Following the nomenclature proposed in Humily et al., submitted to *Environmental Microbiology*.<sup>b</sup> Refers to the number of protein ortholog families calculated in the Cyanorak database accessible at <http://www.sb-roscoff.fr/Phyto/cyanorakk>

### II.1.3 Extraction d'ADN génomique de souches du genre *Synechococcus*

L'ADN génomique a été extrait de 2L de culture de *Synechococcus*, culottée après un suivi quotidien par cytométrie en flux lorsque le taux de contamination de la culture était minimum. Les culots bactériens ont été conservés à -80°C jusqu'à extraction. Le culot a été resuspendu dans 4,575 mL de TE (10 mM Tris pH 8.0, 1 mL EDTA pH 8.0). Les cellules ont ensuite été lysées par ajout de 100 µL de lysozyme (50 mg.mL<sup>-1</sup>, concentration finale 1 mg.mL<sup>-1</sup>). La solution a été incubée 30 min à 37°C sous agitation (500 rpm). Après 30 min d'incubation, 100 µL de lysozyme ont été ajoutés. Les cellules ont ensuite été lysées par ajout de 125 µL de SDS (20% w/v) et 50 µL de protéinase K (20 mg.mL<sup>-1</sup>) afin d'avoir des concentrations finales respectives de 0.5% SDS et 200 µg.mL<sup>-1</sup> de protéinase K. La solution a été incubée à 55 °C, pendant 5 h sous agitation orbitale (500 rpm). En fonction de la concentration cellulaire, 25 µL de protéinase K ont pu être ajoutés après 2 h 30 d'incubation. La solution de cellules lysées a ensuite été placée sur glace et 850 µL de NaCl 5M et 650 µL de TE ont été ajoutés. Les acides nucléiques ont ensuite été extraits par ajout d'un volume de phénol : chloroforme : alcool isoamylique (25 :24 :1 v/v) en utilisant des tubes Phase Lock Gel, puis centrifugés à 5000 g, durant 10 min à 15°C. Les tubes Phase Lock Gel utilisent une résine hydrophobe, se plaçant à l'interface des phases aqueuse et organique. Les traces résiduelles de phénol ont été éliminées par ajout d'un volume de chloroforme : alcool isoamylique (24 :1 v/v), puis centrifugés à 5000 x g, 10 min, 15°C. Cette étape a été répétée 3 fois, et l'absence de phénol a été vérifiée au Nanodrop ND-100 (ThermoScientific).

Les ARNs ont été dégradés par ajout de 1/100<sup>ème</sup> de volume de RNase A (10 mg.mL<sup>-1</sup> afin d'obtenir une concentration finale de 100 µg.mL<sup>-1</sup>). La solution a été incubée à 37°C durant 30 min. L'ADN a ensuite été extrait par ajout d'un volume de phénol : chloroforme : alcool isoamylique (25 :24 :1 v/v), en tubes Phase Lock Gel, puis centrifugés. Toute trace résiduelle de phénol a été éliminée par ajout d'un volume de chloroforme : alcool isoamylique (24 :1 v/v), puis centrifugation. La phase aqueuse supérieure a été transférée dans un nouveau tube et 0.6 vol d'isopropanol froid (-20°C) ont été ajoutés. Les tubes ont été mélangés par retournement et les acides nucléiques ont été précipités une nuit à -20°C. L'ADN a été récupéré par centrifugation à 5000 x g, durant 30 min à 4°C. Les ADN ont ensuite été lavés par ajout respectif de 1 mL d'éthanol 100 % puis 70 %, puis centrifugés à 7500 x g, 5 min, 4°C. Le culot d'ADN a été resuspendu dans 4 mL de TE, au moins une nuit à 4°C. Les petits fragments d'ADN génomique et les ARNs résiduels ont ensuite été éliminés en utilisant des Amicon Ultra-4 (100 000 MWCO), en procédant à 3 lavages avec du TE, puis centrifugés à 2500 g, 10 min à 4°C. La quantité d'ADN a ensuite été évaluée au Nanodrop ND-100 et par dosage au Picogreen (Invitrogen), et la qualité a été vérifiée sur gel d'agarose 1%. Les échantillons ont été conservés à -80°C.

## **II.2. Analyses phylogénétiques des opérons codant pour les sous-unités des phycobiliprotéines**

Des analyses phylogénétiques ont été réalisées à partir des séquences *r/cpcBA*, *cpeBA* et *mpeBA*, codant respectivement pour les sous-unités  $\alpha$  et  $\beta$  de la PC, PEI et PEII, de l'ensemble des génomes de *Synechococcus* marins. Ces séquences ont été utilisées afin de coprendre l'histoire évolutive des phycobiliprotéines au sein de la radiation marine des *Synechococcus*. Les séquences nucléotiques de *mpeBA* et *cpeBA* ont été alignées à l'aide de MEGA 5 (Tamura et al., 2011). Les analyses en inférence bayésienne ont été réalisées sous MrBAYES v 3.1.2 (Huelsenbeck et Ronquist, 2001), en utilisant un "codon model", en démarrant 4 chaînes à partir d'un arbre aléatoire et en réalisant 1,5 million et 2 millions de générations pour *mpeBA* et *cpeBA* respectivement. Afin de n'utiliser que les chaînes stables 3750 et 5000 arbres ont été éliminés. Pour les analyses en ML réalisées avec PhyML (Guindon et al., 2009), les modèles de substitution nucléotique GTR+G et GTR+I+G ont été sélectionnés à l'aide de JModeltest pour *mpeBA* et *cpeBA* respectivement (Darriba et al., 2012). La robustesse des topologies a été évaluée en réalisant 1000 bootstraps. Une topologie bayésienne est montrée pour ces deux gènes.

L'analyse phylogénétique de *r/cpcBA* a été réalisée à partir des séquences protéiques en réalisant 2 millions de générations et en éliminant 5000 arbres. Le modèle LG+G a été sélectionné pour les analyses en ML.

### III. RÉSULTATS ET DISCUSSION

#### III.1. Caractérisation des gènes impliqués dans la capture de lumière au sein des génomes de *Synechococcus*

Certaines caractéristiques originales ont pu être mise en évidence à partir de ces nouveaux génomes, comme la taille du génome BIOS-E4-1, possédant le génome le plus grand de tous les *Synechococcus* marins actuellement séquencés (3,29 Mpb; Annexe E). Les génomes les plus grands au sein du clade 5 (*sensu* Herdman et al., 2001) étaient préalablement affiliés au sub-cluster 5.2 ou au genre *Cyanobium*.

Les régions PBS des différents génomes ont été annotées afin de préciser ou redéfinir le set de gènes spécifique à chacun des types pigmentaires, notamment pour les types 2 et 3a dont un seul génome était préalablement disponible (voir Fig. 33). L'organisation et la synténie des régions PBS des nouvelles souches apparaît très similaire aux régions précédemment décrites par Six et al., 2007b), à l'exception d'un nouveau type, décrit dans l'article ci-après. Cette région PBS ressemble beaucoup à celle des souches de type 3c, WH8102 et CC9605, mais possède 3 gènes supplémentaires.

L'extrémité 3' de la région PBS apparaît très conservée entre le gène de fonction inconnue *unk13* et le gène *ptpA* qui code pour une tyrosine-phosphatase (Fig. 33), à l'exception du gène de lyase situé en amont de ce dernier. Cette lyase peut être l'hétéroduplexe *CpcE/F* ou *RpcE/F* ou la protéine de fusion *RpcG* (Fairchild et al., 1992; Blot et al., 2009). L'extrémité 5', où plusieurs gènes de fonction inconnue sont présents, est également similaire entre toutes les souches possédant de la PEII, sauf pour le gène *unk3* qui n'est présent chez certaines souches (données non montrées). Aucune spécificité n'a été mise en évidence concernant les gènes codant pour les sous-unités  $\alpha$  et  $\beta$  des PBPs, où chacun des résidus cystéine apparaît conservé. Ces gènes ont été utilisés pour des analyses phylogénétiques (voir ci-après et Article II).

Parmi les gènes jouant un rôle important dans la dynamique et la chromophorylation des PBS, une attention particulière a été portée aux polypeptides de liaison (ou "linkers"; Tableau 7) ainsi qu'aux phycobilines-lyases (Tableau 8). Les analyses comparatives ont permis de confirmer la présence chez l'ensemble des souches des linkers participant à la structure du cœur du PBS (*apcC* et *apcE*) et de la liaison PC-cœur (*cpcG1*, *cpcG2*). Les souches du type pigmentaire 1 possèdent de plus 2 à 3 copies de *cpcC* et possèdent également *cpcD*. Certaines souches à PE ont une troisième copie du gène *cpcG* (*cpcG3*), dont la fonction reste inconnue. Toutes les souches à PE possèdent des gènes de linkers de PEI (*cpeC*, *cpeE* ainsi que *mpeD*). *MpeD* a été initialement décrite comme résultant de la

fusion entre un linker de PEI (extrémité C-terminale non chromophorylée) et un linker de PEII (extrémité N-terminale chromophorylée), qui assurerait la liaison entre ces deux types de phycobiliprotéines dans les bras des PBS (Six et al., 2005). Cependant, la présence chez les souches du type 2 (dépourvues de PEII) d'un homologue de MpeD, dont la partie N-terminale est très modifiée par rapport aux souches de type 3, et notamment ne comporte pas de site de fixation du chromophore, reste étonnante et pourrait indiquer un rôle plus global de ce polypeptide de liaison dans la structure des hexamères de PE.

**Tableau 7 : Gènes codant pour les polypeptides de liaison présents dans les différents génomes**

Strain	Pigment type	APC		PC			PEI		PEII				
		<i>apcC</i> (LC)	<i>apcE</i> (LCM)	<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>Cya.</i> PCC7001	1	CC	CC	3 <sup>Ca</sup>	RC	2 <sup>G</sup>	-	-	-	-	-	-	-
<i>Cya.</i> NS01	1	CC	CC	3 <sup>Ca</sup>	RC	2 <sup>G</sup>	-	-	-	-	-	-	-
<i>Cya.</i> WH5701	1	CC	CC	3 <sup>Ca</sup>	RC	2 <sup>G</sup>	-	-	-	-	-	-	-
CB0101	1	CC	CC	2 <sup>Cb</sup>	RC	2 <sup>G</sup>	-	-	-	-	-	-	-
RS9917	1	CC	CC	2 <sup>Cc</sup>	RC	2 <sup>G</sup>	-	-	-	-	-	-	-
RS9909	1	CC	CC	2	RC	2 <sup>G</sup>	-	-	-	-	-	-	-
WH8101	1	CC	CC	2	RC	2 <sup>G*</sup>	-	-	-	-	-	-	-
CB0205	2	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	-	-	-
A15-44	2	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	-	-	-
BMK-MC-1	2	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	-	-	-
WH7805	2	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	-	-	-
PROS-7-1	2	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	-	-	-
WH8016	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	NC	-	-
MVIR-18-1	3a	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	RC	NC	-
ROS8604	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	NC	-	-
M16.1	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	RC	-	-
TAK9802	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	NC	-	-
RS9907	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	-	NC	-	-
WH7803	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC	-	-
NOUM97013	3a	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	NC	-	-
RS9902	3c	CC	G	-	-	2 <sup>G</sup>	RC	RC	RC	RC	NC	-	-
CC9605	3c	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	NC	-	-
BOUM118	3c	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	-
WH8102	3c	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	-
A15-28	3c	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	-
MEDNS5	3c	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	-	-	-
A15-60	3c	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*** / NC	-	-
BIOS-E4-1	3c	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC**	NC	-
A15-127	3c	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	RC*	-	NC
WH8103	3b	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	-
WH8109	3b	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	NC	-	NC
A15-62	3d	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	NC	-	-
PROS-U-1	3d	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	NC	-	NC
RS9915	3d	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	-
MINOS11	3d	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	NC
RCC307	3e	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC*	-	NC
BIOS-U3-1	3d	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC**	NC	-
CC9311	3d	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	RC	NC	-
WH8020	3d	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	RC**	NC	-
BL107	3d	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	RC*	NC	-
CC9902	3d	CC	CC	-	-	3 <sup>G</sup>	RC	RC	RC	RC	RC*	NC	-
RS9916	3d	CC	CC	-	-	2 <sup>G</sup>	RC	RC	RC	RC	RC	-	NC

CC : gène présent au sein du cluster de gènes codant pour le coeur du PBS; RC : gène présent au sein de la région PBS, RC\* : localisé quelques gènes en amont de la région PBS; RC\*\* : en amont de *cpcG2*; RC\*\*\* : en aval de *cpcG2*; NC : gène présent en dehors de la région PBS. <sup>Ca</sup> RC (*cpcI*) RC (*cpcII*) NC (*cpcIII*); <sup>Cb</sup> RC (*cpcI*) NC ; <sup>Cc</sup> RC (*cpcI*) RC (*cpcII*); 2<sup>G</sup> (*cpcG1*) RC (*cpcG2*); 2<sup>G\*</sup> GC (*cpcG1*) RC (*cpcG2*); 3<sup>G</sup> GC (*cpcG1*) RC (*cpcG2*) NC (*cpcG3*).

Une grande variabilité existe au niveau du nombre et de la nature des linkers PEII entre les souches qui possèdent deux PE dans les bras des PBS (Tableau 7). Les types pigmentaires 3a ont un ou plus rarement deux linkers PEII (Tableau 7), sauf MVIR-18-1, qui en a trois mais qui ne correspond pas en réalité à une souche de type 3a "typique" (voir article ci-après). Sous le terme "MpeE" sont regroupées différentes familles de linkers dont les plus divergents sont trouvés chez ROS8604 et WH8016. La souche A15-60 par exemple a deux linkers distincts (donc paralogues) qui sont regroupés au sein du groupe "MpeE". MpeF et MpeG sont deux polypeptides de liaison retrouvés en dehors de la région PBS, généralement trouvés chez les souches capables d'AC4 et qui avaient été définis lorsque le set de génomes était plus restreint. Au vu de la grande diversité existant au sein de ces groupes, il apparaît désormais nécessaire de faire une étude spécifique poussée (biochimique, structurale, évolutive) des linkers PEII, un peu comme nous l'avons fait avec les phycobilines lyases (voir ci-dessous et article Bretaudeau et al., 2012 en annexe B), pour établir une délimitation plus claire au sein des groupes "MpeE", MpeF et MpeG.

Les recherches sur les phycobilines-lyases ont fait l'objet de nombreuses publications et la fonction de plusieurs d'entre elles a été déterminée au cours de ces 5 dernières années, notamment grâce à l'utilisation de systèmes d'expression hétérologues. L'ensemble des génomes de cyanobactéries actuellement disponibles et la connaissance de la pigmentation des PBS de certaines des souches séquencées nous a permis de proposer un certain nombre d'hypothèses quant à la fonction de certaines de ces protéines non encore caractérisées. Par exemple, MpeY n'étant retrouvée qu'au sein des souches possédant de la PEII, cette dernière pourrait agir spécifiquement sur cette PBP, bien que cette spécificité reste à démontrer. De même, l'absence de MpeU chez MVIR-18-1, qui a par ailleurs toutes les autres lyases trouvées chez les adaptateurs chromatiques (voir Tableau 8 et article ci-après) se traduit apparemment par un faible rapport PUB/PEB (type pigmentaire 3a), ce qui pourrait indiquer un rôle clé pour MpeU dans la présence de PUB dans les bras des PBS. Cette enzyme étant systématiquement présente au sein des souches de type pigmentaire 3b, 3c ou 3d, présentant un rapport  $Exc_{495:545nm}$  plus élevé, ce résultat nous amène à suggérer que MpeU serait elle aussi une phycobiline lyase-isomérase, mais seules des études d'expression hétérologue permettraient d'établir sa spécificité exacte.

Un des résultats majeurs de ces comparaisons génomiques aura été la mise en évidence d'un cluster de trois gènes dans des régions PBS dont la composition et l'organisation génique est par ailleurs identique à celle des souches de type 3c, préalablement décrites (Six et al., 2007). Ce cluster de gènes est caractérisé par la présence d'une nouvelle phycobiline lyase, que nous avons nommée MpeW, et de deux autres gènes. En raison de l'homologie apparente entre MpeW et MpeZ, ainsi que

la présence de ce cluster au sein de souches capables d'AC4, nous a amené à suggérer son implication dans le processus d'AC4, ce qui fait l'objet de l'article qui suit.

**Tableau 8 : Gènes codant pour des phycobilines lyases putatives présents dans les différents génomes.**

Strain	Pig- ment type	Phycocyanin						Phycoerythrin I and/or II (CpeF)									
		<i>cpeEF</i>	<i>rpcEF</i>	<i>rpcG</i>	<i>cpcS</i>	<i>cpcT</i>	<i>rpcT</i>	<i>cpeS</i>	<i>cpeT</i>	<i>cpeU</i>	<i>cpeY</i>	<i>cpeZ</i>	<i>mpeV</i>	<i>mpeU</i>	<i>mpeY</i>	<i>mpeZ</i>	<i>mpeW</i>
<i>Cya</i> . PCC7001	1	RC	-	-	RC	RC	RC	-	-	-	-	-	-	-	-	-	-
<i>Cya</i> . NS01	1	RC	-	-	RC	RC	RC	-	-	-	-	-	-	-	-	-	-
<i>Cya</i> . WH5701	1	RC	-	-	RC	RC (2)	RC	-	-	-	-	-	-	-	-	-	-
CB0101	1	RC	-	-	RC	RC	RC	-	-	-	-	-	-	-	-	-	-
RS9917	1	RC	-	-	RC	RC	RC	-	-	-	-	-	-	-	-	-	-
RS9909	1	RC	-	-	RC	RC	-	-	-	-	-	-	-	-	-	-	-
WH8101	1	RC	-	-	RC	RC	-	-	-	-	-	-	-	-	-	-	-
CB0205	2	-	RC	-	RC	-	-	RC	RC	RC	RC	RC	RC	-	-	-	-
A15-44	2	-	RC	-	RC	RC	-	RC	RC	RC	RC	RC	RC	-	-	-	-
BMK-MC-1	2	-	RC	-	RC	RC	-	RC	RC	RC	RC	RC	RC	-	-	-	-
WH7805	2	-	RC	-	RC	RC	-	RC	RC	RC	RC	RC	RC	-	-	-	-
PROS-7-1	2	-	RC	-	RC	RC	-	RC	RC	RC	RC	RC	RC	-	-	-	-
WH8016	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
MVIR-18-1	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	NC	-
ROS8604	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
M16.1	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
TAK9802	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
RS9907	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
WH7803	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
NOUM97013	3a	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	-	RC	-	-
RS9902	3c	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
CC9605	3c	-	-	RC	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
BOUM118	3c	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
WH8102	3c	-	-	RC	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
A15-28	3c	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
MEDNS5	3c	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
A15-60	3c	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
A15-127	3c	-	-	RC	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	-	-
WH8103	3c	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	RC	-
WH8109	3b	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	RC	-
A15-62	3d	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	RC	-
PROS-U-1	3d	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	RC	-
RS9915	3d	-	RC	-	RC	-	RC	RC	RC	RC	RC	-	RC	RC	-	RC	-
MINOS11	3d	-	-	RC	RC	RC	-	RC	RC	RC	RC	-	RC	RC	-	RC	-
BIOS-E4-1	3c	-	-	RC	RC	-	RC	RC	RC	RC	RC	RC	RC	-	NC	-	-
RCC307	3b	-	RC	-	RC	RC	-	RC	RC	RC	RC	RC	RC	RC	RC	NC	-
BIOS-U3-1	3d	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	RC	RC	NC	-
CC9311	3d	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	RC	RC	NC	-
WH8020	3d	-	RC	-	RC	-	RC (2)	RC	RC	RC	RC	RC	RC	RC	RC	NC	-
BL107	3dA	-	-	RC	RC	-	RC	RC	RC	RC	RC	RC	RC	RC	RC	NC	-
CC9902	3dA	-	RC	-	RC	-	RC	RC	RC	RC	RC	RC	RC	RC	RC	NC	-
RS9916	3dA	-	-	RC	RC	RC	RC	RC	RC	RC	RC	RC	RC	RC	RC	NC	-

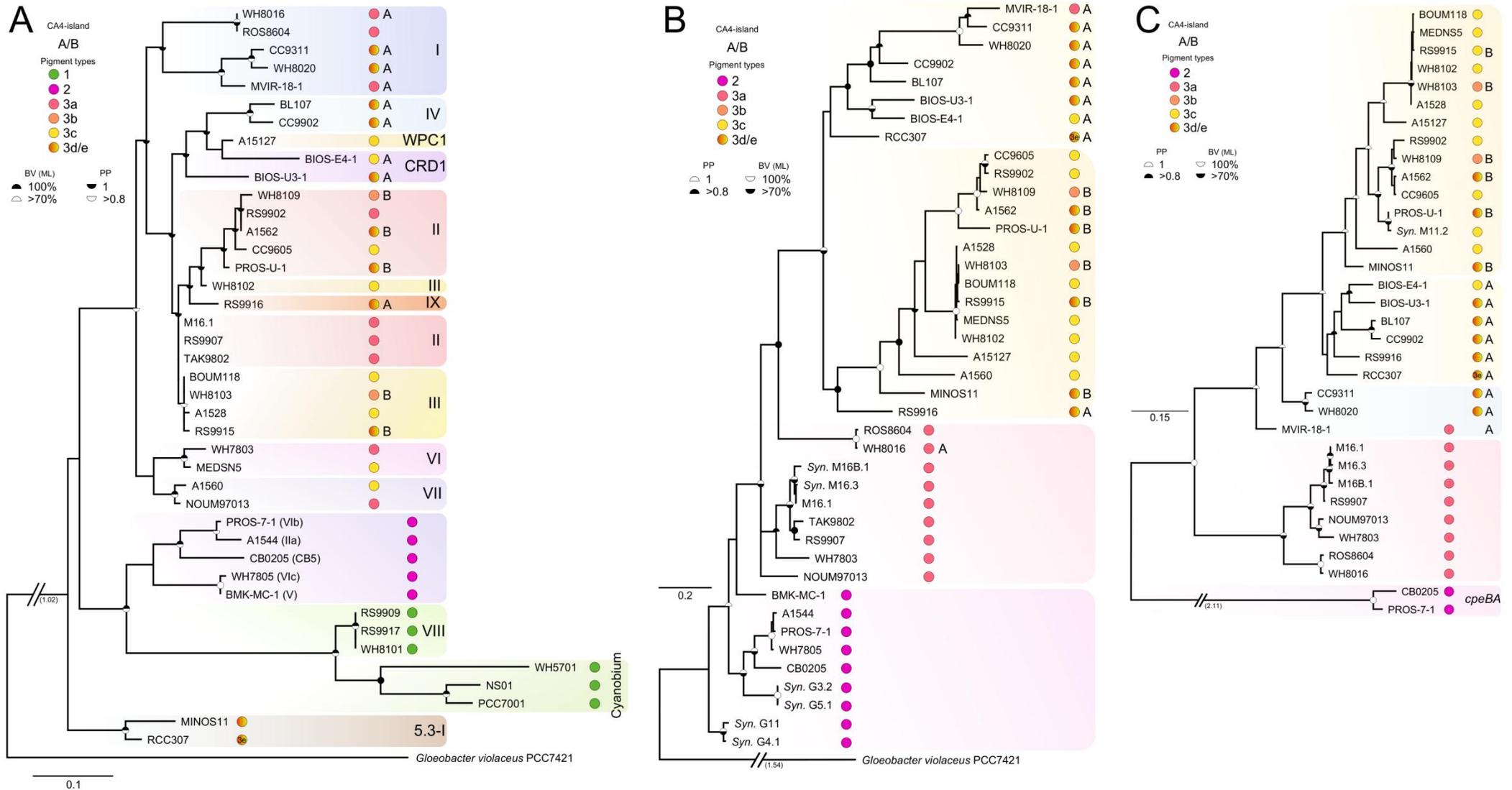
RC : gène présent au sein de la région; NC : gène présent en dehors de la région PBS. Le nombre entre parenthèses réfère au nombre de copies de chaque gène.

### III.2. Processus évolutifs des phycobiliprotéines

La topologie obtenue à partir de l'analyse de l'opéron *c/rpcBA* (codant les sous-unités de PC) sépare les souches vertes (type 1), des souches roses (type 2) et oranges (type 3; Fig. 54A). Cette analyse phylogénétique est congruente avec les analyses utilisant des gènes communs comme le *16S* ou *petB*, car toutes les souches possédant de la PEII sont regroupées selon leur affiliation phylogénétique. D'autre part les souches du subcluster 5.3 sont retrouvées à la base de l'arbre (Fig. 54A). Les phylogénies basées sur les sous-unités  $\alpha$  et  $\beta$  de la PEI et PEII sont congruentes et regroupent les souches de type 2 et de type 3a ensemble, à l'exception de MVIR-18-1, et ce indépendamment de leur affiliation phylogénétique comme cela avait été démontré dans des études précédentes (Six et al., 2007b; Everroad et Wood, 2012).

Les gènes liés à la synthèse et régulation de la PE présentent donc une histoire évolutive différente du reste du génome chez les *Synechococcus* marins (Fig. 54B et C) à l'inverse des gènes impliqués dans la synthèse du coeur d'APC ainsi que des extrémités basales des bras des PBS composés de PC (Fig. 54A). Cette incongruence est renforcée par la dispersion de types pigmentaires similaires au sein de toute la radiation des *Synechococcus* marins (Fig. 53). Ces analyses phylogénétiques corroborent les études précédentes et confirment que les processus évolutifs ont conduit à une complexification des PE, avec l'apparition de davantage de sites de liaison de chromophores et un accroissement du contenu en PUB de ces PBPs, afin de favoriser la capture des photons bleus dominant dans les milieux océaniques. Nous ne pouvons écarter l'hypothèse que l'acquisition de la phycobiline layse putative MpeU ait été concomitante avec les processus évolutifs précédemment décrits, et soit à l'origine du changement de propriétés spectrales des PE.

Nos résultats confirment que la duplication des gènes *mpeBA* a été préalable à l'évolution vers des PBS riches en PUB (Fig. 54C). Ce scénario évolutif a récemment été confirmé de manière statistique (Everroad et Wood, 2012). Everroad et Wood soulignent que les différents sous-types du type pigmentaire 3 formaient des clades phylogénétiques distincts, les amenant à proposer l'idée d'une évolution graduelle vers les phénotypes riches en PUB (Everroad et Wood, 2012). Les topologies des arbres phylogénétiques obtenus, corrélés avec les analyses de génomique comparative réalisées, suggèrent cependant que ces processus évolutifs sont beaucoup plus complexes que ce qui avait été initialement décrit (voir Article II; Everroad et al., 2006; Everroad et al., 2012; Six et al., 2007). Des analyses phylogénétiques plus complètes permettraient de mieux appréhender ces phénomènes et d'évaluer l'importance du transfert latéral de gènes dans les processus ayant conduit à la différenciation de deux types de régions PBS.



**Figure 54: Analyses phylogénétiques des opérons codant pour les sous-unités  $\alpha$  et  $\beta$  de (A) la phycocyanine (PC; *r/cpcBA*), (B) de la phycoérythrine I (PEI; *cpeBA*) et (C) de la phycoérythrine II (*mpeBA*) des différents génomes de *Synechococcus* séquencés.** Les topologies correspondent respectivement à des analyses en maximum de vraisemblance pour l'opéron *r/cpcBA* et inférence bayésienne pour *cpeBA* et *mpeBA*. Pour *cpeBA*, des séquences entières obtenus à partir d'isolats non séquencés, dont le nom commence par Syn., ont également été ajoutées à cette analyse. *Gloeobacter violaceus* PCC 7421 a été utilisé comme groupe externe pour les analyses phylogénétiques de la PC et PEI. Les séquences *cpeBA* de CB0205 et PROS-7-1 ont été utilisées pour enraciner l'arbre de *mpeBA*. Uniquement les valeurs de bootstraps >70% et les probabilités postérieures supérieures à 0.8 sont montrées.

---

---

## Article II : DIVERSITÉ ET ÉVOLUTION DU PHÉNOMÈNE D'AC4 CHEZ *SYNECHOCOCCUS*

---

### I. CONTEXTE DE L'ÉTUDE

Les phénomènes d'acclimation chromatique (AC) chez les cyanobactéries ont été principalement étudiés chez les souches d'eau douce. Ces cellules sont capables d'adapter la composition de leur antenne collectrice de lumière afin de mieux capturer les photons de lumière rouge ou verte (respectivement RL et GL). De nombreuses études se sont attachées à la compréhension de ces phénomènes, permettant de comprendre les bases biochimiques, ainsi que d'identifier les acteurs intervenant dans la régulation au niveau moléculaire. Certaines souches marines de *Synechococcus* sont également capables d'AC, dite de type IV, mais selon des mécanismes bien différents que ceux précédemment décrits chez les souches d'eau douce. Les bases biochimiques de l'AC4 n'impliquent pas un changement de la composition en PBPs du PBS, mais une chromophorylation différentielle de la PE, retrouvée aux extrémités des projections radiales des PBS (Everroad et Wood, 2006). Récemment l'implication de la phycobiline lyase MpeZ dans la CA4 a été démontré (Shukla et al., 2012). La transcription de MpeZ est induite en BL, et cette enzyme bifonctionnelle attache de la PEB au niveau de la Cys-83 de la PEII, et l'isomérisé en PUB (Shukla et al., 2012). Néanmoins, les bases moléculaires de la régulation de ce phénomène restent inconnues.

L'importance écologique de l'AC4 commence tout juste à être appréhendée. Plus d'un quart des souches de *Synechococcus* en culture sont capables de modifier la composition pigmentaire du PBS en réponse à la lumière verte et bleue. Le séquençage de nouveaux génomes de *Synechococcus* marins et leur comparaison doit permettre (1) de déterminer les bases moléculaires du phénomène d'AC4. Les souches capables d'AC4 sont retrouvées au niveau de clades phylogénétiques distincts permettant l'exploration des processus évolutifs associés (Fig. 53). Au cours de cette étude, présentée sous la forme d'un article dont je suis le premier auteur, nous avons également cherché à (2) évaluer la diversité phénotypique de l'AC4, en étudiant les cinétiques d'acclimation entre différentes souches lors de changements de qualité de lumière. Ces expériences ont été réalisées à

deux intensités lumineuses afin d'établir des corrélations entre le taux de croissance cellulaire et la vitesse d'acclimation de ce processus.

Les travaux présentés dans cet article ont été complétés en étudiant (3) la distribution des phycobiline lyases probablement impliquées dans ce phénomène au sein de différentes lignées de *Synechococcus*, en effectuant un criblage par PCR. La représentation et la distribution de ces derniers a également été étudiée *in situ*, en utilisant les métagénomomes actuellement disponibles sur le portail CAMERA. Enfin, des analyses par HPLC de la composition pigmentaire, réalisées par Christophe Six ont été menées sur des souches acclimatées en GL et BL.

## **II. ARTICLE**

## **A gene island with dual evolutionary origin is involved in chromatic acclimation in marine *Synechococcus***

Florian Humily<sup>1,2</sup>, Frédéric Partensky<sup>1,2</sup>, Christophe Six<sup>1,2</sup>, Gregory Farrant<sup>1,2</sup>, Morgane Ratin<sup>1,2</sup>,  
Dominique Marie<sup>1,2</sup>, David M. Kehoe<sup>3</sup>, and Laurence Garczarek<sup>1,2\*</sup>

<sup>1</sup>UPMC-Université Paris VI, Station Biologique, CS 90074, 29688 Roscoff cedex, France

<sup>2</sup>CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment, Oceanic Plankton group, Marine Phototrophic Prokaryotes team, Place Georges Teissier, CS 90074, 29688 Roscoff cedex, France

<sup>3</sup>Department of Biology, 1001 East Third Street, Indiana University, Bloomington, IN 47405, USA

\*For correspondence. E-mail [laurence.garczarek@sb-roscoff.fr](mailto:laurence.garczarek@sb-roscoff.fr); Tel. (+33) 2 9829 2538; Fax (+33) 2 9829 2324.

Running title : Diversity and evolution of Type IV chromatic acclimation

Submitted to Environmental Microbiology on May 2013

## ABSTRACT

*Synechococcus*, the second most abundant oxygenic phototroph in the marine environment, harbors the largest pigment diversity known within a single genus of cyanobacteria, allowing it to exploit a wide range of light niches. Some strains are capable of Type IV chromatic acclimation (CA4), a process by which cells can match the phycobilin content of their light-harvesting complex (or phycobilisomes) to the ambient blue or green light quality. Here, we performed extensive genomic comparisons to explore the diversity of CA4 within the marine *Synechococcus* radiation. A specific gene island was identified in all CA4 strains, involving two genes (*fcIA/B*) coding for putative transcriptional regulators and one gene coding for a phycobilin lyase. However, two distinct configurations of this cluster were observed, depending on the lineage. In CA4-A islands, the lyase gene coded for MpeZ, a recently characterized phycoerythrobilin lyase-isomerase, and the region also encompassed a third short, possible regulator, called *fcIC*. In CA4-B islands, the lyase gene encoded an uncharacterized relative of MpeZ, called MpeW. While *mpeZ* is more highly expressed in BL than in GL, the reverse was true for *mpeW*, although only minor phenotypic differences were found between chromatic adapters possessing either CA4 island type. Furthermore, some strains possessing CA4 islands proved unable to perform CA4, suggesting that they had accumulated mutations that disrupted this process. This study provides many novel insights into understanding both evolution and regulation of CA4.

## INTRODUCTION

Light is one of the main factors affecting growth and photosynthetic rates of marine phytoplankton. In the field, this resource varies not only in quantity, according to diurnal or seasonal cycles, but also in spectral quality both horizontally, along coast-offshore gradients, and vertically, since the different wavelengths of the solar spectrum are differentially filtered out throughout the water column, with blue light penetrating the deepest (Kirk, 1994; Morel et al., 2007). Competition between photosynthetic organisms for light has triggered the evolution of an amazing variety of pigments and chromophorylated proteins and also different capabilities for sensing and responding to changing light conditions. Cyanobacteria evolved a particularly sophisticated light-harvesting system, called the phycobilisome (PBS; Sidler, 2004). PBS are made of a complex combination of phycobiliproteins (PBP), which covalently bind open-chain tetrapyrrole chromophores, known as phycobilins, to cysteinyl residues via thioether bonds. Chromophorylated  $\alpha$ - and  $\beta$ -subunits of PBP assemble into hexamers that are typically incorporated either into the PBS core or in outward-

radiating rods (for reviews, see e.g. MacColl, 1998; Adir, 2005). Each PBP may bind one to three isomeric phycobilins, whose absorption properties fill the large gap between the two absorption bands of chlorophyll *a* ( $A_{max} \sim 440$  and  $680$  nm; Zhao et al., 2000; Blot et al., 2009). In marine cyanobacteria, these chromophores can be either phycocyanobilin (PCB), phycoerythrobilin (PEB) or phycourobilin (PUB). PCB efficiently captures orange photons ( $A_{max} \sim 620$  nm), while PEB absorbs maximally in the green ( $A_{max} \sim 545$  nm) and PUB in the blue ( $A_{max} \sim 495$  nm). The latter pigment is considered to be the most abundant phycobilin in the ocean, due to the prevalence of blue photons in the euphotic zone (Ong et al., 1984; Olson et al., 1990). PBP from oceanic cyanobacteria, including the coccoids *Synechococcus* and *Crocospaera* and the filamentous *Trichodesmium*, are particularly rich in PUB, and this clearly represents an important adaptive event for life in offshore, blue waters (Ong et al., 1984; Swanson et al., 1991; Six et al., 2007b).

Within the cyanobacterial community, *Synechococcus* is numerically the second most abundant oxygenic phototroph (Partensky et al., 1999; Buitenhuis et al., 2012), contributing significantly to global primary productivity (Li, 1994; Liu et al., 1998). This genus harbors the largest diversity of PBS types of all marine cyanobacteria, explaining its ability to colonize a wide range of light niches, from turbid coastal waters to the most transparent waters of open ocean (Olson et al., 1990; Wood et al., 1998; Haverkamp et al., 2008). This remarkably wide diversity of PBS is due to the numerous possible combinations of PBP with various pigments that constitute the rods (Six et al., 2007b). Most marine *Synechococcus* have PBS rods containing two distinct forms of PE, called PEI and PEII, with PEII always binding both PEB and PUB, whereas PEI binds either only PEB or both phycobilins. These *Synechococcus* have been classified into pigment type 3 by (Six et al., 2007b), who have subdivided them into four pigment subtypes according to the PUB:PEB ratio in whole cells. The latter is generally assessed by the  $Exc_{495:545}$  ratio, i.e. the ratio of the relative fluorescence excitation maxima of these two phycobilins (at 495 and 545 nm, respectively), with emission measured at 575 nm. The  $Exc_{495:545}$  ratio can be either low ( $\sim 0.4-0.5$ ), medium ( $\sim 0.6-0.8$ ) or high ( $\sim 1.7-1.9$ ), with such strains classified as pigment subtypes 3a, 3b or 3c, respectively. The fourth subtype (3d) corresponds to the most sophisticated pigmentation of *Synechococcus*, i.e. cells able to tune their pigmentation according to the ambient spectral light quality, a process termed CA4 (for type IV chromatic acclimation; Palenik, 2001; Everroad et al., 2006; Shukla et al., 2012). Most genes involved in the biosynthesis of PBS rods are present in a large genomic cluster called the PBS rod region, the gene content of which is specific to the pigment type (Six et al., 2007b).

CA4 occurs during shifts from green (GL) to blue light (BL) or reciprocally (Palenik, 2001; Everroad et al., 2006). Unlike other CA types, which occur in both freshwater and marine species (i.e. CA2 and CA3; Tandeau de Marsac, 1977; Kehoe and Gutu, 2006; Gutu and Kehoe, 2011), CA4 does not seem to involve a significant alteration of the PBP composition of PBS, but rather changes in

chromophore content at three sites within PE (Everroad et al., 2006; Shukla et al., 2012). In *Synechococcus* sp. RS9916, of the five binding cysteinyl residues of PEI, only  $\alpha$ -139 is altered by this process, while it affects two of the six binding cysteins of PEII,  $\alpha$ -83 and  $\alpha$ -140. These sites are chromophorylated with PUB in BL and PEB in GL. Recently, the function of MpeZ was characterized in *Synechococcus* sp. RS9916 and identified as one of the key enzymes for the CA4 process (Shukla et al., 2012). MpeZ is bi-functional, since it can both attach a PEB chromophore to  $\alpha$ -83 of PEII and isomerize it to PUB. The *mpeZ* gene encoding this lyase-isomerase was expressed six-fold more in BL than GL and its inactivation strongly suppressed the amplitude of variations of the  $EXC_{495:545}$  ratio after shifts of cells from GL to BL or vice versa (Shukla et al., 2012).

CA4 appears to be widespread within the marine *Synechococcus* radiation, since it has been reported in at least four distinct clades of subcluster 5.1, the most abundant and diversified *Synechococcus* group (Palenik, 2001; Six et al., 2007b; Mazard et al., 2012). However the true extent of this process within this lineage, as well as its genetic and phenotypic variability, have not been explored yet. Here, we performed a comparative genomics study of a wide set of marine *Synechococcus* strains predicted to perform CA4 and identified a genomic island whose presence is strongly correlated with the ability to carry out this process. Interestingly, this island exists in two different configurations and gene contents depending on lineages. A physiological characterization of strains that possess part or all of this island was also performed, which allowed us to highlight some interesting phenotypic differences in the acclimation kinetics and/or amplitudes of the CA4 response among strains, and to refine the classification of marine *Synechococcus* spp. into different pigment types. Finally, a study of the expression of phycobilin lyase genes involved or putatively involved in CA4 allowed us to gain insights about the genetic bases of the observed physiological differences.

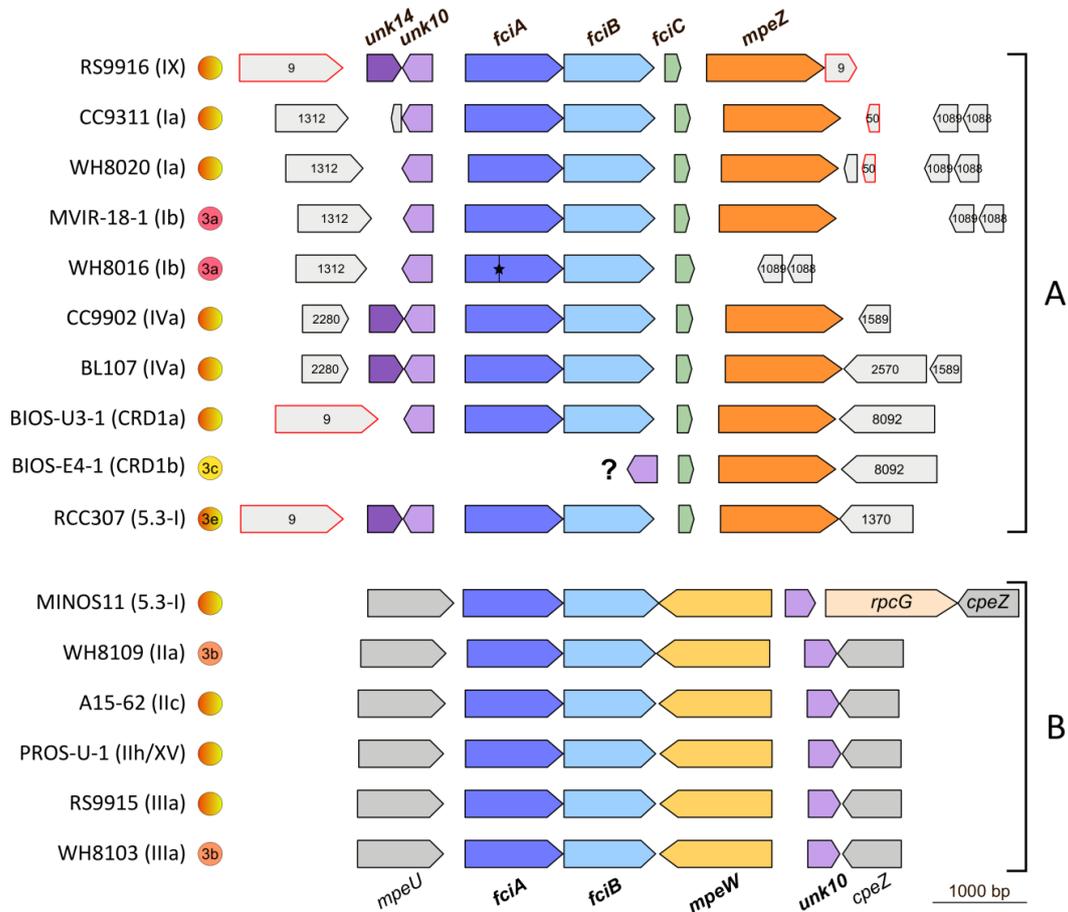
## RESULTS

### *Discovery of a gene island correlated with CA4 in two different configurations*

The availability of genomes of 15 marine *Synechococcus* isolates and preliminary data from 25 additional *Synechococcus* unpublished genomes that will be described in detail elsewhere (G. Farrant and coworkers, unpublished) allowed us to search for genes potentially involved in CA4. Such genes were retrieved in 17 of these 40 sequenced strains, originating from a variety of oceanic environments and depths (Table S1). This set encompasses 15 strains belonging to six distinct clades within subcluster 5.1 (or ten subclades, as defined using the high resolution marker gene *petB*; Mazard et al., 2012) and two strains from a single clade of subcluster 5.3 (sensu Dufresne et al.,

2008). In contrast, there were no strains from subcluster 5.2, since all those included in the initial dataset were lacking phycoerythrin and thus not capable of CA4.

The 17 strains containing genes predominantly found in CA4-capable strains were identified using a two-step bioinformatic approach. First, the 40 genomes were screened for the presence of *mpeZ*, a gene that encodes a phycobilin lyase-isomerase recently shown to have a key role in CA4 (Shukla et al., 2012). Orthologs of this gene were retrieved in 8 genomes, where the *mpeZ* gene was consistently located at the 3'-end of a genomic island (hereafter called "CA4-A island") containing three additional genes, two encoding putative transcriptional regulators and one a short protein of unknown function (Fig. 1). The first two genes encode proteins that both possess a possible C-terminal DNA binding domain (Figs. S1A and S1B). Structural analysis using Phyre2 (Kelley and Sternberg, 2009) predicted that each of these domains consists of seven alpha helices with two weakly interconnected subdomains, each containing one putative  $\alpha$ -helix-turn- $\alpha$ -helix (HTH) motif (IPR018060), similar to *Escherichia coli* AraC, a transcriptional activator involved in the control of the arabinose metabolism (Schleif, 2010). The third gene of the CA4 island is predicted to encode a protein of 55-67 aas and characterized by a N-terminus with a predicted conserved ribbon-helix-helix domain (IPR010985), also found in many bacterial or phage repressors (Fig. S1C; Aravind and Ponting, 1997). Thus, together with the two proteins encoded by the *araC*-like genes, the protein encoded by this gene may be involved in regulating or carrying out CA4. Based on their genomic context and position just upstream of and in the same frame as the CA4-specific phycobilin lyase *mpeZ* (Shukla et al., 2012), we proposed to call these three genes *fciA-C* (where "fci" stands for "type four chromatic acclimation island"; Fig. 1). Another gene of unknown function was consistently found at the 5'-end of the CA4 island, but was encoded on the reverse strand relative to the *fci* genes. This gene is an ortholog of *unk10* (50-98% identity and 31-98% similarity over 106 residues), previously observed in the PBS rod region of the pigment type 3c (i.e. high-PUB) strains WH8102 and CC9605, specifically at the 5'-end of the PEII sub-region (see Fig. 6 in Six et al., 2007b). *Unk10*, predicted to be 107-118 aa in length, contains a Nif11 domain (IPR012903), so-called because it was initially found in the product of an uncharacterized gene of the *nif* cluster of *Azotobacter vinelandii* (Jacobson et al., 1989). In four out of the nine *mpeZ*-containing strains (RS9916, BL107, CC9902 and RCC307), another gene, encoding a predicted 121-123 aas protein of unknown function, was present upstream and head-to-head relative to *unk10* (Fig. 1). This gene, which was designated *unk14* (to extend the numbering of unknown genes in the PBS rod region proposed by Six et al. (2007b)), is specific to these four strains, but the encoded protein is paralogous (50-63% identity and 31-53.4% similarity over 105 residues) to a protein encoded by a gene of unknown function present in all marine *Synechococcus* strains (Cyanorak cluster CK\_00000072; see [www.sb-roscoff.fr/cyanorak/](http://www.sb-roscoff.fr/cyanorak/)). *Unk14* belongs to the highly conserved protein family DUF2237 (or PF09996), which is widely distributed among Bacteria.



**FIG. 1. Gene content and genomic context of the two types of CA4 gene islands found in marine *Synechococcus* genomes.** The CA4-A island can be found at different locations of the genome (though the context is the same within a given clade), whereas the CA4-B island is always located in the main phycobilisome region between *mpeU* and *cpeZ* genes. The gene size and organization are represented to scale and all regions aligned with regard to the *fciA* gene (except for BIOS-E4-1, aligned on *mpeZ* of BIOS-U3-1). The scale bar represents 1000 bp. Genes which are included in CA4 regions are shown in color and orthologs are indicated with the same color. Short CA4 genes of unknown function are named according to Six et al. (2007b), with *unk14* and *fciC* being novel genes, specific of the CA4-A region. Genes found in the immediate genome environment of CA4 regions are shown in grey and designated either by their gene names or their cluster number (with the prefix CK\_ being omitted) in the Cyanorak database of orthologous sequences (<http://www.sb-roscoff.fr/Phyto/cyanorak/>). A stop codon, indicated by a star, interrupted *fciA* of WH8016. The phylogenetic affiliation of strains, as reported in previous studies (Fuller et al., 2003; Ahlgren and Rocap, 2006; Dufresne et al., 2008; Huang et al., 2011; Mazard et al., 2012), is mentioned between brackets. Colored circles represent the pigment type and is specified for strains that do not chromatically adapt. Hotspots for DNA recombination, *hli* (Cluster CK\_00000050; high light inducible protein) and *psbA* (Cluster CK\_00000009) are bordered by a red line.

Surprisingly, two *Synechococcus* strains that are not able to chromatically adapt, BIOS-E4-1 and WH8016 (see physiological characterization below), were each found to possess a partial CA4-A island. Indeed, BIOS-E4-1 lacks the two putative transcriptional regulators *FciA/B*, whereas WH8016 lacks *mpeZ*, suggesting that these strains could represent natural deletion mutants of distinct genes of the CA4 island (see discussion).

As a second step, we searched for the putative transcriptional regulator genes *fciA* and *fciB* in other *Synechococcus* genomes. This led to the discovery of another type of CA4 island in six additional genomes (Fig. 1). In contrast to the first one, this so-called "CA4-B island" lacks both *fciC* and *mpeZ*. The latter gene is replaced by a gene located on the opposite strand of the *fciAB* operon, encoding a putative novel member of the phycobilin lyase E/F family (Schluchter et al., 2010; Bretaudeau et al., 2013). Like *MpeZ*, it is a paralog of *MpeY* (Wilbanks and Glazer, 1993), whose gene is found in all PEII-containing *Synechococcus* strains and located in the PEII subregion of the PBS rod region (Fig. S2; see also Fig. 6 in Six et al., 2007b). Since this gene is not a direct ortholog of *mpeZ* (see phylogenetic analysis below) and co-occurs with *mpeY*, we propose to call it *mpeW*.

Other noticeable differences between CA4-A and CA4-B islands are the different locations of *unk10* with regard to the *fciAB* operon (upstream or downstream, respectively) and their distinct genomic context (Fig. 1). Indeed, whatever the clade, the genomic context of the CA4-B island was always the same, i.e. in the middle of the PBS rod gene region and more specifically between *mpeU* and *cpeZ*, at the 3'-end of the PEII sub-region (Fig. S2). In contrast, even though the location of the CA4-A island was globally conserved among strains of a given clade, it differed between clades and in all strains examined so far was never located in the PBS region (see Figs. 1 and S2). Another important point is that within subcluster 5.1, all strains of a given clade possessed one type of CA4 island but not the other. Indeed, strains belonging to clades I, IV, IX and CRD1 possess a CA4-A island, while a CA4-B island is found in clades II and III. Although the genome of the chromatically adapting *Synechococcus* sp. strain M11.1 is not available yet, we used PCR amplification to determine that it contains *mpeW* and not *mpeZ* (data not shown), as expected for a clade II strain. Interestingly, this clade differentiation does not hold true for subcluster 5.3, since the closely related strains RCC307 and MINOS11 possess a CA4-A and a CA4-B island, respectively. Another unexpected observation was that MINOS11 possesses, downstream of *unk10*, an ortholog of the *rpcG* gene, which is normally located on the 3'-end of the PBS region (Six et al., 2007b). *RpcG* was shown to encode a lyase-isomerase that attaches a PEB molecule at Cys-84 of the  $\alpha$ -PC subunit and isomerizes it to PUB (Blot et al., 2009).

#### *Phylogenetic analyses of genes putatively involved in CA4*

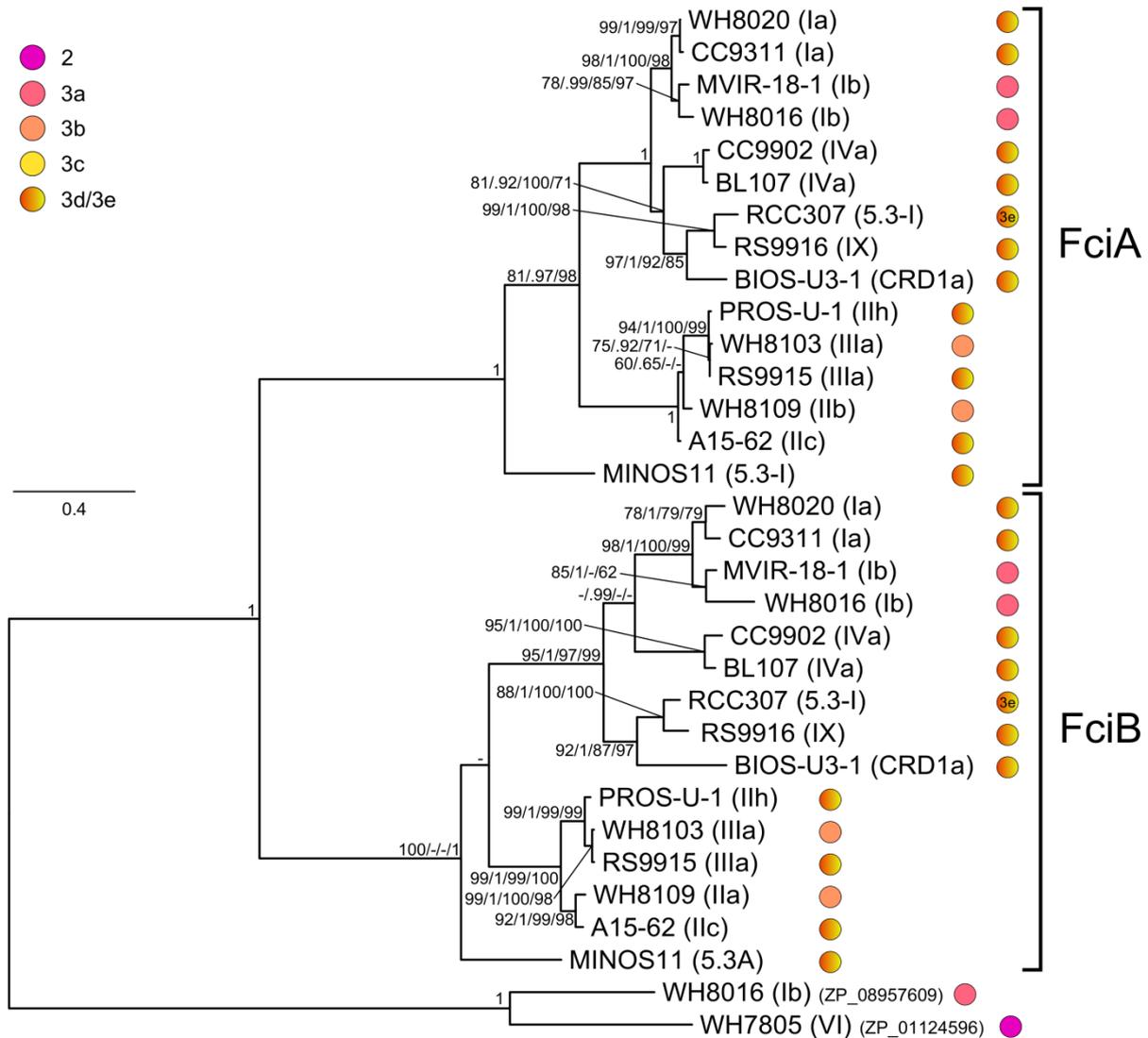
In order to better understand the evolution of the CA4 process and notably the origin of the differentiation into two distinct CA4 islands, we performed a number of phylogenetic analyses using genes that are common to the two island types CA4-A and CA4-B (*fciA*, *fciB*, *unk10* and *mpeW* or *mpeZ*). Figure 2 shows a phylogenetic tree based on *FciA* and *FciB* sequences and rooted with two more distant, uncharacterized members of the AraC family retrieved from marine *Synechococcus*

spp. (Syn8016DRAFT\_2956 and WH7805\_05326; Cyanorak cluster CK\_00002282). As for other cyanobacterial members of the AraC family, their homology with FciA and FciB mostly resides at the C-terminal HTH domain level (Figs. S1A and S1B). Orthologs of both Fci proteins separated into two main branches, the first one including all strains that have a CA4-A island (except BIOS-E4-1, which lacks *fciAB* genes) and the second one encompassing five out of the six strains having a CA4-B island (Figs. 1 and 2). The sixth strain is the subcluster 5.3 strain MINOS11 whose FciA and FciB are split apart from other sequences, consistently with the wide phylogenetic distance between members of subclusters 5.1 and 5.3 (Dufresne et al., 2008). By contrast, the fact that both Fci sequences from the other subcluster 5.3 strain, RCC307, fall within subcluster 5.1 close to clade IX strain RS9916 strongly suggests that RCC307 might have acquired these genes by lateral transfer from a cell belonging to clade IX or a related lineage.

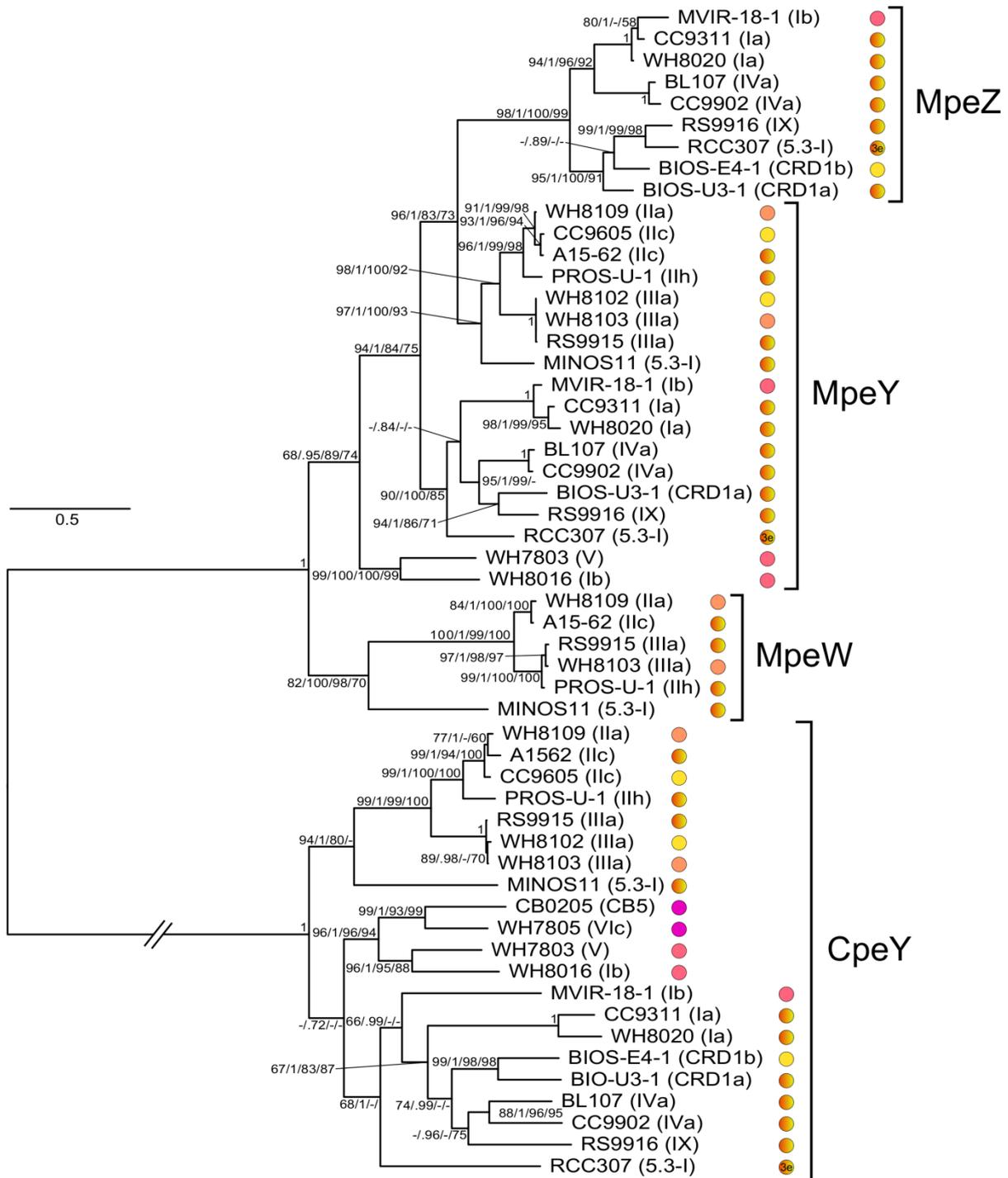
A globally similar tree topology was retrieved for Unk10 (Fig. S3), except that i) BIOS-E4-1 was present and clustered with BIOS-U3-1, another member of the CRD1 clade (Ahlgren and Rocap, 2006; Mazard et al., 2012) and ii) MINOS11 was clearly located at the base of the branch encompassing all strains that also possess *mpeW* (i.e. CA4-B strains). Interestingly, the latter branch also included the pigment type 3c (i.e. high-PUB) strains WH8102 and CC9605, in which the *unk10* gene is located at the 3'-end of the PEII gene subregion (Six et al., 2007b), as in most CA4-B strains (Fig. 1).

We also generated an amino acid alignment comprising the sequences of MpeZ, MpeW and two other related members of the phycobilin lyases E/F clan (Bretaudeau et al., 2013), CpeY and MpeY, occurring in all *Synechococcus* strains possessing PE and PEII, respectively. While CpeY was biochemically characterized to catalyze the attachment of a PEB molecule at Cys-82 of the alpha-PEI subunit (Biswas et al., 2011), MpeY is still uncharacterized (Six et al., 2007b; Bretaudeau et al., 2013). Comparative motif analyses using Protomata (Fig. S4) shows that MpeW, MpeY and MpeZ share several conserved E-Z HEAT-repeat domains, known to facilitate protein-protein interactions (Andrade et al., 2001; Morimoto et al., 2003), but also a few specific motifs, such as the strongly conserved 'HRDE', located around position 285 and not found in CpeY. In the phylogenetic tree shown in Figure 3, sequences of these three proteins grouped together, well apart from the CpeY branch. As for the Unk10 tree, MpeY sequences grouped into two main clusters, a first one containing all strains that possess *mpeZ* (i.e. a CA4-A island) and a second one with *mpeW*-containing and pigment type 3c strains. A third MpeY branch was also present and contained the pigment type 3a (i.e. low-PUB) strains WH7803 and WH8016. The occurrence of the latter strain at this position in the tree may seem surprising given that it possesses a partial (*mpeZ*-lacking) CA4-A island (Fig. 1), but it is worth noting that its *mpeY* gene is located in a PBS region whose gene content and organization is similar to that of WH7803 (Fig. S2; see also Fig. 6 in Six et al., 2007b).

A similar tree topology as for MpeY was found for CpeY (Fig. 3), except that the third branch contained two pigment type 2 (i.e. PUB and PEII-less) strains, CB0205 and WH7805, which lack the *mpeY* gene.



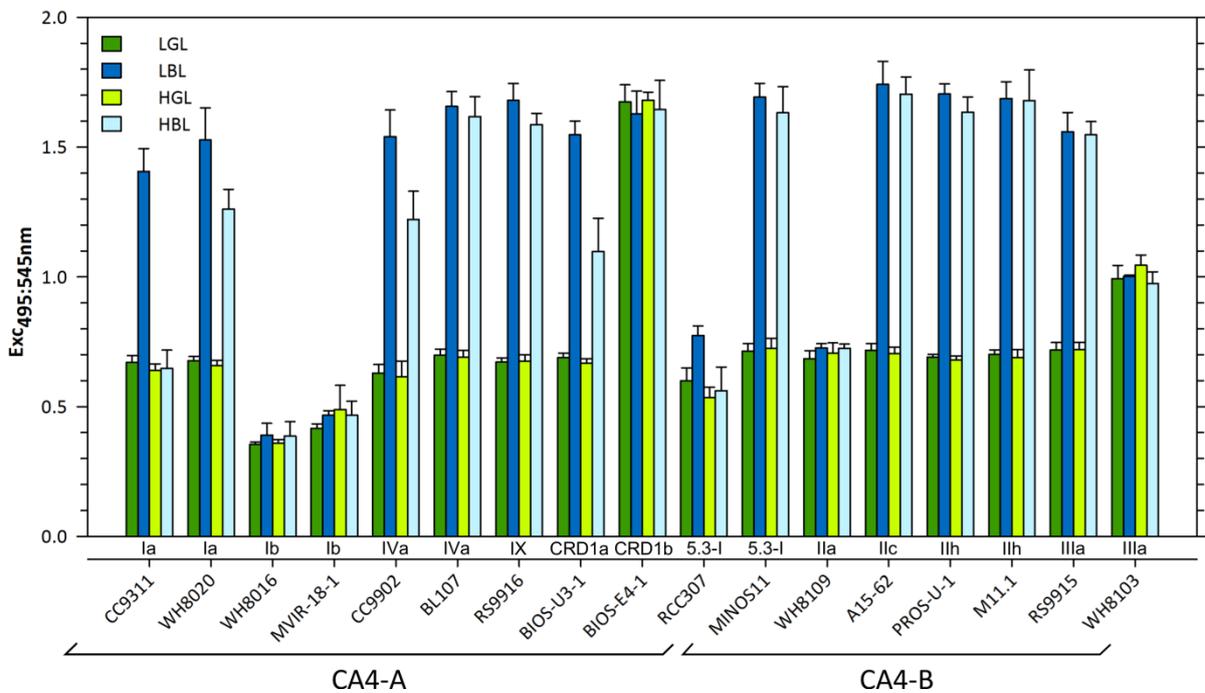
**FIG. 2. Maximum likelihood analysis of AraC-like proteins FciA and FciB (289 aligned positions) from marine *Synechococcus*.** For each strain, the phylogenetic affiliation, is mentioned between brackets and the pigment type is indicated by colored circles. Series of four numbers shown at nodes correspond to bootstrap values for ML analyses, Bayesian posterior probabilities (PP, ranging between 0 and 1), and bootstraps values for Neighbor-Joining and Parsimony methods, respectively. Bootstraps, represented as a percentage, were obtained through 1,000 repetitions and PP from 1,500,000 generations. Only values higher than 0.60 for PP and 60% for bootstrap values are shown on the phylogenetic tree. The scale bar represents 0.4 substitutions per nucleotide.



**FIG. 3. Maximum likelihood analysis of five phycobilin lyase sequences of the E/F clan (380 aligned positions) from marine *Synechococcus*.** For each *Synechococcus* strain, the phylogenetic affiliation is mentioned between brackets and the pigment type is indicated by colored circles. Series of four numbers shown at nodes correspond to bootstrap values for ML analyses, Bayesian posterior probabilities (PP, ranging between 0 and 1), and bootstraps values for Neighbor-Joining and Parsimony methods, respectively. Bootstraps, represented as a percentage, were obtained through 1,000 repetitions and PP from 1,500,000 generations. Only values higher than 0.60 for PP and 60% for bootstrap values are shown on the phylogenetic tree. The scale bar represents 0.6 substitutions per nucleotide. The branch bearing the CpeY cluster has been cut for readability (its full length is 3.85 substitutions per nucleotide).

*Physiological characterization of mpeW- or mpeZ-containing strains acclimated to blue or green light*

In order to evaluate the ability of the different strains listed in Table S1 to change their pigmentation in response to ambient light, we determined the  $Exc_{495:545}$  ratio, a proxy for the PUB to PEB ratio, from exponentially growing cultures acclimated to continuous blue light (BL) or green light (GL) for over 20 generations (Fig. 4). Measurements were done both on cultures grown at low light (LL;  $20 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ ) and high light (HL;  $75 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ ). The  $Exc_{495:545}$  ratio was generally stable in GL and more variable in BL (Fig. 4). This observation held particularly true at high irradiance, at which the  $Exc_{495:545}$  ratio of BL controls often declined notably after cultures had reached the early stationary phase.



**FIG. 4.** Average  $Exc_{495:545}$  ratios of CA4 genes-containing *Synechococcus* strains acclimated to blue light (BL) and green light (GL) at two different irradiances ( $20$  and  $75 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ ). Low and high light conditions are shown by dark and light colors, respectively. A minimum of 5 replicates were used to calculate the mean and standard deviation.

Despite the presence of genes potentially involved in CA4 in their genomes (see above), five strains (WH8103, BIOS-E4-1, MVIR-18-1, WH8016 and WH8109) were unable to chromatically acclimate, since they displayed a similar  $Exc_{495:545}$  ratio in GL and BL at both irradiances tested (Table 1). An additional strain seemingly had a strongly altered capacity for CA4, since it exhibited only a small (29% for RCC307) increase of this ratio between LGL and LBL (significant at  $P < 0.01$ ; Fig. 4). In contrast, for the latter strain, the  $Exc_{495:545}$  ratio was not significantly different between HGL and HBL ( $P > 0.1$ ). However, the mean value of the  $Exc_{495:545}$  ratio (both irradiances considered) was highly

variable among these strains since it was i) as low as measured for the pigment type 3a strain WH7803 (Six et al., 2007b) in WH8016 and MVIR-18-1 ( $0.37\pm 0.04$ ;  $n=30$  and  $0.46\pm 0.06$ ;  $n=27$ , respectively), ii) at a level intermediate between pigment type 3a and GL-grown CA4 strains for RCC307 ( $0.62\pm 0.11$ ,  $n=115$ ), iii) similar to that of GL-grown CA4 strains for WH8109 ( $0.71\pm 0.03$ ,  $n=109$ ), iv) intermediate between values measured in GL- and BL-grown CA4 strains for WH8103 ( $1.00\pm 0.05$ ,  $n=27$ ) and v) comparable to that of BL-grown CA4 strains in BIOS-E4-1 ( $1.66\pm 0.08$ ,  $n=27$ ).

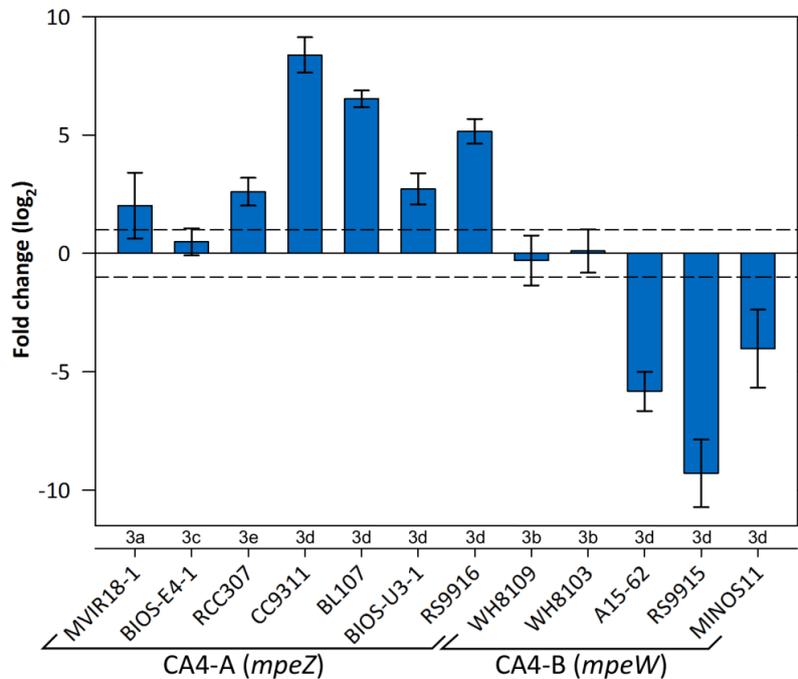
**TABLE 1.** Summary of the genotypic and phenotypic pigment characteristics of the *Synechococcus* strains used in this study

Strains	Subclade <sup>a</sup>	Pigment type (Six et al., 2007)	Pigment phenotype (this study)	CA4 island type	CA4 phenotypic group	Exc <sub>495:545</sub> amplitude variations	Pigment type (this study)
WH8016	Ib	3a	3a	A	na	None (ratio ~ 0.4)	3aA
MVIR-18-1	Ib	3a	3a	A	na	None (ratio ~ 0.5)	3aA
WH8103	IIIa	3c	3b	B	na	None (ratio ~ 1.0)	3bB
WH8109	IIa	3b	3b	B	na	None (ratio ~ 0.7)	3bB
BIOS-E4-1	CRD1b	3c	3c	B	na	None (ratio ~ 1.7)	3cB
RS9916	IX	3d	3d	A	2	Small variation in HBL	3dA
CC9311	Ia	3d	3d	A	2	Small variation in HBL	3dA
WH8020	Ia	3d	3d	A	2	Small variation in HBL	3dA
BIOS-U3-1	CRD1a	3d	3d	A	2	Small variation in HBL	3dA
CC9902	IVa	3d	3d	A	3	Large variation but delayed in BL to GL	3dA
BL107	IVa	3d	3d	A	3	Large variation but delayed in BL to GL	3dA
MINOS11	5.3-I	3d	3d	B	1	Large variation	3dB
A15-62	IIc	3d	3d	B	1	Large variation	3dB
PROS-U-1	IIh	3d	3d	B	1	Large variation	3dB
RS9915	IIIa	3d	3d	B	1	Large variation	3dB
M11.1	IIh	3d	3d	B?	1	Large variation	3dB?
RCC307	5.3-I	3b	3e	A	4	Small variation	3eA

All eleven other strains analyzed in this study were able to match their pigmentation to the ambient light color with an average  $134\pm 13\%$  increase of their Exc<sub>495:545</sub> ratio between LGL and LBL, consistently with previous reports on CA4 (Palenik, 2001; Everroad et al., 2006; Shukla et al., 2012). A similar variation was also observed at high irradiance ( $119\pm 43\%$ ), except for CC9311, which maintained a low Exc<sub>495:545</sub> ratio at HBL (Fig. 4). It is noteworthy that the average Exc<sub>495:545</sub> ratios were more variable and generally lower ( $P<0.01$ ) for CA4-A strains than for their CA4-B counterparts in all four light conditions tested, though the difference was much more conspicuous at HL than LL ( $0.67\pm 0.02$  vs.  $0.71\pm 0.01$  at LGL;  $1.58\pm 0.10$  vs.  $1.67\pm 0.08$  at LBL;  $0.66\pm 0.03$  vs.  $0.71\pm 0.02$  at HGL and  $1.41\pm 0.25$  (excluding CC9311) vs.  $1.63\pm 0.06$  at HBL, respectively).

*CA4 induces a change in the expression of phycobilin lyase genes*

The inability of *Synechococcus* spp. WH8016 and BIOS-E4-1 to acclimate to changes in light quality may be due to the absence of key CA4 genes in these genomes (i.e., *mpeZ* and *fciAB*, respectively; Fig. 1). More unexpectedly, strains WH8109, WH8103, MVIR-18-1 and RCC307, which possess a complete set of CA4 genes (Fig. 1), are also affected in their ability to perform CA4 (see above). Comparative analyses of protein sequences did not reveal any obvious mutation/deletion or insertion in the most conserved regions of FciA/B (data not shown) or MpeW/Z (Fig. S3) that could explain this loss of CA4 phenotype. To check whether this could result from an alteration of the control of the expression of the known and putative CA4-specific lyase-encoding genes, *mpeZ* or *mpeW* transcript levels were measured for exponentially growing strains acclimated to either LGL or LBL in a selection of strains.



**FIG. 5.** Differential gene expression of *mpeW* or *mpeZ* of *Synechococcus* strains acclimated under GL of BL. Transcripts levels were quantified using RT-qPCR. The mean and standard deviation are calculated from triplicates and normalized to levels of transcripts under GL. Dotted line denote the significance threshold (>1, upregulated; <1, downregulated).

All four tested CA4-A strains that exhibited a large variation of their  $Exc_{495:545}$  ratio between LGL and LBL (Fig. 4) showed a significantly higher *mpeZ* expression in LBL than in LGL, ranging from 6.7-fold [ $\log_2(FC)=2.73\pm0.66$ ] in BIOS-U3-1 up to 333-fold [ $\log_2(FC)=8.38\pm0.75$ ] in CC9311 (Fig. 5). RCC307, which showed only a small variation of its  $Exc_{495:545}$  ratio between LGL and LBL (Fig. 4), also displayed a slightly higher expression of *mpeZ* in LBL than LGL, comparable to that of BIOS-U3-1 (Fig. 5). In contrast, all three tested CA4-B strains that were capable of CA4 exhibited a significantly lower

*mpeW* expression in LGL than in LBL, ranging from 35.9-fold [ $\log_2(\text{FC})=4.03\pm 1.65$ ;  $n=3$ ] in MINOS11 up to 1,150-fold [ $\log_2(\text{FC})=9.29\pm 1.43$ ;  $n=3$ ] in RS9915 (Fig. 5). All but one strains that exhibited no significant change in their  $\text{Exc}_{495:545}$  ratio between BL and GL (Fig. 4) also showed no significantly different expression of either *mpeZ* or *mpeW* gene between BL and GL in gene expression (i.e. lower than 2-fold between the two light colors). The only exception was MVIR-18-1, for which in one out of the three biological replicates *mpeW* was significantly more expressed in BL than GL.

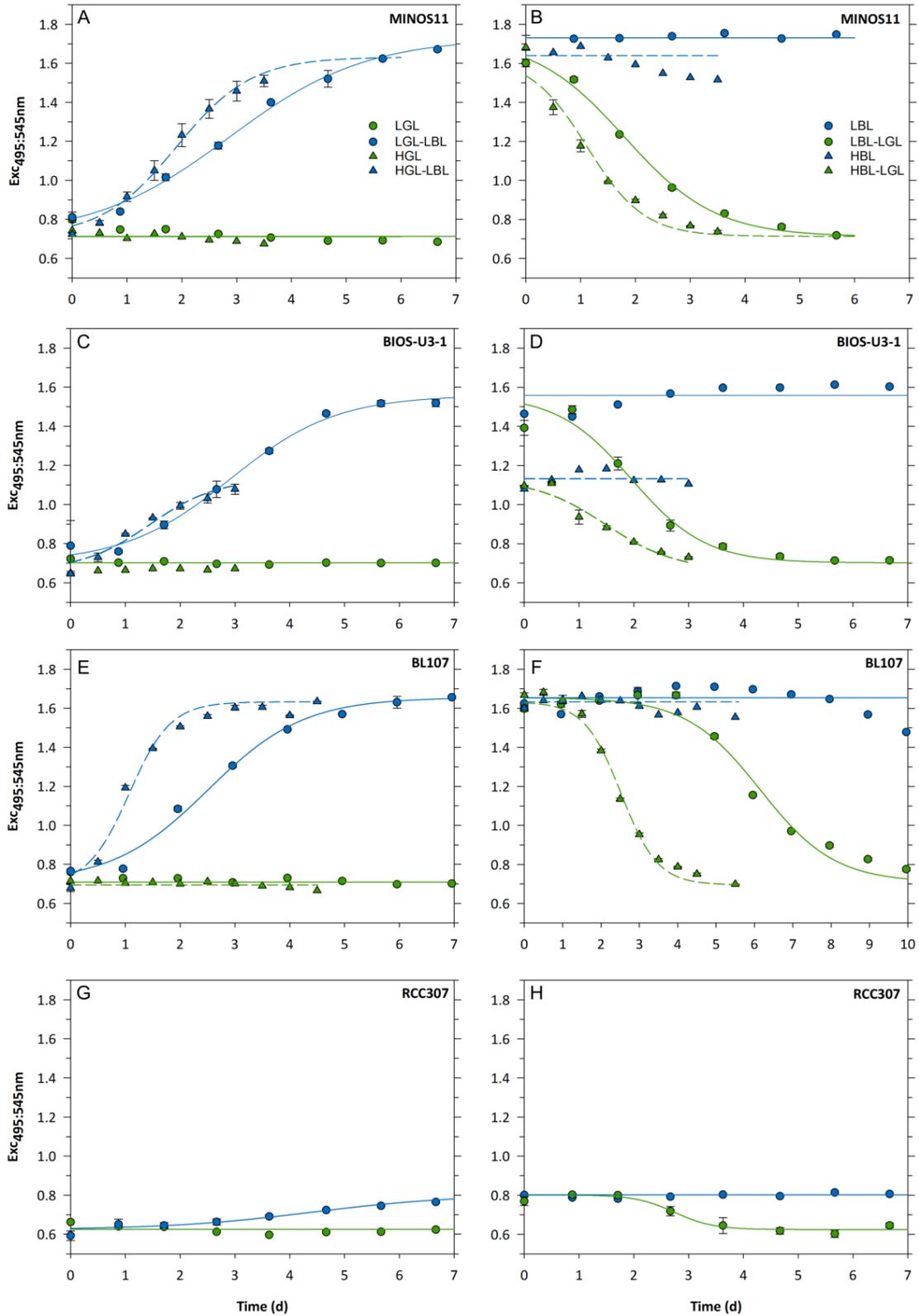
#### *Variability of growth and acclimation kinetics of CA4 strains during light quality shifts*

The kinetics and/or amplitudes of chromatic acclimation of twelve strains, representative of six distinct clades of the subcluster 5.1 and one clade of the subcluster 5.3, were analyzed in order to reveal possible phenotypic differences. All strains were allowed to acclimate for at least one month in either BL or GL and at two irradiances, LL and HL, prior to being shifted to the other light color condition (same irradiance). Growth rates ( $\mu$ ) and the variation of the  $\text{Exc}_{495:545}$  ratio were then assessed from flow cytometric cell counts and by spectrofluorimetry, respectively. The latter measurements allowed us to assess the half acclimation time ( $T_{50}$  in days) for each strain and each type of light quality shift at both irradiances tested (see Table S2 and Figs. 6 and 7).

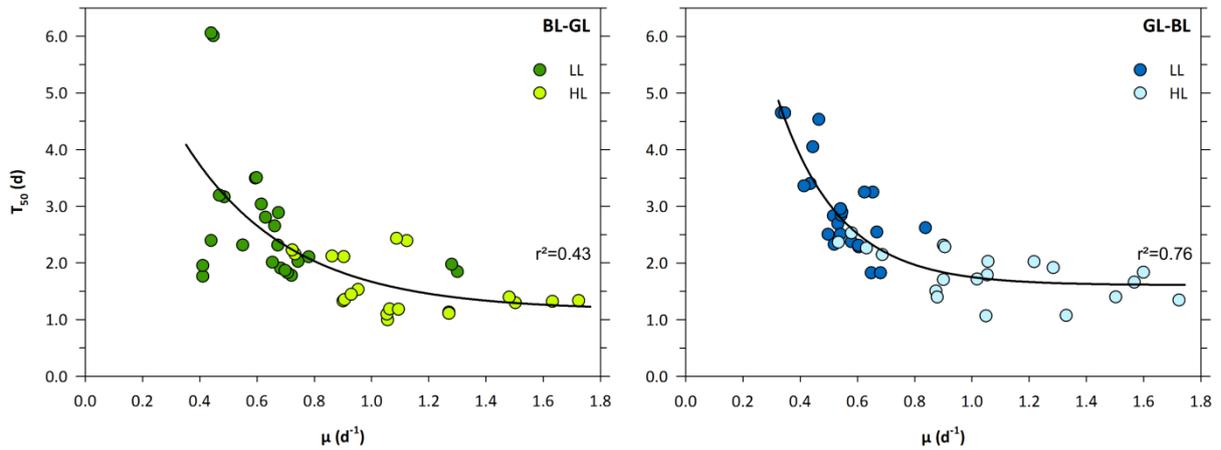
All strains displayed a significantly higher growth rate and a shorter  $T_{50}$  when grown in HL than in LL (Fig. 7 and Table S2). Based on kinetics studies, four CA4 phenotypic groups can be defined among strains that showed a variable  $\text{Exc}_{495:545}$  ratio between GL and BL and Figure 6 illustrates one representative example (or the sole case) of each group. Group 1 that comprises strains A15-62, PROS-U-1, M11.1, RS9915, and MINOS11 exhibited a 'typical' CA4 kinetics at both growth irradiances (Fig. 6A-B and Table 1). This group exhibited i) significant changes in the  $\text{Exc}_{495:545}$  ratio within 1-2 days after the light color shift, ii)  $T_{50}$  values of  $2.01\pm 0.24$  d for the LBL to LGL shift,  $2.72\pm 0.37$  d for the reverse shift and  $1.22\pm 0.13$  d and  $1.85\pm 0.36$  d for the corresponding shifts at HL and iii) the complete acclimation to the new color took about 5.8 d at LL and this time was reduced to about 3.4 to 4 d at HL (Fig. 6).

The second group, comprising the CA4-A strains BIOS-U3-1, CC9311, WH8020 and RS9916 was defined by the fact that cultures reached a significantly lower  $\text{Exc}_{495:545}$  at HBL than LBL, though their  $T_{50}$  was comparable to the previous group ( $P>0.01$ , Fig. 6C-D and Table S2).

The third group comprises the two closely related strains BL107 and CC9902, which exhibited a systematic lag compared to the first two groups in the initiation of the CA4 process after being shifted from BL to GL. Indeed, the  $\text{Exc}_{495:545}$  ratio decreased only after 3-4 days in LL and 1.5-2 days in HL (Fig. 6E).



**FIG. 6. Phenotypic variability of chromatic acclimation in marine *Synechococcus*.** Temporal changes of the EXC<sub>495:545</sub> after shifts from BL to GL (and vice versa) at two irradiances: LL (circles) and HL (triangles). The color (blue or green) of lines and symbols matches the ambient light color under which cultures were shifted at time zero (or left, for controls). Error bars indicate standard deviation for two biological replicates. Four distinct CA4 phenotypic groups were observed: group 1 (A,B) for strains MINOS11, A15-62, PROS-U-1, M11.1 and RS9915; group 2 (C,D) for BIOS-U3-1, CC9311, WH8020, CC9902 and RS9916; group 3 (E,F) for BL107 and CC9902; group 4 (G,H) for the sole strain RCC307. Note the different x-axis scale for BL107.



**FIG. 7. Correlations between growth rate ( $\mu$ , days) and half maximal acclimation time ( $T_{50}$ , days) of twelve *Synechococcus* strains.** Shifts to another light quality (BL to GL and vice versa) were performed under two light intensities. LL and HL corresponds respectively to 20 and 75  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  of BL or GL. Data represents two biological replicates for each strain and each light condition.

To better understand this behavior, we tested whether the delay in the CA4 initiation observed during the BL to GL shift was depending on the acclimation time to BL before the shift to GL. We found that the delay was the same for BL107 cultures acclimated to BL for weeks as for cultures shifted to GL either immediately, or after 1 or 3 days after they reached the BL value of the  $\text{Exc}_{495:545}$  ratio following a GL to BL shift (Fig. S5A and Table 1). Furthermore, BL107 cultures that were subjected to successive shifts, i.e. first from BL to GL, then back to BL just before or after being fully GL-acclimated, then again shifted from BL to GL immediately upon reaching the BL value of the  $\text{Exc}_{495:545}$  ratio, also exhibited the same delay (Fig. S5B). Altogether, these experiments clearly demonstrate that this delay is independent from the light history of the culture. In contrast, no significant lag occurred after a shift of BL107 and CC9902 cultures from GL to BL (Fig. 6F).

The fourth CA4 group comprises the sole strain RCC307, for which the amplitude of variation of the  $\text{Exc}_{495:545}$  ratio was very limited (0.63-0.80 for a shift from GL to BL; Fig. 6G-H and Table 1). It is interesting to note that the  $\text{Exc}_{495:545}$  ratio in GL of this strain was slightly lower than that of other strains analyzed in this study.

## DISCUSSION

### *Type 4 chromatic acclimation in marine Synechococcus requires a genomic island*

CA4 is a sophisticated light-controlled process by which *Synechococcus* spp. cyanobacteria can efficiently modify the phycobilin content of their PBS in response to fluctuations in light quality occurring in the marine environment along nutrient and/or light gradients (Palenik, 2001; Everroad et al., 2006; Shukla et al., 2012). In the present study, we have identified a small gene cluster comprising 4 to 6 genes, which we have found to be correlated with the capacity of cells to carry out CA4. Unexpectedly, depending on *Synechococcus* strains, this cluster could exhibit two distinct configurations (so-called CA4-A and CA4-B) in terms of gene content, gene order and genomic context (Fig. 1 and Table 1). Physiologically, chromatically acclimating strains containing either island type displayed a similar kinetics of variation of the  $Exc_{495:545}$  ratio in response to color shifts, indicating that CA4-A and CA4-B regions are functionally equivalent. However, while all CA4-B strains displayed a similar amplitude of variation of their  $Exc_{495:545}$  ratio at LL as at HL (CA4 phenotypic group 1; Fig. 6), CA4-A strains exhibited more variability with three possible behaviors: a lower  $Exc_{495:545}$  ratio at HL than LL (group 2), a delay in the initiation of the CA4 process after a shift from BL to GL (group 3) or a reduced amplitude of variation of the ratio at LL and no variation at HL (group 4). The difference between groups 1 and 2 at HL could be due to the loss of one or two distal PEII discs from the PBS rods of CA4-A strains, as previously observed in other *Synechococcus* strains either grown at high white irradiance (Six et al., 2004) or submitted to a ultraviolet stress (Six et al., 2007a). This hypothesis is consistent with the longer acclimation time needed to reach the maximal  $Exc_{495:545}$  ratio in HBL for CA4-A than for CA4-B strains (not shown).

CA4-A and CA4-B islands have in common to contain an operon of two putative transcriptional regulator genes (*fciAB*) and one gene encoding a characterized (MpeZ; Shukla et al., 2012) or a putative (MpeW) phycobilin lyase of the E/F clan (Bretaudeau et al., 2013), that is either on the same or the reverse strand as *fciAB* (Fig. 1). MpeW is not an ortholog of MpeZ since the best hit of both sequences is MpeY, a paralog of both proteins (Fig. 3). However, the combination of genomic and physiological data suggests that these mutually exclusive enzymes could have the same function, i.e. to catalyze the attachment of a PEB chromophore at Cys-83 of the PEII alpha-chain (MpeA) and to isomerize it into PUB (Shukla et al., 2012). This hypothesis is seemingly contradicted by expression studies, which indicate that *mpeZ* is significantly more expressed in BL than in GL (see (Shukla et al., 2012 and this study), while *mpeW* exhibits the reverse behaviour (Fig. 5). An alternative hypothesis would therefore be that *mpeW* encodes a GL-induced PEB lyase that catalyzes

the binding of a PEB either at Cys-83 of MpeA (although the CpeY/Z complex is another candidate for this function; Biswas et al., 2011; Shukla et al., 2012).

Transcriptional regulators of the AraC family are found in a wide range of microorganisms and regulate a variety of cellular pathways (Gallegos et al., 1997). They are characterized by the presence of HTH domains, which bind DNA target sequences and activate or repress transcription from cognate promoters. Most are single genes and the signal receptor resides in the same polypeptide as the regulatory function (e.g. in the case of AraC itself; Schleif, 2010). Occurrence of a cluster of two *araC*-like genes, as is the case for *fciAB*, is more rarely observed, e.g. in rhamnose catabolism which was shown to be regulated in a two-step activation cascade by the *rhaSR* operon (Egan and Schleif, 1993; Wickstrum et al., 2010). Interestingly, *Synechococcus* sp. BIOS-E4-1 was found to possess a partial CA4-A island lacking the *fciAB* genes (Fig. 1), providing us some additional insights into the putative regulatory role of these genes. Indeed, this strain is seemingly locked into the BL phenotype ( $\text{Exc}_{495:545} \sim 1.65$ ), independently of the ambient light color (Fig. 4) and its *mpeZ* is not differentially expressed between GL and BL (Fig. 5). This suggests that in strains in which *fci* genes are functional, the regulatory complex they encode might repress the expression of the *mpeZ* gene in GL.

In *E. coli* cells grown in the absence of arabinose, the AraC homodimer represses transcription of the *araBAD* operon, which encodes three enzymes involved in the L-arabinose catabolism, by binding to two DNA half-sites and forming a DNA loop (for a review, see Schleif, 2010). When arabinose is present, it attaches itself to the AraC N-terminal domain, modifying the structural conformation of the AraC homodimer and causing it to bind to different DNA sites; this modification releases the DNA loop, thus allowing transcription from the  $P_{\text{BAD}}$  promoter (Rodgers and Schleif, 2009). By analogy, it is possible that homo- or heterodimers of FciA and/or FciB may repress the transcription from the  $P_{\text{mpeZ}}$  promoter in GL. However, the fact that *mpeW* exhibited an opposite expression pattern with regard to *mpeZ* suggests that the regulatory machinery controlling the CA4 process is in fact more complex. In this context, the short *fciC* gene, which is found in CA4-A but not in CA4-B islands and that codes for a ribbon-helix-helix domain-containing protein previously found in bacterial and phage repressors (Aravind and Ponting, 1997), might well play a role in the differential expression of *mpeW* and *mpeZ*.

Another interesting finding from this study is that several *Synechococcus* strains that possess a complete CA4-A or B island were only partially (RCC307) capable or completely unable (WH8109, MVIR-18-1, WH8103) to chromatically acclimate (Figs. 1 and 4; Table 1), suggesting that these strains could be natural mutants for this process. The case of RCC307 is particularly interesting in this context, since this strain exhibits small but reversible changes of the  $\text{Exc}_{495:545}$  ratio after light color shifts, and the amplitude of variation is similar to that previously reported for a RS9916 mutant in

which *mpeZ* has been inactivated (Shukla et al., 2012). However, transcripts accumulation data showed that this gene is significantly expressed in RCC307. Although it was only about 6-fold more expressed in BL than GL conditions, this was quite comparable to the difference observed in BIOS-U3-1, a typical CA4-A strain (Fig. 5). Thus, it is possible that the expression of other, still unidentified, phycobilin lyase gene(s) involved in the chromophorylation of the two other cysteinyl binding sites that change during CA4 (i.e., Cys-139  $\alpha$ -PEI and Cys-140  $\alpha$ -PEII; Shukla et al., 2012) could also be affected in RCC307.

Several strains exhibited no significant changes in the  $Exc_{495:545}$  ratio between BL and GL, indicating that their CA4 process is affected at either structural or regulatory levels. WH8016 constitutes an obvious example of the first type of alteration, since this strain lacks *mpeZ* and has a stop codon in *fciA* (Fig. 1), two characteristics that we predict should completely suppress its ability to perform CA4. Another noteworthy example is MVIR-18-1, which despite having a complete CA4-A island, displayed a very low and invariable  $Exc_{495:545}$  ratio (Fig. 5), comparable to that of the pigment type 3a strain WH7803, which has been shown to have a molar PUB:PEB ratio of 0:5 for PEI and 1:5 for PEII (Ong and Glazer, 1991; Six et al., 2007b). Interestingly, the PE gene region of MVIR-18-1 resembles that of other CA4-A strains except that it lacks the *mpeU* gene (Fig. S2), which codes for an uncharacterized phycobilin lyase of the E/F clan (Bretaudeau et al., 2013). This suggests that MpeU could have a critical role for the PUB chromophorylation of phycoerythrins.

Lastly, WH8103 and WH8109 are two strains that possess both a complete PE region (Fig. S2) and a complete CA4-B island, but are unable to perform CA4. Both of these showed no differential expression of *mpeW* between GL and BL. Given their different phenotypes ( $Exc_{495:545}$  ratios of  $\sim 1.00 \pm 0.05$  and  $0.71 \pm 0.03$ , respectively; Fig. 1), it is possible that these strains are affected at different levels of the CA4 regulatory network.

#### *CA4-A and CA4-B island types have different evolutionary histories*

Unlike motility, a physiological trait found only in clade III strains (Toledo et al., 1999), CA4 is not restricted to one specific clade, but appears to be widely distributed within the marine *Synechococcus* radiation. Indeed, CA4 strains are found in both subclusters 5.1 and 5.3, while none has been isolated yet from the strictly coastal subcluster 5.2. As for structural genes of the PBS region (Six et al., 2007b), phylogenetic analyses using genes of the CA4 cluster are globally inconsistent with those made with classical marker genes, including the 16S rRNA, the 16S-23S rRNA internal transcribed sequence (ITS), the cytochrome *b6* subunit (PetB) or diverse combinations of concatenated core proteins (Dufresne et al., 2008; Zwirgmaier et al., 2008; Mazard et al., 2012). This suggests that genes of both CA4 cluster types form a genomic island that has been laterally

transferred between lineages at some point during the evolution of marine *Synechococcus*. Furthermore, different clades of subcluster 5.1 were found to possess distinct CA4 island types, with clades I, IV, IX and CRD1 having a CA4-A island, whereas clades II and III contain a CA4-B island. This suggests that this separation is relatively ancient, as confirmed by phylogenetic analyses with coding sequences common to both region types (*FciA/B* or *Unk10*), which consistently split apart these two groups of strains (Figs. 2 and S2).

In order to gain more insights about how the different *Synechococcus* lineages have acquired the CA4 process by obtaining one or the other island type, we examined their genomic context. The latter is remarkably conserved around CA4-B islands, which are always situated in the middle of the PBS region and more specifically at the 3'-end of the PEII subregion (Fig. S2). It is worth noting that CA4-B strains possess the longest PBS region reported so far (about 30 Kbp) and we suggest it represents the ultimate degree of sophistication of this highly specialized region. Indeed, the length of the PBS region seems to have expanded during evolution as the complexity of *Synechococcus* PBS rod composition and chromophorylation progressively increased (Six et al., 2007b).

Contrary to CA4-B islands, the genomic context around CA4-A islands varies widely from clade to clade, though it is more or less constant between subclades (Fig. 1). Interestingly, *psbA* or *hli* genes, which code respectively for the D1 protein of photosystem II and high light-inducible proteins, are frequently found immediately upstream or downstream the latter island. In the case of RS9916, for instance, a complete *psbA* sequence is found upstream the CA4 island while a partial *psbA* sequence, 100% identical to the latter, is found downstream. Both *psbA* and *hli* genes are highly conserved genes and known to be hotspots for intragenomic homologous recombination (Garczarek et al., 2008). Both are also frequently found in cyanophages and are known to have a key role in phage-host interactions during infection (Lindell et al., 2004; Lindell et al., 2005; Sullivan et al., 2006). Another important point is that phylogenetic analyses of CA4 genes (*mpeZ*, *fciA*, *fciB* and *unk10*) from the subcluster 5.3 strain RCC307, consistently place it very close to RS9916 (subcluster 5.1, clade IX; Figs. 2-3 and S2). This strongly suggests a lateral transfer of a CA4-B island from a clade IX strain to RCC307 or one of its close ancestors.

Altogether, this information can be integrated into a coherent evolutionary scenario in which an ancestral CA4 island would have initially occurred by duplication and divergence of a previously existing PEB lyase-isomerase gene, possibly the common ancestor of *mpeW*, *Y* and *Z*, followed by the acquisition of a *fciA/B*-like gene, coding for an AraC-like transcriptional regulator, which was also duplicated before the two gene copies diverged. This gene cluster would then have split between two distinct lineages and evolved separately to form a CA4-A island in some lineages and CA4-B in others. Given the presence of the latter island in both MINOS11 (subcluster 5.3) and several strains of subcluster 5.1 and the congruence of the phylogeny obtained with its constitutive genes with a

vertical inheritance, the CA4-B island likely occurred prior to the differentiation of subclusters 5.1 and 5.3.

The comparison of the genomic context around both island types also indicates that the CA4-B type might be the most stable, while CA4-A islands could represent a more dynamic form of transmission of the CA4 process. Indeed, CA4-A islands are likely to be more easily transferred between *Synechococcus* lineages than their CA4-B counterparts, because of the presence of recombination hotspots in their close vicinity, facilitating genetic transfer by conjugation or via viral intermediates (Lindell et al., 2004; Garczarek et al., 2008). Acquisition of such a CA4-A island by lateral transfer could have conferred the ability to chromatically acclimate to strains that previously had a fixed PBS pigmentation. However, it is also probable that such a transfer only worked if the recipient strain possesses all other found in the PBS region of actual CA4 strains (notably *cpeF*; see Fig. 4 in Six et al., 2007b). In contrast, pigment type 3a strains likely miss too many PBS genes to be able to chromatically acclimate after acquiring a CA4-B island. Strain WH8016 might represent an example of such a failed lateral transfer. Its PBS region is clearly that of a pigment type 3a strain (Fig. S2) and this is independently confirmed by phylogenetic analyses of various PBS genes, such as *cpeY* or *mpeY* (Fig. 3), which unambiguously group WH8016 with WH7803. The absence of *mpeZ* and the presence of a stop codon in *fciA* in WH8016 (Fig. 1) may thus appear as signs of degeneration of the likely useless CA4 island, although we cannot exclude that the transfer was initially incomplete and/or imperfect.

#### *A revised classification of Synechococcus pigment types*

A further outcome from this work was that it allowed us to refine the classification of *Synechococcus* pigment types initially proposed by Six et al. (2007b) and which was based on the phycobiliprotein composition of PBS rods: PC only in type 1; PC and PEI in type 2; PC, PEI and PEII in type 3 (Table 1). Additionally, the  $Exc_{495:545}$  ratio measured at LBL and LGL was used to delineate four subtypes within pigment type 3:  $<0.5$  for type 3a;  $\sim 0.6-0.8$  for type 3b;  $>1.6$  for type 3c; variable between  $\sim 0.7$  in GL and  $\sim 1.6$  in BL for type 3d (or CA4; Six et al., 2007b).

Based on the phenotypic characterization performed in the present study, we propose i) to set the limit of type 3a to  $<0.6$  to fit the complete range of  $Exc_{495:545}$  values observed in MVIR-18.1 (Fig. 4), ii) to extend the range of  $Exc_{495:545}$  ratio for pigment type 3b ( $0.6 \leq Exc_{495:545} < 1.6$ ) to take into account CA4 (and possibly pigment type 3c) strains altered in their PUB chromophorylation, and iii) to distinguish CA4 strains displaying a high (type 3d) or low (type 3e) amplitude of variation of the  $Exc_{495:545}$  ratio during shift from LGL to LBL and reciprocally. Additionally, we propose to complement this phenotype by genomic information by adding the suffix A or B to indicate the CA4 island type,

when present, even for strains not capable to chromatically acclimate likely due to various alterations that totally or partially block the CA4 process. For instance, MVIR-18-1 is now classified as 3aA and MINOS11 as 3dB (Table 1). Note that RCC307 and WH8103, initially classified as 3b and 3c, are now referred to as 3eA and 3bB, respectively.

#### *Novel physiological insights into the functioning of the CA4 process*

The kinetics of the CA4 process depends to some extent upon the population growth rate. Indeed, the half time of the CA4 process ( $T_{50}$ ) is inversely proportional to growth rate for  $\mu$  values lower than about  $0.9 \text{ day}^{-1}$  (Fig. 7). One can therefore predicts that this kinetics will be influenced by a variety of environmental factors that impact on the growth rate, i.e. not only irradiance changes (as shown here), but also temperature or nutrient deficiency. Nitrogen-poor waters are thus likely particularly inhospitable to marine *Synechococcus* performing chromatic acclimation.

The delay in the initiation of CA4 observed in *Synechococcus* spp. BL107 and CC9902 after a shift from BL to GL is intriguing. Either the GL signal is slow to be transduced to the receptors or the BL signal persists for some time, preventing the triggering of the response to GL. In the former hypothesis, the slowing down may originate from an impairment of the GL photoreception system and/or of one of the transduction chain component (e.g. a kinase) of the GL signal. In the second hypothesis, the lifetime of the BL-signaling molecule(s) could be longer than in other strains, possibly because of low degradation rate of the molecule bearing this signal. Because of this unique behavior, this strain could be an interesting model for future research that will aim at deciphering the regulation network of the CA4 process.

## **EXPERIMENTAL PROCEDURES**

### *Strains and growth conditions*

Clonal strains of marine *Synechococcus* used in this study were obtained from the Roscoff Culture Collection (<http://www.sb-roscoff.fr/Phyto/RCC/index.php>). The characteristics and origins of strains are listed in Table S1. They were grown at  $22 \pm 1$  °C in 50 mL polycarbonate culture flasks in the seawater-based medium PCR-S11 (Rippka et al., 2000b), supplemented with 1 mM  $\text{NaNO}_3$  as an additional source of nitrogen. Seawater was reconstituted from Red Sea Salts™ (Houston, Texas, USA) using distilled water. Lee filters (Andover, Hampshire, UK) were used to cover plastic boxes surrounding culture flasks in order to expose them either to blue (Mikkel BLue, #716, 448 nm peak

transmission, 422-484 nm half-height width) or green light (Jas Green, #738, 511 nm peak transmission, 488-593 nm half-height width). Continuous light was provided by Daylight F58W/54-765<sub>TB</sub> fluorescence tubes (Sylvania, Gennevilliers, France). Experiments were performed under both low light (LL) and high light (HL). Irradiances were adjusted under the colored boxes to  $20 \pm 2$  or  $75 \pm 5$   $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  in LL and HL conditions, respectively, as measured with a QSL-2100 Biospherical Instruments quantameter (San Diego, CA).

*Kinetics measurements, control of growth and optical properties*

All cultured strains were acclimated to experimental light conditions for at least one month under continuous BL or GL, corresponding to at least 20 and 40 generations in LL and HL, respectively (see results). For kinetics studies, an exponentially growing batch culture ( $2 \cdot 10^7$  to  $2 \cdot 10^8$  cells.mL<sup>-1</sup>) was diluted into fresh medium at a final concentration of  $10^6$  cells.mL<sup>-1</sup> then split into three subcultures. In LL conditions, *Synechococcus* strains CC9311, WH8020 and CC9902, which seem to be more sensitive to dilution effect, were inoculated at a final concentration of  $\sim 5 \cdot 10^6$  cells.mL<sup>-1</sup>. For each strain acclimated in BL or GL, one aliquot was kept in the original condition (control), whereas the two experimental subcultures were shifted to the other light color condition, when the  $\text{Exc}_{495:545\text{nm}}$  (see below) was measurable with a sufficient accuracy. Shifts were generally performed at time zero in LL, and 24 h after inoculation in HL. Measurements were then made every 12 h in HL and every day in LL. Fluorescence excitation spectra were recorded for an emission at 575 nm using a Perkin Elmer LS-50B spectrofluorimeter and used to calculate the  $\text{Exc}_{495:545}$  ratio, a proxy to the PUB to PEB ratio. Moreover for determining cell concentrations, a 500  $\mu\text{L}$  aliquot was fixed with glutaraldehyde (0.25% final concentration, Grade II, Sigma Aldrich, Saint-Louis, MO, USA), incubated for 10 min in the dark and frozen at  $-80^\circ\text{C}$  until flow cytometric analyses. Thawed samples were analyzed after dilution into sterile medium, using a FACS Canto (Becton Dickinson Biosciences, San Jose, CA, USA) in the presence of 0.95  $\mu\text{m}$  standard fluorescent microspheres (Polysciences Warrington, PA, USA) as previously described (Marie et al., 1999). Growth rates ( $\mu$  in d<sup>-1</sup>) were computed from cell number variations using the following calculation:

$$\mu = \text{Ln} (N(t_2) / N(t_1)) / (t_2 - t_1) \text{ (Eq. 1),}$$

where  $N(t)$  is the cell concentration at time  $t_2$  and  $t_1$  taken at 24 h interval. The half maximal acclimation time ( $T_{50}$  in d) was used as a proxy of the acclimation time.  $T_{50}$  was calculated using using the regression wizard of SigmaPlot (Systat Software, San Jose, CA) by fitting a sigmoidal non-linear regression to data with the following equation:

$$f = \text{min} + (\text{max} - \text{min}) / (1 + 10^{\log T_{50} - x} \times a) \text{ (Eq. 2),}$$

where  $a$  is the slope of the curve. Minimum (*min*) and maximum (*max*) values were defined as average values of the Exc<sub>495:545</sub> ratio of the GL and BL controls during the experiment.

Kinetics of acclimation over successive shifts was also monitored in *Synechococcus* strain BL107. During the whole experiment, the two biological replicates were maintained in exponential growth phase by diluting cultures every two days to a final concentration of  $\sim 8.10^6$  cells.mL<sup>-1</sup>. Fluorescence emission spectra were measured between 540 and 600 nm with excitation at 515 nm. The dilution did not affect growth and chromatic acclimation rates.

#### *Comparative genomics and sequence analysis*

The presence of genes involved or potentially involved in chromatic acclimation was checked in marine *Synechococcus* genomes, by BLASTP searches (Altschul et al., 1997) using BioEdit v7.0.9.0 (Hall, 1999), and both CA4 and PBS gene region were visualized using the Artemis software (Rutherford et al., 2000). The dataset comprised all available marine *Synechococcus* genomes, draft genomes of *Synechococcus* sp. WH8109 (ACNY00000000) and WH8016 (AGIK00000000), as well as unpublished draft genomes of marine *Synechococcus*, coming from preliminary data of unassembled contigs. The presence of protein domains and motifs was checked using InterProScan v4.8 on the EMBL-EBI database, SMART 7 (Letunic et al., 2012) and Phyre2 (Kelley and Sternberg, 2009). CA4 regions of unpublished genomes were deposited in the GenBank nucleotide sequence database under the following accession numbers : XXXXXXXX–YYYYYY (Table S1).

#### *Phylogenetic analyses*

Protein sequences were retrieved for BLASTP searches on marine *Synechococcus* genomes and on the non-redundant protein database (nr) from the National Center for Biotechnology Information (NCBI; release 192.0). Sequences were aligned using MAFFT (G-INS-I option, v 6.953; Katoh et al., 2005) with gap opening penalty and offset set at default values, and 1000 iterations. Poorly aligned regions and gaps were removed by trimAL (v1.4; Capella-Gutiérrez et al., 2009). Phylogenetic reconstructions were performed using 4 different methods: Neighbor Joining (NJ), Maximum Parsimony (MP), Maximum likelihood (ML) and Bayesian inference (BI). NJ and PARS analyses were performed using Phylip 3.69 as previously described (Mella-Flores et al., 2011). ProtTest 3.2 was used for selection of best-fit models of amino acid replacement (Darriba et al., 2011) according to the Akaike information criterion (AIC; Posada and Buckley, 2004) and Bayesian Information Criterion (BIC; Schwarz, 1978). ML reconstructions were performed using PhyML (v3.0; Guindon et al., 2009) with the Le and Gascuel substitution model with an estimated  $\Gamma$  distribution

parameter (LG+G model; Le and Gascuel, 2008), 1000 bootstrap replicates, four substitution-rate categories and a BIONJ starting tree.

Posterior probability values (PP) were generated using MrBayes (v3.1.2; Huelsenbeck and Ronquist, 2001), a mixed-model option, a swap frequency of 1, with four chains, a random starting tree and sampling every 100th generation. PP were generated from 1,000,000 generations for *unk10* and the first 100,000 generations were removed as burn-in. For genes encoding phycobilin lyases and genes containing an AraC-type domain, 1,500,000 generations were performed and the first 1,500 trees were discarded. Trees were visualized using FigTree v1.3.1.

#### *Real time quantitative PCR*

For transcript abundances, experiments were performed on triplicates on LL acclimated strains. Exponentially growing cultures, at a density of approximately  $1-5 \times 10^7$  cells mL<sup>-1</sup>, were harvested and extracted as previously described (Kolowrat et al., 2010). Briefly, after addition of Pluronic (as described above), 150 mL cultures were centrifuged at 9000 x g for 7 min. Cell pellets were resuspended in 500 µL Trizol (Invitrogen, Carlsbad, CA, USA), immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. Samples were kept on ice during all steps. RNA extractions were performed with the miRNAeasy kit (Qiagen), following manufacturer instructions, in order to recover both large and small RNAs. Residual DNA was removed by two successive DNase treatments on the miRNAeasy columns with the Qiagen RNase-free DNase set (Qiagen). RNA samples were eluted in DEPC-treated water, quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DA), and kept at -80°C until analysis.

Real time qPCR analyses were carried out on 12 *Synechococcus* strains harboring *mpeZ* or *mpeW* genes. Primer Express™ (v2.0; Applied Biosystems, Foster City, CA) was used to design gene-specific primers (Table S3) and optimization was performed by checking for each primer set the linearity of the cycle at threshold (CT) vs. cDNA content within a dilution range, as previously described (Six et al., 2007a). These primers were used at a final concentration of 300 mM, excepted for RS9915 for which this latter was increased to 600 mM. Reverse transcription was carried out using SuperScriptII reverse transcriptase (Gibco-BRL, Gaithersburg, MD, USA) on 100 ng RNA. qPCR was performed on each biological replicate, using the 480 II (Roche, Boulogne-Billancourt, France) and the absolute SYBR Green ROX Mix (Abgene, Epsom, UK) as previously described (Garczarek et al., 2008). Resulting data were analysed using the comparative CT ( $\Delta\Delta CT$ ) and normalized using the aperiodic gene *rnpB* as an internal standard (Mary and Vaultot, 2003; Zinser et al., 2009). This latter was amplified using primers qProSyn\_rnpB\_F (GGCAAGGGTGCAARGGTG) and qProSyn\_rnpB\_R (GCCGGGGTTTACCKAGCC) at a final concentration of 600 mM.

## AKNOWLEDGEMENTS

This work was supported by the French “Agence Nationale de la Recherche” Microbial Genomics Programme (PELICAN, ANR-09-GENM-030), the European Union's Seventh Framework Programmes FP7 MicroB3 and MaCumba (grant agreements 287589 and 311975, respectively). We are grateful to Anton Post who provided unpublished genomic data from *Synechococcus* sp. WH8020. We thank the Genoscope, particularly Karine Labadie and Jean-Marc Aury, for whole genome sequencing of 25 additional *Synechococcus* strains and assembling of preliminary data, in the framework of the METASYN project. We warmly thank the Roscoff Culture Collection for maintaining the *Synechococcus* strains used in this study.

None of the authors have a conflict of interest

## REFERENCES

- Adir, N.** (2005) Elucidation of the molecular structures of components of the phycobilisome: reconstructing a giant. *Photosynth Res* **85**: 15-32.
- Ahlgren, N.A., and Rocap, G.** (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* **72**: 7193-7204.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25**: 3389-3402.
- Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W., and Bork, P.** (2001) Comparison of ARM and HEAT protein repeats. *J Mol Biol* **309**: 1-18.
- Aravind, L., and Ponting, C.P.** (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Bioch Sci* **22**: 458.
- Biswas, A., Boutaghou, M.N., Alvey, R.M., Kronfel, C.M., Cole, R.B., Bryant, D.A., and Schluchter, W.M.** (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.
- Blot, N., Wu, X.J., Thomas, J.C., Zhang, J., Garczarek, L., Bohm, 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.
- Bretaudeau, A., Coste, F., Humily, F., Garczarek, L., Le Corguillé, G., Six, C., Ratin, M., Collin, O., Schluchter, W.M., and Partensky, F.** (2013) CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. *Nucl Acids Res* **41**: D396-401.
- Buitenhuis, E., Li, W., Lomas, M., Karl, D., Landry, M., Partensky, F., Karl, D.M., Ulloa, O., Jacquet, S., Lantoine, F., Chavez, F., Macias, D., Gosselin, M. and McManus, G.B.** (2012) Picophytoplankton biomass distribution in the global ocean. *Earth System Science Data* **4**: 37-46.
- Capella-Gutiérrez, S., Silla-Martinez, J.M., and Gabaldon, T.** (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D.** (2011) ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**: 1164-1165.

- Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A.F., Hess, W.R., and Partensky, F. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: 16.
- Egan, S.M., and Schleif, R.F. (1993) A regulatory cascade in the induction of *rhaBAD*. *J Mol Biol* **234**: 87-98.
- Everroad, C., Six, C., Partensky, F., Thomas, J.C., Holtzendorff, J., and Wood, A.M. (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp. *J Bacteriol* **188**: 3345-3356.
- Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* **69**: 2430-2443.
- Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J.L. (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* **61**: 393-410.
- Garczarek, L., Dufresne, A., Blot, N., Cockshutt, A.M., Peyrat, A., Campbell, D.A., Joubin, L., and Six, C. (2008) Function and evolution of the *psbA* gene family in marine *Synechococcus*: *Synechococcus* sp. WH7803 as a case study. *ISME J* **2**: 937-953.
- Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V., and Gascuel, O. (2009) PhyML: Fast and accurate phylogeny reconstruction by Maximum Likelihood. *Infect Genet Evol* **9**: 384-385.
- Gutu, A., and Kehoe, D.M. (2011) Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. *Mol Plant* **5**: 1-13.
- Hall, T. (1999) BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95-98.
- Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., and Stal, L.J. (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* **10**: 174-188.
- Huang, S., Wilhelm, S.W., Rodger Harvey, H., Taylor, K., Jiao, N., and Chen, F. (2011) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6**: 285-297.
- Huelsenbeck, J.P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Jacobson, M.R., Brigle, K.E., Bennett, L.T., Setterquist, R.A., Wilson, M.S., Cash, V.L., Beynon, J., Newton, W.E., and Dean, D.R. (1989) Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *J Bacteriol* **171**: 1017-1027.
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucl Acids Res* **33**: 511-518.
- Kehoe, D.M., and Gutu, A. (2006) Responding to color: The regulation of complementary chromatic adaptation. *Ann Rev Plant Biol* **57**: 127-150.
- Kelley, L.A., and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363-371.
- Kirk, J.T.O. (1994) Light and photosynthesis in aquatic ecosystems. Cambridge, UK: Cambridge University Press.
- Kolowrat, C., Partensky, F., Mella-Flores, D., Le Corguillé, G., Boutte, C., Blot, N., Ratin, M., Férréol, M., Lecomte, X. and Gourvil, P. (2010) Ultraviolet stress delays chromosome replication in light/dark synchronized cells of the marine cyanobacterium *Prochlorococcus marinus* PCC9511. *BMC Microbiol* **10**: 204.
- Le, S.Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**: 1307-1320.
- Letunic, I., Doerks, T., and Bork, P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **40**: D302-D305.

- Li, W.K.W.** (1994) Primary productivity of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169-175.
- Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W.** (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86-89.
- Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W.** (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA* **101**: 11013-11018.
- Liu, H., Campbell, L., Landry, M.R., Nolla, H.A., Brown, S.L., and Constantinou, J.** (1998) *Prochlorococcus* and *Synechococcus* growth rates and contributions to production in the Arabian Sea during the 1995 Southwest and Northeast monsoons. *Deep Sea Res II* **45**: 2327-2352.
- MacColl, R.** (1998) Cyanobacterial phycobilisomes. *J Struct Biol* **124**: 311-334.
- Marie, D., Brussaard, C., Partensky, F., Vaultot, D., and Wiley, J.** (1999) Flow cytometric analysis of phytoplankton, bacteria and viruses. *Current Protocols in Cytometry Supplement* **10**: 11.11.11-11.11.15.
- Mary, I., and Vaultot, D.** (2003) Two-component systems in *Prochlorococcus* MED4: Genomic analysis and differential expression under stress. *FEMS Microbiol Lett* **226**: 135-144.
- Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J.** (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372-386.
- Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahe, F., Bariat, L., Courties, C., Marie, D., Ras, J., Mauriac, R., Jeanthon, C., Bendif, E.M., Ostrowski, M., Scanlan, D.J. and Garczarek, L.** (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences* **8**: 2785-2804.
- Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., and Tieche, F.** (2007) Optical properties of the "clearest" natural waters. *Limnol Oceanogr* **52**: 217-229.
- Morimoto, K., Sato, S., Tabata, S., and Nakai, M.** (2003) A HEAT-repeats containing protein, IaiH, stabilizes the iron-sulfur cluster bound to the cyanobacterial IscA homologue, IscA2. *J Bioch* **134**: 211-217.
- Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V.** (1990) Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* **35**: 45-58.
- Ong, L.J., and Glazer, A.N.** (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, L.J., Glazer, A.N., and Waterbury, J.B.** (1984) An unusual phycoerythrin from a marine cyanobacterium. *Science* **224**: 80.
- Palenik, B.** (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* **67**: 991-994.
- Partensky, F., Blanchot, G., and Vaultot, D.** (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters : a review. *Bull Inst Océanogr* **19**: 457-475.
- Posada, D., and Buckley, T.R.** (2004) Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* **53**: 793-808.
- Rippka, R., Coursin, T., Hess, W., Lichtlé, C., Scanlan, D.J., Palinska, K.A., Itean, I., Partensky, F., Houmard, J. and Herdman, M.** (2000) *Prochlorococcus marinus* Chisholm et al. 1992 subsp *pastoris* subsp nov strain PCC 9511, the first axenic chlorophyll  $a_2/b_2$ -containing cyanobacterium (Oxyphotobacteria). *Intl J Syst Evol Microbiol* **50**: 1833-1847.
- Rodgers, M.E., and Schleif, R.** (2009) Solution structure of the DNA binding domain of AraC protein. *Proteins: Structure, Function, and Bioinformatics* **77**: 202-208.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B.** (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.
- Schleif, R.** (2010) AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol Rev* **34**: 779-796.

- Schluchter, W.M., Shen, G., Alvey, R.M., Biswas, A., Saunée, N.A., Williams, S.R., Mille, C.A. and Bryant, D.A.** (2010) Phycobiliprotein biosynthesis in cyanobacteria: Structure and function of enzymes involved in post-translational modification. In *Recent Advances in Phototrophic Prokaryotes*: Springer New York, pp. 211-228.
- Schwarz, G.** (1978) Estimating the dimension of a model. *Annal Stat* **6**: 461-464.
- Shukla, A., Biswas, A., Blot, N., Partensky, F., Karty, J.A., Hammad, L.A., Garczarek, L., Gutu, A., Schluchter, W.M. and Kehoe, D.M.** (2012) Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci USA* **109**: 20136–20141.
- Sidler, W.A.** (1994) Phycobilisome and phycobiliprotein structures. In *The molecular biology of cyanobacteria*. Bryant, D.A. (ed). Netherlands: Kluwer Academic Publishers, pp. 139-216.
- Six, C., Thomas, J.C., Brahamsha, B., Lemoine, Y., and Partensky, F.** (2004) Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquat Microb Ecol* **35**: 17-29.
- Six, C., Joubin, L., Partensky, F., Holtzendorff, J., and 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., Scanlan, D.J. and Partensky, F.** (2007b) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.
- Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W.** (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: e234.
- Swanson, R.V., Ong, L.J., Wilbanks, S.M., and Glazer, A.N.** (1991) Phycoerythrins of marine unicellular cyanobacteria. II. Characterization of phycobiliproteins with unusually high phycourobilin content. *J Biol Chem* **266**: 9528-9534.
- Tandeau de Marsac, N.** (1977) Occurrence and nature of chromatic adaptation in cyanobacteria. *J Bacteriol* **130**: 82-91.
- Toledo, G., Palenik, B., and Brahamsha, B.** (1999) Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* **65**: 5247-5251.
- Wickstrum, J.R., Skredenske, J.M., Balasubramaniam, V., Jones, K., and Egan, S.M.** (2010) The AraC/XylS family activator RhaS negatively autoregulates *rhaSR* expression by preventing cyclic AMP receptor protein activation. *J Bacteriol* **192**: 225-232.
- Wilbanks, S.M., and Glazer, A.N.** (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, A.M., Phinney, D.A., and Yentsch, C.S.** (1998) Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol Prog Ser* **162**: 25-31.
- Zhao, K.H., Deng, M.G., Zheng, M., Zhou, M., Parbel, A., Storf, M., Meyer, M., Strohmam, B. and Scheer, H.** (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.
- Zinser, E.R., Lindell, D., Johnson, Z.I., Futschik, M.E., Steglich, C., Coleman, M.L., Wright, M.A., Rector, T., Steen, R. and McNulty, N.** (2009) Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS One* **4**: e5135.
- Zwirgmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaultot, D., Not, F., Massana, R., Ulloa, O. and Scanlan, D.J.** (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147-161

## SUPPLEMENTARY INFORMATION

- Chen, F., Wang, K., Kan, J.J., Bachoon, D.S., Lu, J.R., Lau, S., and Campbell, L.** (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.
- Gomis-Rüth, F.X., Solà, M., Acebo, P., Párraga, A., Guasch, A., Eritja, R., Gonzalez, A., Espinosa, M., del Solar, G. and Coll, M.** (1998) The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J* **17**(24): 7404-7415.
- Le Gall, F., Rigaut-Jalabert, F., Marie, D., Garczarek, L., Viprey, M., Gobet, A., and Vaultot, D.** (2008) Picoplankton diversity in the South-East Pacific Ocean from cultures. *Biogeosciences* **5**: 203-214.
- Palenik, B., Ren, Q.H., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., Durkin, A.S., Daugherty, S.C., Sullivan, S.A., Khouri, H., Mohamoud, Y., Halpin, R., and Paulsen, I.T.** (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* **103**: 13555-13559.
- Waterbury, J.B., Watson, S.W., Valois, F.W., and Franks, D.G.** (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Can Bull Fish Aquat Sci* **214**: 71-120.

**TABLE S1.** Characteristics of marine *Synechococcus* strains and accession numbers of sequences used in this study.

Strain	Strain characteristics									Genomic information			Protein accession numbers							
	RCC # <sup>a</sup>	Sub-cluster <sup>b</sup>	Sub-clade <sup>b</sup>	Isolation site	Lat. (N)	Long. (E)	Date	Depth (m)	Reference	Genome status <sup>c</sup>	Genome accession #	CA4 island accession #	Unk10	FciA	FciB	CpeY	MpeY	MpeZ	MpeW	
CC9311	1086	5.1	Ia	California current	32	-124.5	1-Jan-93	95	Palenik <i>et al.</i> , 2006	Complete	NC_008319		NC_008319	YP_731428	YP_731427	YP_729726	YP_729733	YP_731426	-	
WH8020	2437	5.1	Ia	Sargasso Sea	38.68	-69.34	26-Jun-80	50	Waterbury <i>et al.</i> , 1986	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	-
MVIR-18-1	2385	5.1	Ib	North Sea	61	1.98	23-Jul-07	25	This study	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	-
WH8016	2535	5.1	Ib	Atlantic Ocean	41.52	-70.67	1-Jun-80	nd	Waterbury <i>et al.</i> , 1986	WGS	NZ_AGIK01000001		ZP_08956116	ZP_08956113	ZP_08956112	ZP_08957406	ZP_08957409	-	-	
WH8109	2033	5.1	Ila	Sargasso Sea	39.29	-70.28	1-Jun-81	nd	Waterbury <i>et al.</i> , 1986	WGS	NZ_ACNY00000000		ZP_05789211	ZP_05788624	ZP_05790361	ZP_05789451	ZP_05788234	-	ZP_05790143	
CC9605	753	5.1	Iic	California current	33.3	-124.2	1-Jan-96	51	Dufresne <i>et al.</i> , 2008	WGS	CP000110		YP_380762	-	-	YP_380760	YP_380767	-	-	
A15-62	2374	5.1	Iic	Atlantic Ocean	17.62	-20.97	4-Oct-04	15	Mazard <i>et al.</i> , 2012	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	-	XXX	
PROS-U-1	2369	5.1	Iih	Atlantic Ocean, Moroccan upwelling	30.13	-10.05	12-Sep-99	5	Mazard <i>et al.</i> , 2012	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	-	XXX	
M11.1	2531	5.1	Iih	Gulf of Mexico	27.7	-91.32	15-Feb-04	275	Everroad <i>et al.</i> , 2006	Not seq.	/		-	-	-	-	-	-	-	
RS9915	2553	5.1	IIla	Gulf of Aqaba	29.47	34.94	18-Oct-99	10	Fuller <i>et al.</i> , 2003	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	-	XXX	
WH8102	539	5.1	IIla	Atlantic Ocean	22.48	65.6	15-Mar-81	nd	Waterbury <i>et al.</i> , 1986	Complete	BX548020		BX548020	-	-	NP_898104	NP_898098	-	-	
WH8103	2366	5.1	IIla	Atlantic Ocean	28.5	-67.4	17-Mar-81	nd	Waterbury <i>et al.</i> , 1986	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	-	XXX	
CC9902	2673	5.1	IVa	California current	32.9	-117.3	1-Jan-99	5	Dufresne <i>et al.</i> , 2008	Complete	NC_007513		YP_377567	YP_377568	YP_377569	YP_377899	YP_377893	YP_377570	-	
BL107	515	5.1	IVa	Mediterranean Sea	41.72	3.57	n.d.	1800	Dufresne <i>et al.</i> , 2008	WGS	NZ_AATZ01000000		ZP_01469559	ZP_01469558	ZP_01469557	ZP_01469889	ZP_01469895	ZP_01469555	-	
WH7803	28	5.1	V	Atlantic Ocean	33.73	-67.48	3-Jul-78	25	Waterbury <i>et al.</i> , 1986	Complete	CT971583		-	-	-	YP_001224214	YP_001224217	-	-	
WH7805	1085	5.1	Vic	Atlantic Ocean	33.75	-67.5	30-Jun-78	nd	Waterbury <i>et al.</i> , 1986	Complete	AAOK00000000		-	-	-	ZP_01123337	-	-	-	
RS9916	555	5.1	IX	Gulf of Aqaba	29.28	34.55	22-Nov-99	10	Fuller <i>et al.</i> , 2003	WGS	NZ_AAUA00000000		ZP_01472953	ZP_01472952	ZP_01472951	ZP_01473085	ZP_01473091	ZP_01472949	-	
BIOS-U3-1	2533	5.1	CRD1a	Pacific Ocean, Chile upwelling	-33.98	-73.39	6-Dec-04	5	Le Gall <i>et al.</i> , 2008	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	-
BIOS-E4-1	2534	5.1	CRD1b	South East Pacific Ocean	-31.87	-91.44	28-Nov-1994	40	This study	WGS	unpublished	XXX	XXX	-	-	XXX	-	XXX	-	
CB0205	-	5.2	CB5	Atlantic Ocean, Chesapeake bay	37.46	-76.14	nd	nd	Chen <i>et al.</i> , 2004	WGS	ADXM00000000		-	-	-	ZP_07970129	-	-	-	
RCC307	307	5.3	5.3-I	Mediterranean Sea	39.1	6.1	26-May-1996	15	Dufresne <i>et al.</i> , 2008	Complete	NC_009482		NC_009482	YP_001228263	YP_001228262	YP_001228309	YP_001228303	YP_001228260	-	
MINOS11	2319	5.3	5.3-I	Mediterranean Sea	34	18.02	19-Jun-96	20	Mazard <i>et al.</i> , 2012	WGS	unpublished	XXX	XXX	XXX	XXX	XXX	XXX	-	XXX	

<sup>a</sup> Strain number in the Roscoff culture collection (<http://www.sb-roscoff.fr/Phyto/RCC>)

<sup>b</sup> Phylogenetic affiliation following the nomenclature reported in previous studies (Fuller *et al.*, 2003; Ahlgren *et al.*, 2006; Dufresne *et al.*, 2008; Huang *et al.*, 2012; Mazard *et al.*, 2012a)

<sup>c</sup> WGS, Whole Genome Sequencing (scaffolds only); Not seq., Not sequenced. See results of genomic comparisons

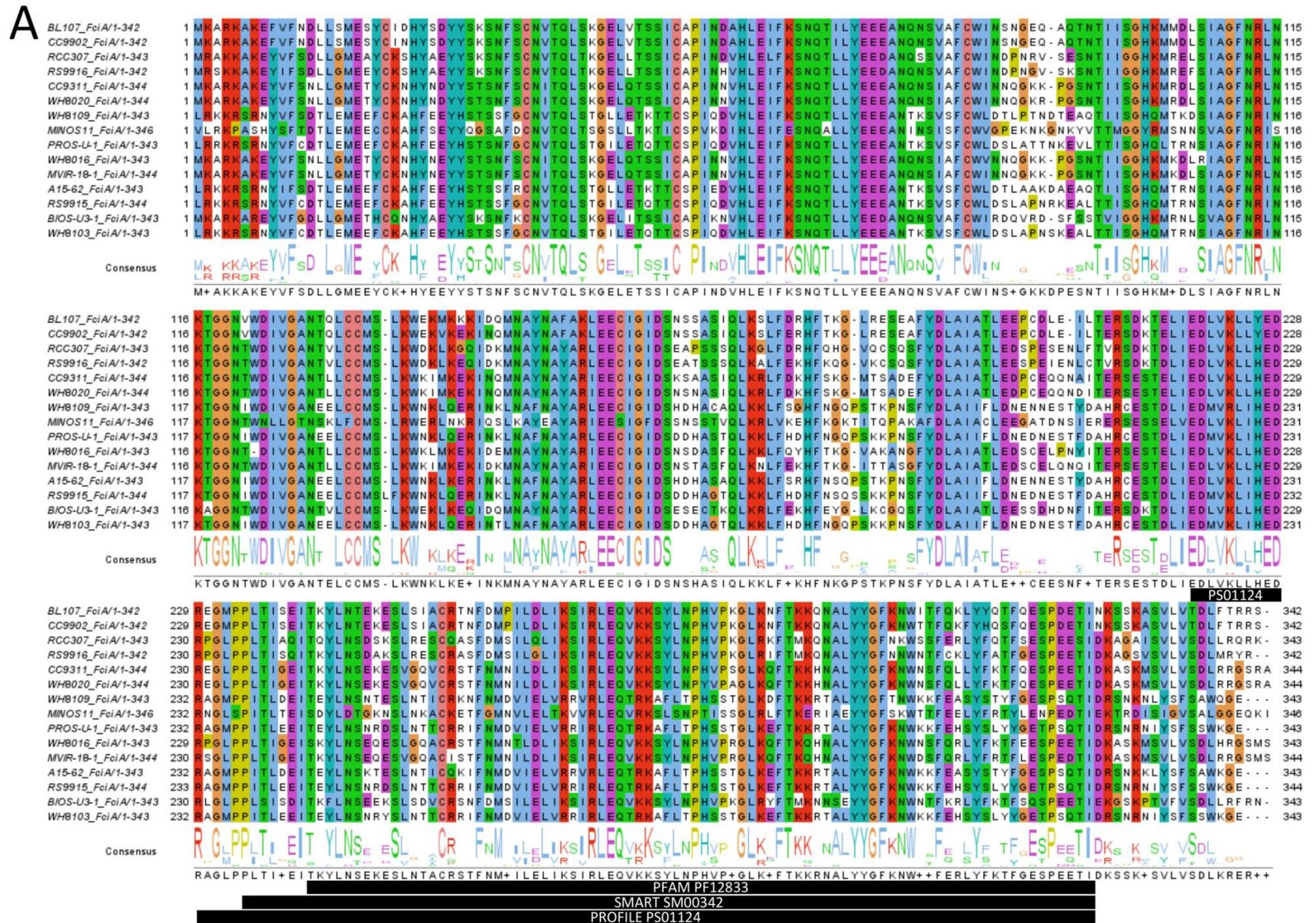
**TABLE S2. Kinetics of acclimation of twelve marine *Synechococcus* strains after shifting cultures to the other light color.** Growth rate ( $\mu$ ) were assessed from flow cytometric counts while  $T_{50}$  was calculated from  $Exc_{495:545}$  spectrofluorimetric measurements. Values indicated for individual strains are averages of two biological replicates ( $\pm$  mean deviation) calculated from two biological replicates. Averages ( $\pm$  standard deviation) were also calculated for the different sets of strains gathered by CA4 type or physiological behavior group (see text).

Strain	LL				HL			
	BGL		GBL		BGL		GBL	
	$\mu$ ( $d^{-1}$ )	$T_{50}$ (d)						
<b>A15-62</b>	0.67 $\pm$ 0.01	1.96 $\pm$ 0.05	0.64 $\pm$ 0.01	3.25 $\pm$ 0.00	0.90 $\pm$ 0.00	1.34 $\pm$ 0.01	0.90 $\pm$ 0.00	2.30 $\pm$ 0.01
<b>BIOS-U3-1</b>	0.72 $\pm$ 0.02	1.95 $\pm$ 0.08	0.54 $\pm$ 0.00	2.90 $\pm$ 0.06	0.94 $\pm$ 0.01	1.49 $\pm$ 0.04	0.88 $\pm$ 0.00	1.45 $\pm$ 0.05
<b>BL107</b>	0.44 $\pm$ 0.00	6.03 $\pm$ 0.03	0.52 $\pm$ 0.02	2.51 $\pm$ 0.00	1.11 $\pm$ 0.02	2.41 $\pm$ 0.02	1.19 $\pm$ 0.14	1.07 $\pm$ 0.00
<b>CC9311</b>	0.65 $\pm$ 0.02	2.85 $\pm$ 0.04	0.55 $\pm$ 0.03	2.35 $\pm$ 0.02				
<b>CC9902</b>	0.60 $\pm$ 0.00	3.50 $\pm$ 0.00	0.54 $\pm$ 0.00	2.93 $\pm$ 0.03	0.88 $\pm$ 0.02	2.12 $\pm$ 0.01	0.61 $\pm$ 0.08	2.21 $\pm$ 0.06
<b>M11.1</b>	0.49 $\pm$ 0.06	2.36 $\pm$ 0.04	0.60 $\pm$ 0.00	2.30 $\pm$ 0.01	1.08 $\pm$ 0.02	1.19 $\pm$ 0.00	0.96 $\pm$ 0.06	1.71 $\pm$ 0.01
<b>MINOS11</b>	0.71 $\pm$ 0.01	1.80 $\pm$ 0.02	0.52 $\pm$ 0.01	2.76 $\pm$ 0.07	1.05 $\pm$ 0.00	1.05 $\pm$ 0.05	1.05 $\pm$ 0.00	1.91 $\pm$ 0.12
<b>RCC307</b>	0.64 $\pm$ 0.02	2.85 $\pm$ 0.19	0.45 $\pm$ 0.01	4.29 $\pm$ 0.24				
<b>RS9915</b>	0.73 $\pm$ 0.05	2.21 $\pm$ 0.10	0.75 $\pm$ 0.09	2.58 $\pm$ 0.04	1.60 $\pm$ 0.00	1.32 $\pm$ 0.02	1.61 $\pm$ 0.11	1.37 $\pm$ 0.03
<b>RS9916</b>	1.29 $\pm$ 0.01	1.91 $\pm$ 0.07	0.66 $\pm$ 0.02	1.83 $\pm$ 0.00	1.56 $\pm$ 0.08	1.36 $\pm$ 0.04	1.58 $\pm$ 0.02	1.75 $\pm$ 0.09
<b>PROS-U-1</b>	0.41 $\pm$ 0.00	1.86 $\pm$ 0.10	0.42 $\pm$ 0.01	3.38 $\pm$ 0.02	1.27 $\pm$ 0.00	1.12 $\pm$ 0.01	1.25 $\pm$ 0.03	1.97 $\pm$ 0.05
<b>WH8020</b>	0.48 $\pm$ 0.01	3.18 $\pm$ 0.02	0.34 $\pm$ 0.01	4.65 $\pm$ 0.00	0.73 $\pm$ 0.00	2.19 $\pm$ 0.03	0.56 $\pm$ 0.02	2.45 $\pm$ 0.08
<b>All strains</b>	0.65 $\pm$ 0.23	2.71 $\pm$ 1.17	0.55 $\pm$ 0.11	2.98 $\pm$ 0.80	1.11 $\pm$ 0.28	1.56 $\pm$ 0.48	1.06 $\pm$ 0.36	1.82 $\pm$ 0.43
<b>CA4-A island</b>	0.69 $\pm$ 0.27	3.18 $\pm$ 1.34	0.52 $\pm$ 0.10	3.07 $\pm$ 1.00	1.04 $\pm$ 0.30	1.91 $\pm$ 0.44	0.96 $\pm$ 0.41	1.79 $\pm$ 0.53
<b>CA4-B island</b>	0.60 $\pm$ 0.14	2.04 $\pm$ 0.23	0.59 $\pm$ 0.12	2.86 $\pm$ 0.43	1.18 $\pm$ 0.25	1.20 $\pm$ 0.12	1.16 $\pm$ 0.28	1.85 $\pm$ 0.33
<b>Phenotypic group 1</b>	0.60 $\pm$ 0.14	2.04 $\pm$ 0.23	0.59 $\pm$ 0.12	2.86 $\pm$ 0.43	1.18 $\pm$ 0.25	1.20 $\pm$ 0.12	1.16 $\pm$ 0.28	1.85 $\pm$ 0.33
<b>Phenotypic group 2</b>	0.75 $\pm$ 0.30	2.68 $\pm$ 0.68	0.53 $\pm$ 0.11	2.93 $\pm$ 1.00	1.03 $\pm$ 0.34	1.79 $\pm$ 0.40	0.91 $\pm$ 0.44	1.97 $\pm$ 0.42
<b>Phenotypic group 3</b>	0.52 $\pm$ 0.09	4.77 $\pm$ 1.46	0.53 $\pm$ 0.02	2.72 $\pm$ 0.24	0.99 $\pm$ 0.13	2.26 $\pm$ 0.17	0.90 $\pm$ 0.36	1.64 $\pm$ 0.66

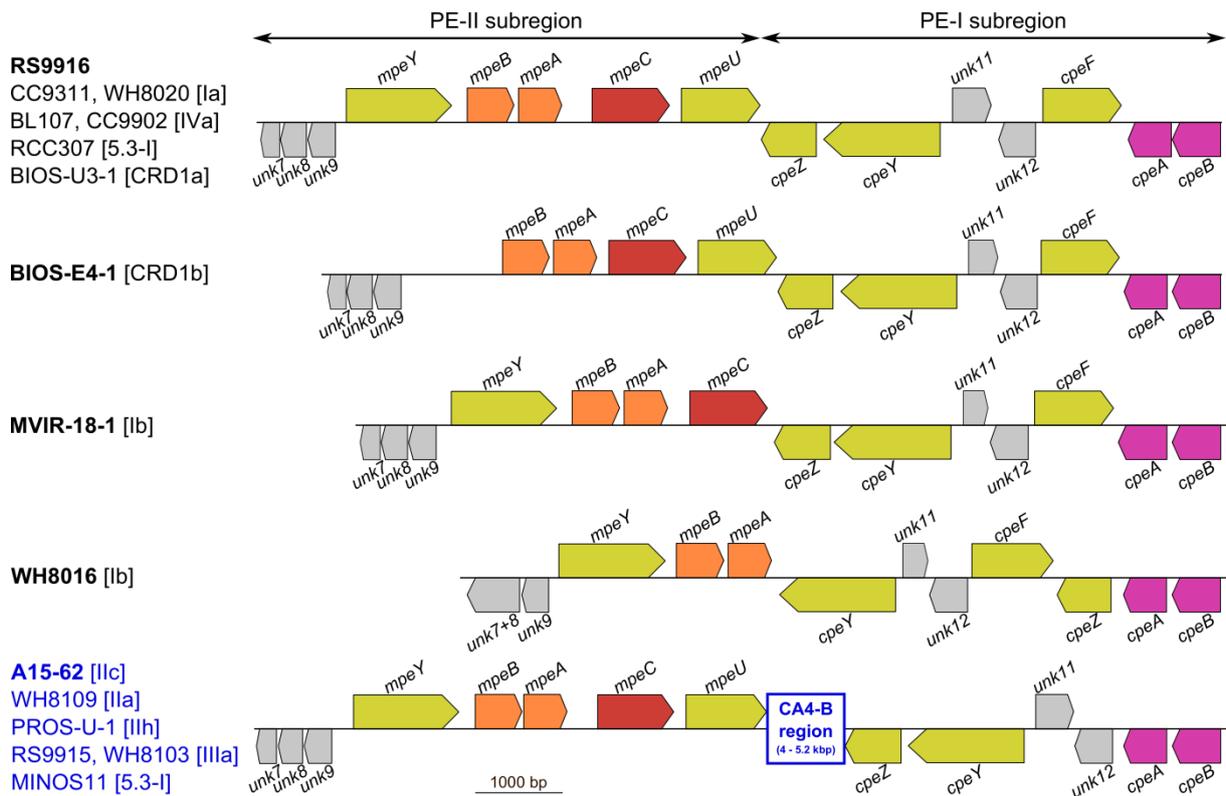
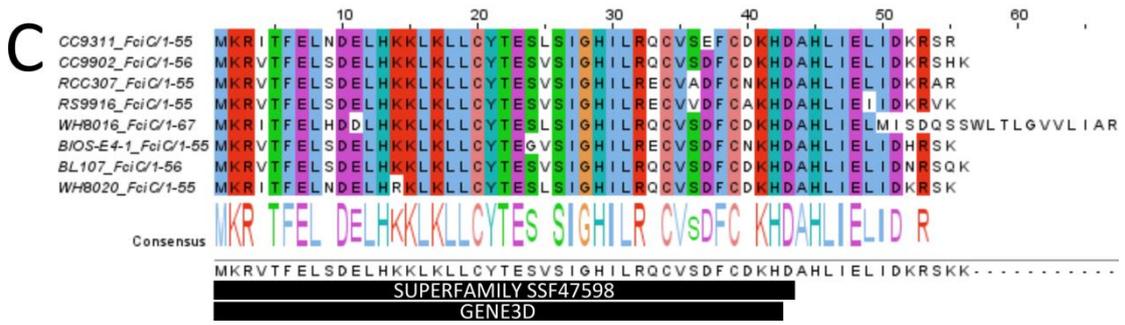
**TABLE S3. List of primers used for real time PCR reactions.**

Gene name	<i>mpeW</i>			
Strain name	Primer_Fw_name	Primer_Fw_sequence	Primer_Rev_name	Primer_Rev_sequence
WH8109	qWH8109_mpeW-110F	CAACACGGTTGGGAGCTTGT	qWH8109_mpeW-172R	CGGTAGCCGCTGCGACTA
A15-62	qA15-62_mpeW-817F	TGTGATTTCCTTAAGCATAGAGATGAG	qA15-62_mpeW-892R	TTGAAGCTGGCCAACTGAATAA
RS9915	qRS9915_mpeW-558F	AGATCACCTCGAAAATACCAATGTT	qRS9915_mpeW-629R	GGCTCACCAGCATCACC AAT
WH8103	qWH8103_mpeW-627F	GCCGGCTCTTTTGTGCAATA	qWH8103_mpeW-704R	GGAGTTATCTTGAACGCACCTTTTAGC
MINOS11	qMinos11_mpeW-512F	CCCTTGCCTATTTGGTTTCTGA	qMinos11_mpeW-574R	GGCGTCACTCAAGCTCTCT
	<i>mpeZ</i>			
CC9311	qCC9311_mpeZ-558F	TCCTTTGGTCCACAACCTTACTGA	qCC9311_mpeZ-633R	ACCAGCATCACCTAAGTCAATCAC
MVIR18-1	qMVIR18-1_mpeZ-1124F	TGATCTACGAACTTTCTAATTCTGCTTT	qMVIR18-1_mpeZ-1195R	GCCGTGCCATCACCTGTT
BL107	qBL107_mpeZ-53F	TCTCCTATTATTGGCTCAAACG	qBL107_mpeZ-123R	AAGTCGAGTGGCAGCGAAAT
RS9916	qRS9916_mpeZ-84F	GCCACGCCCTCAGATC	qRS9916_mpeZ-146R	GTCGCCCACTAGAACACAA
BIOS-U3-1	qBIOS-U3-1_mpeZ-1084F	TGGGCGTGCCTTGAGTT	qBIOS-U3-1_mpeZ-1150R	GTGCCGATGAAGAGAGCTCAT
BIOS-E4-1	qBIOS-E4-1_mpeZ-629F	CTGGCGTTGAAGCTCTCCT	qBIOS-E4-1_mpeZ-691R	TAGCCTGAGGGACATTGAAA
RCC307	qRCC307_mpeZ-1039F	TTAGCGGAAACAACCTCAGTAC	qRCC307_mpeZ-1108R	GTTGCAGCCGCACACAAG

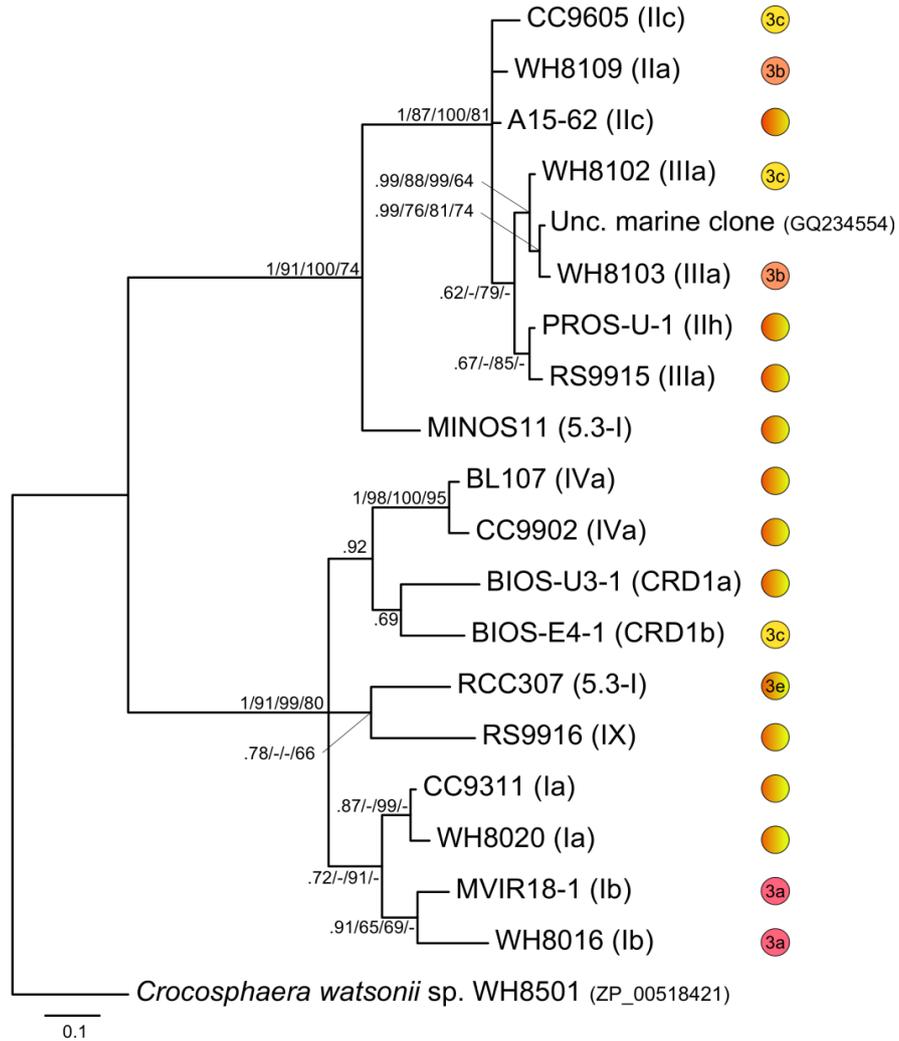
**FIG. S1.** Multiple sequence alignments of putative transcriptional regulators FciA (A) and FciB (B) containing an  $\alpha$ -helix-turn- $\alpha$ -helix (HTH) domain of AraC-type and FciC with a predicted ribbon-helix-helix of marine *Synechococcus*. Shading represents identical amino acid (>70%) using Jalview. Specific proteins signatures, framed by a black rectangle, were determined using InterProScan and sequences from *Synechococcus* sp. RS9916. *Ab initio* modelling using Phyre2 indicates that both 77% of FciA (263 of 342 residues) and 50% of FciB (158 of 316 residues) from *Synechococcus* sp. RS9916 were modelled with >90% confidence to the AraC/XylS family. Furthermore, 78% of FciC (42 of 55 residues) were modelled with the CopG-like family (Gomis-Rüth et al., 1998).



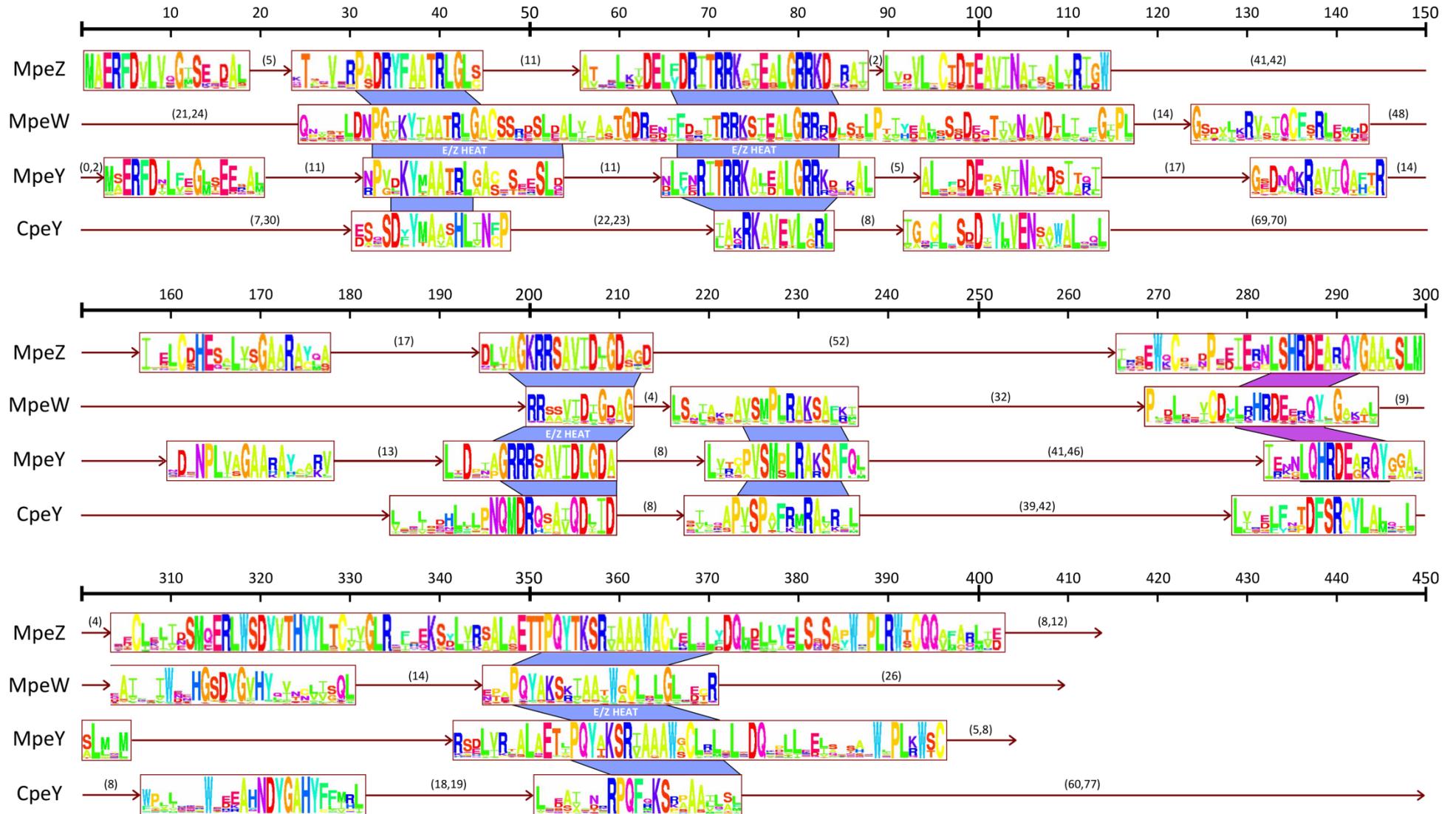




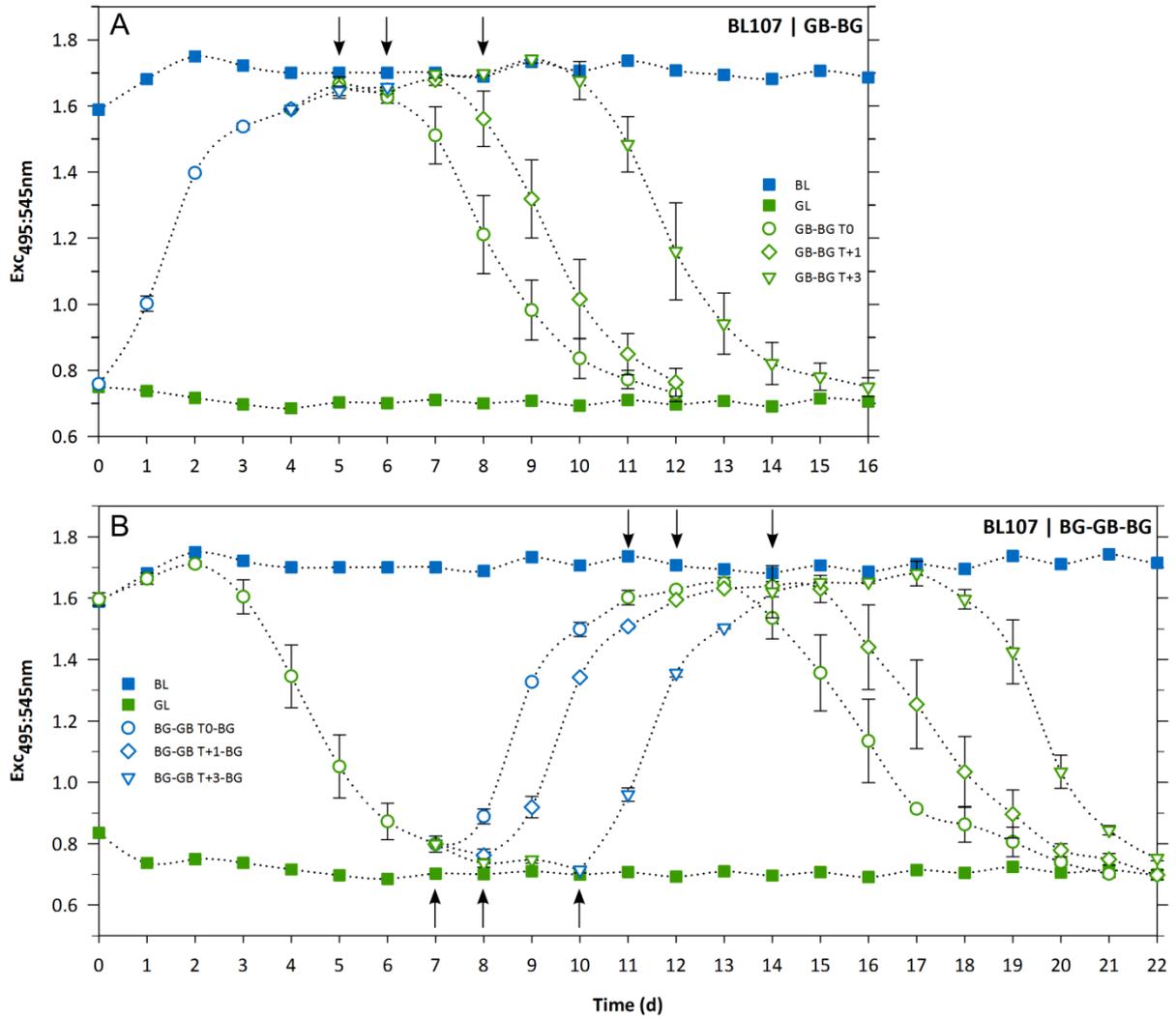
**FIG. S2.** Genomic comparisons of PE rod gene regions (including the whole PEII subregion) of marine *Synechococcus* with a CA region. Note that the gene organization of WH8016 PBS region is identical to that of the pigment type 3a strain WH7803. For readability, the whole PBS region is not shown. See Fig.6 in Six et al. (2007b).



**FIG. S3. Bayesian analysis of *unk10* amino acid sequences (108 aligned positions) from marine *Synechococcus*.** For each *Synechococcus* cultured strain phylogenetic affiliation is mentioned in brackets and the pigment type is indicated by colored circles. The tree is rooted using the sequence from *Crocosphaera watsonii* sp. WH8501. Series of four numbers shown at nodes correspond to Bayesian posterior probabilities (PP, ranging between 0 and 1), bootstrap values for ML analyses, Neighbor-Joining and Parsimony methods, respectively. Bootstraps, represented as a percentage, were obtained through 1,000 repetitions and PP from 1,000,000 generations. Only values higher than 0.60 for PP and 60% for bootstrap values are shown on the phylogenetic tree. The scale bar represents 0.1 substitutions per nucleotide.



**FIG. S4.** Motifs for phycobilin lyases members of the E/F clan of marine *Synechococcus*. Motifs were created by Protomata learner (v 0.07, <http://tools.genouest.org/tools/protomata/learn/>). E/Z-repeats domains shared by these proteins were highlighted in blue color. A motif shared by the first set of proteins, but of unknown function, is indicated in pink. All sequences available for each protein were used for motif design. Numbers into brackets refer to number of residues between the different motifs.



**FIG. S5.** Effect of successive color shifts on the kinetics of chromatic acclimation of *Synechococcus* sp. BL107. **A** : Effects of the time of BL acclimation on the delay of CA4 initiation, as assessed by significant changes in the  $Exc_{495:545}$  ratio. For this experiment, GL acclimated cultures were first shifted to BL and once they had reached the  $Exc_{495:545}$  ratio in BL, they were shifted back to GL either immediately or after 1 or 3 days; **B** : Two successive shifts from BL acclimated cultures. Cultures were shifted to GL and once reached  $Exc_{495:545}$  values in GL, cultures were shifted to BL at T0, or after 1 or 3 days. Finally cultures were again shifted to GL once reached  $Exc_{495:545}$  values in BL.

### III. DONNÉES COMPLÉMENTAIRES A CETTE ÉTUDE

#### EXPERIMENTAL PROCEDURES

##### *Lincomycin assay*

To test the possible involvement of protein synthesis in the CA4 process, two exponentially growing cultures of *Synechococcus* sp. BL107 were submitted to a shift from GL to BL, then after 1.5 day, corresponding approximatively to  $T_{50}$ , were split into 4 aliquots. Two of them were kept in BL, but one was supplemented with  $10 \text{ ng.mL}^{-1}$  lincomycin, an antibiotic interacting with ribosome to inhibit protein synthesis. The two other flasks were shifted to GL and darkness, respectively. Samples were collected for excitation and emission spectra analysis at 3 h, 6 h then every 12 h. This dose of lincomycin was determined after testing a range of concentration in *Synechococcus* sp. RS9916 and selecting the minimal dose necessary to fully inhibit the CA4 process while being non lethal in the time frame of the experiment. This is a much lower dose than was used in previous studies on marine *Synechococcus* (Zhao et al., 2000; Blot et al., 2009).

##### *Pigments measurements (performed by Christophe Six)*

Liposoluble-pigment were extracted by methanolic extraction (95% final concentration, Sigma-Aldrich) and analysed using an HP1100 HPLC system (Agilent Technologies, Palo Alto, CA) as previously described (Everroad et al., 2006). Analyses were performed on same samples than qPCR assays, in triplicates. Photosynthetic pigments were identified according to their retention times and absorbance properties.

##### *PCR screening of phycobilin lyase genes potentially involved in CA4*

To complement genomic analyses, the presence of CA4-specific lyase genes either formally characterized (*mpeZ*; Shukla et al., 2012) or identified by genomic analyses (*mpeW*; see results) was checked by PCR in a number of unsequenced *Synechococcus* isolates. Specific primers were designed for *mpeZ* (*mpeZ*\_2F \_ TGGCHGARMGWTTTGTATGT and *mpeZ*\_310R \_ GAYACNGAGGCYGTNATTA) and *mpeW* (*mpeW*\_175F \_ KCYACYGGGGAYAGAGA and *mpeW*\_960R \_ GAACRCCATARTCRGAKCCATG) using all available sequences and the Primalclade webserver (Gadberry et al., 2005), and manually refined. Numbers mentioned in the primer names refers to hybridization position in multiple alignments of individual genes. PCR reactions were performed in 25  $\mu\text{L}$  final volume with mix

consisting of 1.25 U GoTaq polymerase (Promega, Madison, WI), 1 X buffer, 1.5 mM MgCl<sub>2</sub>, 1.25 mM of each dNTP, 2 μM of each primer. PCR reactions were carried out from 1 μL of DNA of each culture. DNA was extracted from 25 mL of these strains using Qiagen Blood and Tissue kit (Qiagen, Courtaboeuf, France), after centrifugation in presence of Pluronic F68 solution (0.005% final concentration, Sigma Aldrich), as previously described (Balzano et al., 2012). The reaction cycle consisted of an initial denaturation step of 5 min at 95°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 50°C and 1 min at 72°C with a final extension step of 10 min. Despite optimization, a fairly low efficiency of these highly degenerated primers was noticed on some strains harboring either *mpeZ* or *mpeW* (see Table 9).

#### *Distribution of mpeZ and mpeW sequences in metagenomes and phylogenetic analyses*

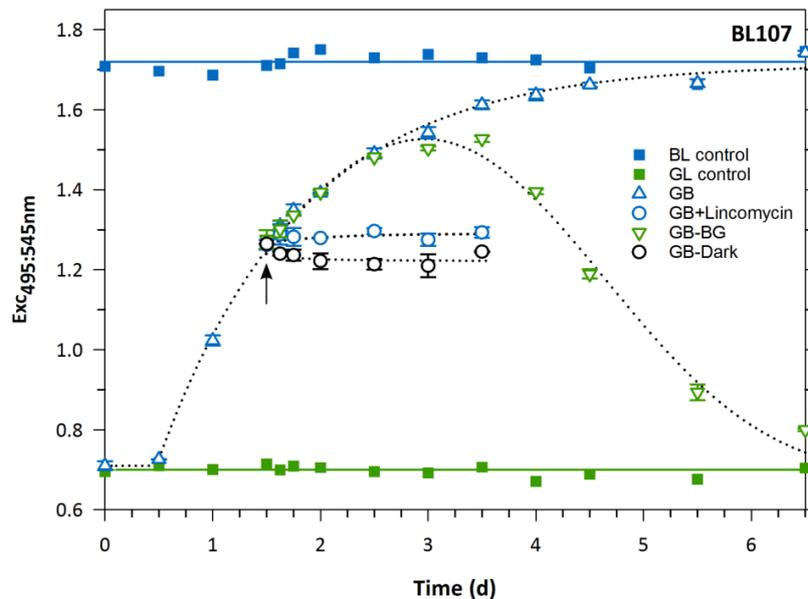
The occurrence of *mpeZ* or *mpeW* sequences was checked in all marine metagenomes available in public databases in October 2012 (Table 10). All datasets were obtained from the CAMERA website (<http://camera.calit2.net/index.shtm>; Sun et al., 2011) and even stations or metagenomes with a low number of reads were used. *mpeZ* sequence from *Synechococcus* sp. RS9916 (IX) and *mpeW* sequence from *Synechococcus* sp. WH8109 (IIa) were used as baits to retrieve significant hits by BLASTP searches (Altschul et al., 1997). Due to strong similarity between *mpeZ*, *mpeW* and *mpeY*, assignation of positive hits was manually checked using the Cyanorak BlastP Tool (<http://www.sb-roscoff.fr/Phyto/cyanorak/>). Metadata were obtained from the CAMERA Portal. Stations with at least one positive sequence were plotted on a world map using Ocean data View.

A phylogenetic tree was constructed by Maximum likelihood (model LG+G), as previously described (see Humily et al. submitted to *Environmental Microbiology*), using all *mpeZ* and *mpeW* sequences from *Synechococcus* isolates. Short metagenomic reads and nucleotidic reference sequences were aligned using MAFFT-GINSI (Katoh et al., 2005) and manual refinement, then added to the phylogenetic tree using the ADD\_BY\_PARSIMONY algorithm implemented in ARB {Ludwig, 2004 #1289}.

## RESULTS AND DISCUSSION

### *Lyncomycin assay*

In order to determine whether the CA4 process requires a continuous supply of photon energy and/or proteins or can run autonomously for some time without such supply, LL-acclimated cultures of *Synechococcus* sp. BL107 were shifted from GL to BL and after 1.5 days (corresponding approximately to the time needed to reach  $T_{50}$ ), they were aliquoted then subjected to different treatments. A first experimental flask was supplemented with a small dose of lincomycin (an inhibitor of the ribosomal function used to block protein synthesis) and kept in the initial light condition, while a second flask was simply placed in the dark (Fig. 55). In both treatments, the  $Exc_{495:545}$  ratio ceased increasing shortly after the treatment then remained constant, unlike the control culture (GL-BL), in which it kept increasing to reach values typically expected in BL for this strain (Fig. 55). A third flask was shifted back to GL after reaching  $T_{50}$  (Fig. 55). In that case, the  $Exc_{495:545}$  ratio continued to increase at the same rate as the control in BL for one day, then remained constant for another day before going down. These results clearly indicate that protein synthesis and photosynthesis are both needed for CA4 to proceed. Furthermore, once the CA4 process is going on, at least one generation time is necessary before a change in light quality starts to affect the pigmentation of *Synechococcus* sp. BL107 cells.



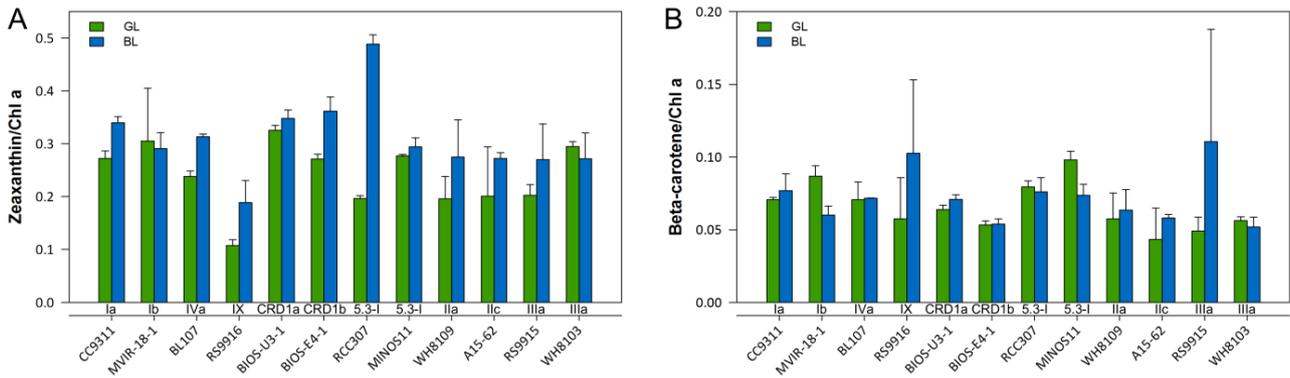
**Figure 55 : Role of protein synthesis in CA4 process.** Kinetics of CA was studied on marine *Synechococcus* sp. BL107 (Phenotypic group 3). Cells were shifted from GL to BL and shifted in a new light condition after 1.5 days (corresponding approximately to  $T_{50}$ ). Cultures were kept in the initial light condition (i.e. BL) and either supplemented with lincomycin (an inhibitor of protein synthesis) or shifted to GL (GB-BG) or darkness (GB-Dark).

Since the treatment occurred in the middle of the CA4 process time course ( $T_{50}$ ), there was likely a stock of phycobilin lyases-isomerases in the cell and they were therefore probably not the limiting factor of the process. These data strongly suggest that CA4 does not consist in a mere replacement within assembled PBS of extant PEI and PEII chromophores by new chromophores better matching the new ambient light quality. Indeed, such a mechanism would likely require minutes to hours to proceed (rather than days as observed our study) and would not be sensitive to lincomycin. Instead, the CA4 process likely requires a *de novo* light-dependent synthesis of PEI and PEII subunits with a new chromophorylation. This process starts in *Synechococcus* cells that initially perceived the ambient light color change signal and continues in their progeny, the complete replacement of the PBS population usually requiring 6 to 7 generations (see Everroad et al., 2006 and this study).

#### *Effect of light quality on liposoluble pigment ratios*

Liposoluble pigment content was analysed on *Synechococcus* strains acclimated for several weeks to 20 photons  $m^{-2} s^{-1}$  of GL or BL, then harvested during the exponential growth. In all strains, 3 major pigments were identified by HPLC : zeaxanthin, Chl  $a$  and  $\beta$ -carotene (Fig. 56; Kana and Glibert, 1987; Kana et al., 1988). Traces of  $\beta$ -cryptoxanthin and Mg-2,4-divinyl pheoporphyrin  $a_5$  (MgDVP), which are intermediates in the zeaxanthin and Chl biosynthesis pathways, respectively, were also detected in some strains. Carotenoids, including xanthophylls, play crucially important physiological and structural roles in cyanobacteria, in which they are involved in both light harvesting and photoprotection. In particular, these pigments help to prevent the production of singlet oxygen by quenching Chl triplet states and dissipating energy as heat (Frank and Cogdell, 1996; Zhu et al., 2010). The role of zeaxanthin in marine *Synechococcus* remains unclear but implication of this xanthophyll in photoprotection has been reported in several studies, mainly in freshwater cyanobacteria (Deeming-Adams and Adams, 1993).

Comparative variations of molar ratios of the main pigments are reported in Figure 56. The amount of zeaxanthin per cell remained constant between the two light qualities, in contrast to Chl  $a$  for which a decrease was observed in BL in the majority of strains, inducing a higher zeaxanthin/Chl  $a$  ratio in BL than GL. Interestingly *Synechococcus* sp. RS9916 and RCC307 showed very large variation of this ratio. Even if it remains to be confirmed, this result may reflect either a light quality-dependent differential synthesis of zeaxanthin under the two light colors or a specific degradation of Chl  $a$  in GL. In contrast no obvious variation was observed for the  $\beta$ -carotene/Chl  $a$  ratio.



**Figure 56 :** Variations of the ratios of the two major carotenoids (zeaxanthin and  $\beta$ -carotene) to Chl *a* for *Synechococcus* strains acclimated to 20 photons  $m^{-2} s^{-1}$  of GL or BL. Error bars indicated standard deviation for triplicates.

*Distribution of mpeZ and mpeW among unsequenced isolates*

Presence of genes coding for MpeZ or MpeW was screened in *Synechococcus* isolates of various lineages for which no genome sequence is available (Table 9). Our primers were not able to amplify *mpeZ* and *mpeW* in some control strains known to possess either gene (Table 9). The *mpeZ* gene was found in several isolates belonging to the clade CRD1 isolated during the Biosope cruise in the South Pacific ocean. All these strains exhibited a high  $Exc_{495:545}$  ratio and are also unable to change their pigmentation according to light quality. Thus, it is likely that, like for BIOS-E4-1, one or the two putative transcriptional regulators FciA/B are lacking in genomes of these isolates (see Humily et al. submitted to *Environmental Microbiology*). With the exception of BIOS-U3-1 isolated from upwelling waters, all CRD1 strains analysed here were isolated in oligotrophic waters (Morel et al., 2007) and it therefore possible that the loss of the *cayAB* operon may constitute an evolutionary advantage for the capture of blue photons in these waters where they penetrate very deeply (Kirk, 1994). CA4 strains harboring *mpeZ* were also identified in all strains belonging to the clade XVI, for which no isolate has been sequenced to date (Ahlgren and Rocard, 2006). To our knowledge all strains belonging to this clade are able to perform CA4.

**Table 9 : Presence of *mpeZ* or *mpeW* genes in *Synechococcus* strains.** For strains for which no genome was available, presence of these genes was screened by PCR.

Strain	RCC # <sup>a</sup>	Sub-Cluster <sup>b</sup>	Clade/subclade <sup>b</sup>	Pigment type <sup>c</sup>	Site of origin	Lat. (N)	Long. (E)	Date	Depth (m)	Genome sequenced	Reference	Physiology		Screening by PCR	
												Phenotypic group <sup>d</sup>	<i>mpeW</i> <sup>e</sup>	<i>mpeZ</i> <sup>e</sup>	
CC9311	1086	5.1B	Ia	3dA	California current	32	-124.52	1-Jan-93	95	Yes	Palenik et al., 2006	2	-	Y	
WH8020	2437	5.1B	Ia	3dA	Sargasso Sea	38.68	-69.34	26-Jun-80	50	Yes	Waterbury et al., 2006	2	-	●	
Biosope_211	2532	5.1B	Ia	3d	Pacific Ocean - Chile upwelling	-33.87	-73.33	8-Dec-04	30	No	Le Gall et al., 2008	-	-	-	
MVIR-18-1	2385	5.1B	Ib	3a	North Sea	61	1.98	23-Jul-07	25	Yes	This study	-	-	●	
WH8016	2535	5.1B	Ib	3a	Atlantic Ocean	41.52	-70.67	1-Jun-80	nd	Yes	Waterbury et al., 2006	-	-	-	
ALMO-3	2432	5.1B	Ib	3a	Mediterranean Sea	36.18	-1.85	8-May-91	nd	No	Fuller et al., 2003	-	-	-	
SYN20	2035	5.1B	Ib	3a	North Sea	60.37	5.41	Nd	0	No	Mazard et al., 2012a	-	-	-	
MICROVIR 1CR_4-1	1649	5.1B	Ib	3d	North Sea - Norway Coast	59.19	4.19	24-Jul-07	50	No	This study	-	-	-	
MICROVIR 22CR_1	1661	5.1B	Ib	3d	North Sea - French Coast	50.12	0.19	8-Jul-07	10	No	This study	-	-	-	
Biosope_141 D	2568	5.1B	Ic	3b	South East Pacific Ocean	-31.87	-91.42	28-Nov-04	40	No	Le Gall et al., 2008	-	-	-	
WH8109	2033	5.1A	IIa	3b	Sargasso Sea	39.29	-70.28	1-Jun-81	nd	Yes	Waterbury et al., 2006	-	Y	-	
A15-62	2374	5.1A	IIc	3dB	Atlantic Ocean	17.62	-20.97	4-Oct-04	15	Yes	Mazard et al., 2012a	1	Y	-	
A15-19	2525	5.1A	IIc	3c	Atlantic Ocean	38.88	-21.35	24-Sep-04	75	No	Mazard et al., 2012a	-	-	-	
PROS-U-1	2369	5.1A	IIh	3dB	Atlantic Ocean Moroccan upwelling	30.13	-10.05	12-Sep-99	5	Yes	Mazard et al., 2012a	1	●	-	
PROSOPE 2-16	2370	5.1A	IIh	3d	Atlantic Ocean Moroccan upwelling	30.13	-10.05	12-Sep-99	5	No	Mazard et al., 2012a	-	-	-	
M11.1	2531	5.1A	IIh	3d	Gulf of Mexico	27.7	-91.32	15-Feb-04	275	No	Mazard et al., 2012a	1	Y	-	
A15-22	1097	5.1A	IIIa	3c	Atlantic Ocean	35.08	-20.85	25-Sep-04	15	No	Mazard et al., 2012a	-	-	-	
A15-24.2	1087	5.1A	IIIa	3c	Atlantic Ocean	35.08	-20.85	25-Sep-04	35	No	Mazard et al., 2012a	-	-	-	
BOUM113-2	2415	5.1A	IIIa	3c	Mediterranean Sea	34.11	-22.9	22-Jun-08	5	No	This study	-	-	-	
RS9915	2553	5.1A	IIIa	3dB	Gulf of Aqaba	29.47	34.94	18-Oct-99	10	Yes	Fuller et al., 2003	1	●	-	
WH8103	2366	5.1A	IIIa	3bB	Atlantic Ocean	28.5	-67.4	17-Mar-81	nd	Yes	Waterbury et al., 2006	-	●	-	
CC9902	2673	5.1A	IVa	3dA	California current	32.9	-117.25	1-Jan-99	5	Yes	Dufresne et al., 2008	2	-	Y	
BL107	515	5.1A	IVa	3dA	Sargasso Sea	41.72	3.57	n.d.	1800	Yes	Dufresne et al., 2008	3	-	Y	
BOUM31-10	2457	5.1A	IVa	3c	Mediterranean Sea	33.38	32.38	27-Jun-08	100	No	This study	-	-	-	
PROSOPE_153-3	325	5.1B	VIa	3c	Mediterranean Sea	43.4	7.82	30-Sep-99	25	No	Mazard et al., 2012a	-	-	-	
PROSOPE_53-19	326	5.1B	VIa	3c	Mediterranean Sea	36.48	13.32	18-Sep-99	5	No	Mazard et al., 2012a	-	-	-	
RS9916	555	5.1B	IX	3dA	Gulf of Aqaba	29.28	34.55	22-Nov-99	10	Yes	Fuller et al., 2003	1	-	Y	
RS9921	559	5.1B	IX	3d	Gulf of Aqaba	29.28	34.55	22-Nov-99	150	No	Fuller et al., 2003	-	-	Y	
PROSOPE_37-2	316	5.1B	XVI	3d	Mediterranean Sea	38	3.83	16-Sep-99	110	No	Mazard et al., 2012a	-	-	Y	
PROSOPE_85-6	2567	5.1B	XVI	3d	Mediterranean Sea	37.4	15.62	26-Sep-99	90	No	Mazard et al., 2012a	-	-	Y	
PROSOPE_41-3	318	5.1B	XVI	3d	Mediterranean Sea	36.48	13.32	18-Sep-99	65	No	Mazard et al., 2012a	-	-	Y	
M21B.3	794	5.1B	XVI	3d	Gulf of Mexico	27.7	-91.3	20-Apr-04	275	No	Everroad and Wood, 2006	-	-	Y	
BIOS-U3-1	2533	5.1B	CRD1a	3dA	Pacific Ocean - Chile upwelling	-33.98	-73.39	6-Dec-04	5	Yes	Le Gall et al., 2008	2	-	Y	
Biosope_112 B6	1023	5.1B	CRD1a	3c	Pacific Ocean	-30.78	-95.43	22-Nov-04	100	No	Le Gall et al., 2008	-	-	Y	
BIOS-E4-1	2534	5.1B	CRD1b	3cA	South East Pacific Ocean	-31.87	-91.44	28-Nov-04	40	Yes	This study	-	-	Y	
Biosope_45 C4Y	1020	5.1B	CRD1b	3c	Pacific Ocean	-9.07	-136.98	2-Nov-04	100	No	Le Gall et al., 2008	-	-	Y	
Biosope_48 A2Y	1030	5.1B	CRD1b	3c	Pacific Ocean	-9.07	-136.98	2-Nov-04	30	No	Le Gall et al., 2008	-	-	Y	
Biosope_48 B6Y	1031	5.1B	CRD1b	3c	Pacific Ocean	-9.07	-136.98	2-Nov-04	30	No	Le Gall et al., 2008	-	-	Y	
RCC307	307	5.3	5.3-I	3eA	Mediterranean Sea	39.1	6.1	26-May-96	15	Yes	Dufresne et al., 2008	4	-	●	
MINOS11	2319	5.3	5.3-I	3dB	Mediterranean Sea	34	18.02	19-Jun-96	20	Yes	Mazard et al., 2012a	1	Y	-	

<sup>a</sup> Strain number in the Roscoff culture collection (<http://www.sb-roscoff.fr/Phyto/RCC>)

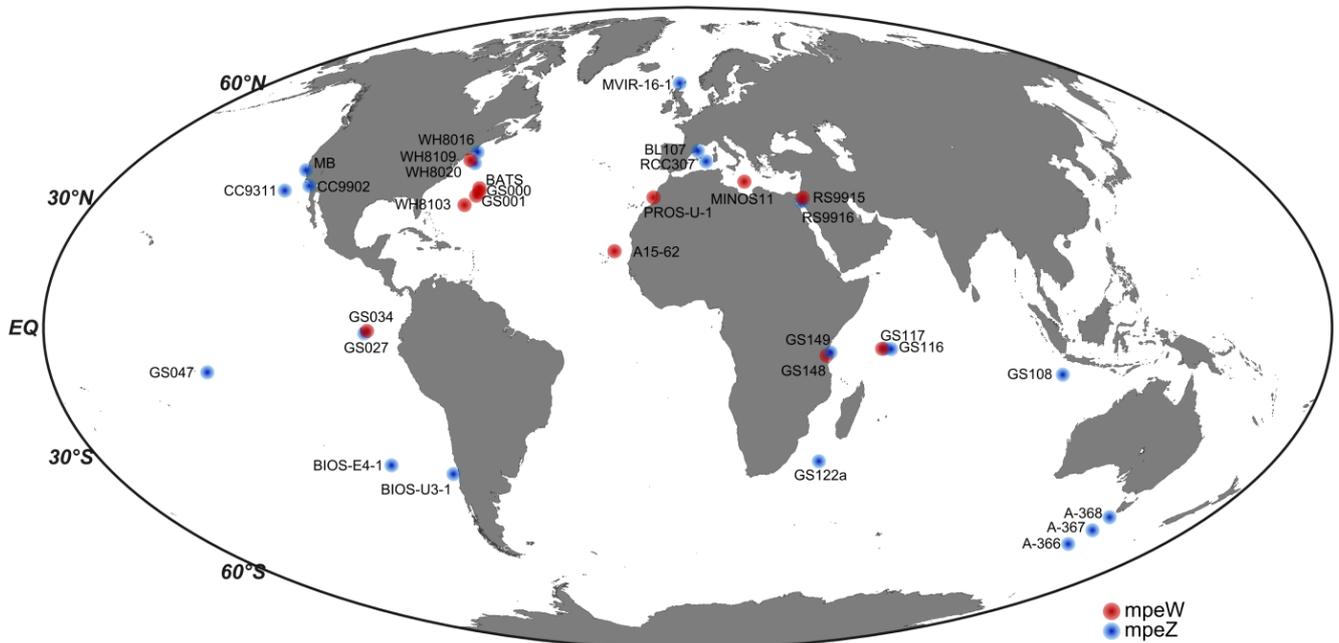
<sup>b</sup> Classification by subcluster, clade or subclade, following the nomenclature reported in previous studies (Fuller et al., 2003; Ahlgren et Rocap, 2006; Dufresne et al., 2008; Huang et al., 2012; Mazard et al., 2012b).

<sup>c,d</sup> Following nomenclature proposed in see Humily et al., submitted to *Environmental Microbiology*.

<sup>e</sup> ● No amplification despite presence of the gene on sequenced genome / Y : positive amplification by PCR

*Diversity and distribution among metagenomes*

The availability of numerous metagenomes in public databases allowed us to search for the occurrence of *mpeZ* and *mpeW* sequences in natural marine samples and to obtain new insights about the distribution of strains able to perform type IV chromatic acclimation (CA4) in the field. Sequences of WH8109 and RS9916 were used to screen metagenomes using BLASTP searches. Relatively few sequences were found in the different metagenomes and the majority was affiliated to MpeZ (Table 10). Despite this apparent low abundance, strains harboring either of these two gene exhibited a worldwide spatial distribution encompassing the Atlantic, Pacific, Indian and Southern oceans, as well as the northern part of the North Sea, the Mediterranean Sea and the Red Sea. Additionally, sequences were retrieved both in nutrient-rich waters (i.e. California coast) and in the ultra-oligotrophic regime of the South-Pacific Ocean. Another striking observation was the co-occurrence of *mpeZ* and *mpeW*-containing strains in the Galapagos Island (Waterbury et al., 1986). These preliminary data indicate a worldwide occurrence of chromatic adapters. They also suggest either a possible differential distribution of CA4-A (containing *mpeZ*) and CA4-B populations (containing *mpeW*), with the latter generally occurring at higher latitudes than the former or possibly a narrower latitudinal distribution of the latter (Fig. 57).

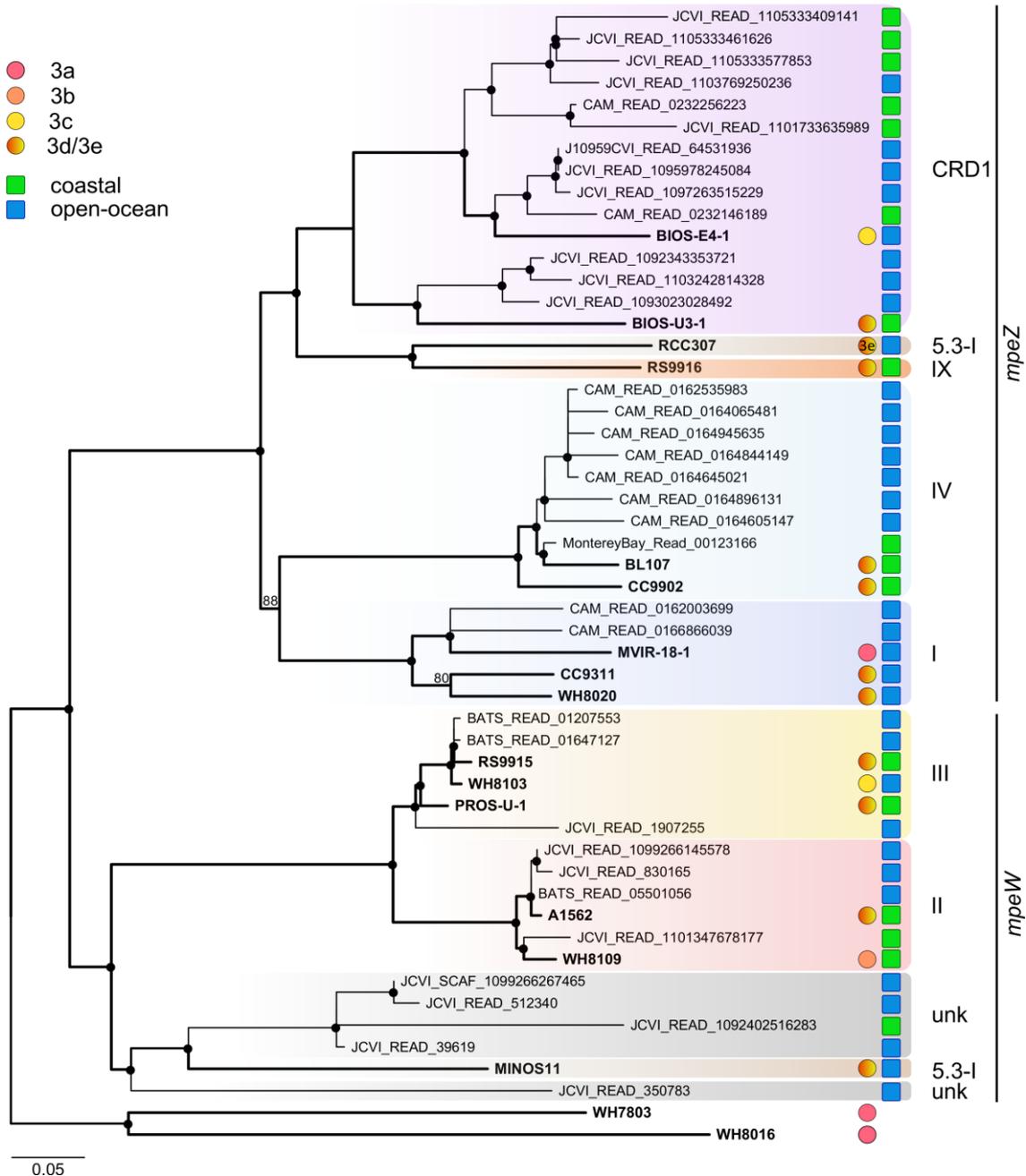


**Figure 57** : World map illustrating position of oceanographic stations with at least one positive sequence affiliated to *mpeZ* (blue circles) or *mpeW* (red circles) and *Synechococcus* strains used in this study. Screening were done in October 2012. Stations with prefix GS correspond to stations from the GOS expedition, while MB and A referred to Monterey Bay and Antarctica respectively (see Table 10).

**Table 10 : List of positive reads affiliated to *mpeZ* or *mpeW* sequences and oceanographic stations. Associated metadata were obtained from the Camera Portal.**

CAMERA Accession number	Sequence affiliation	Metagenomic database	Sample	Site name and description	Region	Country name	Latitude	Longitude	Sampling depth	Water column depth (m)	Sampling date	T°C	Salinity (psu)	Blast-P E <sub>val</sub>	Sequence length (bp)	Adjacent genes (Cyanorak cluster number)
BATS_READ_01647127	<i>mpeW</i>	BATS	BATS-167-0	North Atlantic Spring Bloom	Sargasso Sea	Bermuda	31.645	-64.259	1	>4200	19-Aug-02	27.867	36.319	2.55E-32	243	
BATS_READ_01207553	<i>mpeW</i>	BATS	BATS-167-0	North Atlantic Spring Bloom	Sargasso Sea	Bermuda	31.645	-64.259	1	>4200	19-Aug-02	27.867	36.319	1.09E-26	236	
BATS_READ_05501056	<i>mpeW</i>	BATS	BATS-173-40	North Atlantic Spring Bloom - Oceanic gyre	Sargasso Sea	Bermuda	31.682	-64.173	40	>4200	20-Feb-03	20.24	30.7	1.30E-37	245	
JCVI_SCAF_1099266267465	<i>mpeW</i>	GOS Scaffolds and Unassembled sequences	GS000a	Sargasso Sea - Stations 11 & 13	Sargasso Sea	Bermuda	31.175	-64.324	5	>4200	26-Feb-03	20	36.6	4.79E-97	1383	<i>fciB</i>
JCVI_READ_512340	<i>mpeW</i>	GOS: All Metagenomic Reads	GS000a	Sargasso Sea - Stations 11 & 13	Sargasso Sea	Bermuda	31.175	-64.324	5	>4200	26-Feb-03	20	36.6	2.74E-84	1048	
JCVI_READ_39619	<i>mpeW</i>	GOS: All Metagenomic Reads	GS000a	Sargasso Sea - Stations 11 & 13	Sargasso Sea	Bermuda	31.175	-64.324	5	>4200	26-Feb-03	20	36.6	2.44E-42	1022	<i>fciB</i>
JCVI_READ_350783	<i>mpeW</i>	GOS: All Metagenomic Reads	GS000a	Sargasso Sea - Stations 11 & 13	Sargasso Sea	Bermuda	31.175	-64.324	5	>4200	26-Feb-03	20	36.6	1.28E-30	975	<i>fciB</i>
JCVI_SCAF_1099266145578	<i>mpeW</i>	GOS Scaffolds and Unassembled reads	GS000b	Sargasso Sea - Stations 11 & 13	Sargasso Sea	Bermuda	31.175	-64.324	5	>4200	26-Feb-03	20	36.6	7.59E-146	1719	<i>fciB</i>
JCVI_READ_830165	<i>mpeW</i>	GOS: All Metagenomic Reads	GS000b	Sargasso Sea - Stations 11 & 13	Sargasso Sea	Bermuda	31.175	-64.324	5	>4200	26-Feb-03	20	36.6	1.62E-171	1000	
JCVI_READ_1907255	<i>mpeW</i>	GOS: All Metagenomic Reads	GS001a	Hydrostation S	Sargasso Sea	Bermuda	32.167	-64.5	5	>4200	15-May-03	22.9	36.7	8.87E-29	986	<i>fciB</i>
JCVI_READ_1092343353721	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS027	Devil's Crown, Floreana Island	Galapagos Islands	Ecuador	-1.216	-90.423	2.2	2.3	4-Feb-04	34.9	25.5	2.94E-71	1057	<i>gap-A</i> (CK_8092)
JCVI_READ_1092402516283	<i>mpeW</i>	GOS: All Metagenomic Reads	GS027	Devil's Crown, Floreana Island	Galapagos Islands	Ecuador	-1.216	-90.423	2.2	2.3	4-Feb-04	34.9	25.5	3.11E-42	1039	<i>unk10</i>
JCVI_READ_1093023028492	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS034	North Seamore Island	Galapagos Islands	Ecuador	-0.383	-90.280	2.1	35	19-Feb-04	27.5		1.34E-79	1161	<i>gap-A</i> (CK_8092)
JCVI_SCAF_1097263515229	<i>mpeZ</i>	GOS Scaffolds and Unassembled reads	GS047	201 miles from F. Polynesia	Tropical South Pacific	French polynesia	-10.131	-135.450	30	2400	28-Mar-04	28.6	37.2	1.70E-143	1841	<i>gap-A</i> (CK_8092)
JCVI_READ_1095978245084	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS047	201 miles from F. Polynesia	Tropical South Pacific	French polynesia	-10.131	-135.450	30	2400	28-Mar-04	28.6	37.2	8.50E-54	1057	<i>gap-A</i> (CK_8092)
JCVI_READ_1095964531936	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS047	201 miles from F. Polynesia	Tropical South Pacific	French polynesia	-10.131	-135.450	30	2400	28-Mar-04	28.6	37.2	5.90E-97	1047	
CAM_READ_0232146189	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS108	Coccos Keeling, Inside Lagoon	Indian Ocean	Australia	-12.093	96.882	1.8	7	3-Aug-05	25.8	32.4	8.86E-37	438	
CAM_READ_0232256223	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS108	Coccos Keeling, Inside Lagoon	Indian Ocean	Australia	-12.093	96.882	1.8	7	3-Aug-05	25.8	32.4	2.99E-19	461	<i>gap-A</i> (CK_8092)
JCVI_READ_1103769250236	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS116	Outside Seychelles	Indian Ocean	Seychelles	-4.635	96.836	1.5	2150	17-Aug-05	33.1	26.2	2.45E-72	723	
JCVI_READ_1105333577853	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS117b	St. Anne Island	Indian Ocean	Seychelles	-4.614	55.509	1.8	14	9-Sep-05	26.4	35.5	1.56E-115	1067	
JCVI_READ_1105333461626	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS117b	St. Anne Island	Indian Ocean	Seychelles	-4.614	55.509	1.8	14	9-Sep-05	26.4	35.5	6.47E-87	1055	<i>gap-A</i> (CK_8092)
JCVI_READ_1105333409141	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS117b	St. Anne Island	Indian Ocean	Seychelles	-4.614	55.509	1.8	14	9-Sep-05	26.4	35.5	4.61E-45	1099	<i>unk10</i>
JCVI_READ_1103242814328	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS122a	Between Madagascar and South Africa	Indian Ocean	International waters	-30.898	40.420	1.9	4921	30-Sep-05	20.2	35.8	3.21E-84	1015	<i>gap-A</i> (CK_8092)
JCVI_READ_1101347678177	<i>mpeW</i>	GOS: All Metagenomic Reads	GS148	West coast Zanzibar	Indian Ocean	United republic of tanzania	-6.317	39.010	0.3	1	11-Sep-05	nbsp		1.24E-149	970	
JCVI_READ_1101733635989	<i>mpeZ</i>	GOS: All Metagenomic Reads	GS149	West coast Zanzibar	Indian Ocean	United republic of tanzania	-6.117	39.117	1.5	5	12-Sep-05			9.61E-45	989	<i>gap-A</i> (CK_8092)
MontereyBay_READ_00123166	<i>mpeZ</i>	MontereyBay	298_1	Monterey Bay, California	USA West Coast - Pacific Ocean	United states	36.747	-122.022	0		15-Oct-00			4.90E-27	251	
CAM_READ_0162003699	<i>mpeZ</i>	Antartica	Station 366	Open Ocean transect	Antarctica Ocean		-52.023	144.066	2	3180	24-Jan-08	7.6	33.84	6.57E-30	361	
CAM_READ_0164945635	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	1.75E-47	508	
CAM_READ_0164065481	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	1.38E-41	439	
CAM_READ_0162535983	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	4.00E-41	473	
CAM_READ_0164896131	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	4.57E-41	494	
CAM_READ_0164605147	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	5.14E-39	405	
CAM_READ_0164645021	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	1.16E-36	436	
CAM_READ_0164844149	<i>mpeZ</i>	Antartica	Station 367	Open Ocean transect	Southern Ocean		-48.249	145.805	2	3490	25-Jan-08	11	34.43	6.89E-34	473	
CAM_READ_0166866039	<i>mpeZ</i>	Antartica	Station 368	Open Ocean transect	Southern Ocean		-44.718	145.755	2	3201	26-Jan-08	14.75	34.96	6.77E-42	436	

A phylogenetic tree was constructed using reference sequences from *Synechococcus* isolates and metagenomic reads (Fig. 57). Phylogenetic analyses revealed that most of the metagenomic reads grouped with isolates affiliated to CRD1 and IV clades (37% and 22% of the sequences respectively). Few sequences grouped with the most abundant clades II and III. Interestingly some metagenomic reads grouped profoundly in the tree, and were closely related to MINOS11, whose *mpeW* sequence is more divergent.



**Figure 58** : Phylogenetic tree showing the relationships between *mpeZ* and *mpeW* sequences from *Synechococcus* isolates (in bold) and metagenomic reads retrieved in CAMERA portal. Phylogenetic inference a performed with amino sequences sequences from isolates, using maximum likelihood, and short sequences were added by ARB\_parsimony. Numbers at nodes represent branch confidence values from 1000 bootstraps. Filled circles indicated bootstraps values >90%. The tree is rooted on *mpeY* sequences from WH7803 and WH8016. The scale bar represents 0.05 substitutions per nucleotide.

The distribution pattern follows the distribution of the four most ecologically important clades of *Synechococcus* (I to IV) from which biogeography is well documented. Indeed, sequence affiliated to clade I and IV were retrieved in cold waters of coastal environments at high latitude, while clades II and III are confined to warm waters. To date *mpeZ* was only described in strains of clades I and IV, while the opposite trend was found in clades II and III for *mpeW* (Humily et al., submitted to *Environmental Microbiology*)

It must be noted that some long reads allowed us to examine adjacent genes of *mpeZ* or *mpeW*, but no new gene organization was observed, when comparing these reads to genomes sequenced to date (data not shown). Indeed, in some sequences from stations GS000a, GS000b and GS0001a, the end of *ficiA* gene was found just upstream *mpeW*, while *unk10* was present in another read of station GS027 (JCVI\_READ\_1092402516283). It must also be noted that the beginning of *gapA* (cluster CK\_8092) was found in several reads from the GOS expedition, a gene organization previously described on strains belonging to the CRD1 clade.

Despite the apparent correlation between biogeographic pattern of major clades and distribution of *mpeW* and *mpeZ*, this study is only preliminary and more work is needed to define obvious conclusions about distribution of CA phenomenon. Indeed, in this study all metagenomes have been considered, even those with low numbers of sequence and short reads. With the increasing number of metagenomic data, in particular from the TARA-OCEANS cruise that are currently being analyzed, and technical improvements of high-throughput sequencing methods, a more resolute view of the distribution pattern of CA4 populations should be obtained in future and might also reveal CA regions with an original gene content and organization.

---

CHAPITRE IV

DISCUSSION GÉNÉRALE  
ET  
PERSPECTIVES

---



---

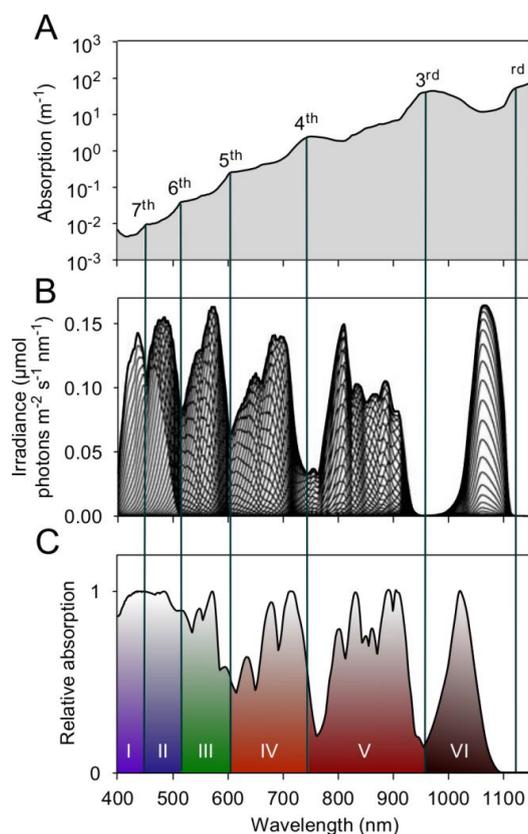
---

## DISCUSSION GÉNÉRALE ET PERSPECTIVES

---

Une théorie de l'écologie classique, dite loi de Gause, prédit que deux espèces ne peuvent pas coexister de manière stable si ces dernières utilisent les mêmes ressources (Gause, 1934), menant à l'exclusion de l'espèce la moins compétitive (Hardin, 1960; MacArthur et Levins, 1967). Cette compétition a conduit les organismes et microorganismes à se spécialiser pour l'utilisation de certaines ressources, un phénomène appelé différenciation de niche (MacArthur et Levins, 1967). Cette théorie est également valable dans le milieu océanique (Hutchinson, 1961). Parmi ces ressources, la qualité et l'intensité de l'énergie lumineuse est particulièrement variable. En plus des phénomènes d'alternance jour-nuit et des changements annuels de photopériode, les conditions hydrologiques du milieu marin (charge en substances organiques, particules, etc.), ainsi que l'abondance et la nature des communautés phytoplanctoniques en présence conditionnent la disponibilité et la qualité de cette ressource. Le spectre lumineux est ainsi divisé en différentes régions, déterminées par les différents harmoniques de vibration des molécules d'eau (Fig. 59; Stomp et al., 2007), accessibles de manière différentielle aux microorganismes en fonction de leur composition pigmentaire.

Les *Synechococcus* marins présentent une très grande diversité pigmentaire, leur permettant de capturer de nombreuses longueurs d'onde du spectre visible (Fig. 30; Six et al., 2007b). Cette diversité est bien plus importante que celle retrouvée chez d'autres genres de cyanobactéries marines comme *Prochlorococcus* (Goericke et Repeta, 1992), *Crocospaera* (Swanson et al., 1991), *Trichodesmium* (Fujita et Shimura, 1974) ou *Acaryochloris* (Marquardt et al., 1997). Les longueurs d'onde de couleur bleu-violette et rouge sont efficacement capturées par la Chl *a* alors que les photons retrouvés entre ces deux régions sont absorbés par les chromophores fixés au niveau des PBS. La singularité de cette cyanobactérie réside dans la complexité et la plasticité de son appareil collecteur de lumière, qui apparaît plus efficace que celui de *Prochlorococcus*, sauf au bas de la couche euphotique où les propriétés d'absorption des DV-Chl *a* et *b* présentes chez ce dernier conviennent parfaitement à la capture des photons bleus parvenant en profondeur (Ting et al., 2002).



**Figure 59** : Partitionnement du spectre solaire dans l'environnement océanique. **(A)** Spectre d'absorption de l'eau pure. Les différents harmoniques de vibrations des molécules d'eau (indiqués par des traits gris) conditionnent les niches lumineuses utilisables par les différents organismes. **(B)** Superposition de 100 spectres lumineux collectés dans différents environnements, des eaux claires de l'océan oligotrophe, aux écosystèmes turbides. **(C)** Superposition de 100 spectres d'absorption collectés sur 20 espèces phototrophes incluant les bactéries pourpres sulfureuses (absorption dans les zones I, II, V), vertes sulfureuses (I, IV), pourpres non sulfureuses (II, V, VI), cyanobactéries (I, II, III, IV) et algues vertes, rouges et diatomées (I, II, III, IV). Modifié d'après Stomp et al. (2007).

*Synechococcus* est facilement cultivable et constitue à l'heure actuelle le genre cyanobactérien pour lequel le plus de génomes ont été séquencés (Dufresne et al., 2008; Shih et al., 2013). Cependant, les souches en culture fournissent souvent une vision restreinte de la diversité réellement présente dans l'environnement (Amann et al., 1995; Rappé et Giovannoni, 2003). D'autre part, la répartition des types pigmentaires *in situ* a essentiellement été étudiée en utilisant la cytométrie en flux ou la spectrofluorimétrie qui ne permettent pas de déterminer avec précision leur abondance relative. Au cours de ce travail de thèse, nous avons développé une approche métagénomique originale, permettant le séquençage de régions PBS directement à partir de l'environnement. Le principal intérêt de cette région est qu'elle nous permet de déduire le type pigmentaire de la cellule qui contient cette région, sans connaissance préalable de sa pigmentation. La taille de la région PBS n'excédant pas 30 Kpb dans l'état actuel des connaissances (Six et al., 2007; Article II), une approche utilisant des bibliothèques de fosmidés a été retenue.

De nombreuses études, notamment en SCG, ont utilisé le couplage entre la cytométrie en flux et la WGA (voir revues de Stepanauskas, 2012; Blainey, 2013). Le couplage entre WGA et fabrication de bibliothèques de fosmidés a également été utilisé pour identifier un cluster de gènes impliqué dans l'oxydation du méthanol (Chen et al., 2008; Neufeld et al., 2008). Notre approche est la première à combiner ces trois technologies, permettant de cibler des régions génomiques d'intérêt. Néanmoins, des biais techniques sont associés à chacune de ces méthodologies sophistiquées (sélectivité du tri, amplification différentielle de taxons, formation de chimères, etc.) et nous avons donc essayé d'en évaluer les effets, afin d'en diminuer au maximum l'incidence, en suivant l'évolution de la diversité génétique et pigmentaire des échantillons par une approche T-RFLP au cours des différentes étapes du protocole. Cette approche nous a permis d'obtenir des régions PBS partielles à partir d'un échantillon naturel prélevé au large de Roscoff, validant ainsi la procédure mise en œuvre. Les régions obtenues présentent un contenu et une organisation génique très proche de celle des souches en culture. Il semble donc que la méthode de sélection des fosmidés à séquencer ait été efficace. L'analyse de ces séquences a par ailleurs montré un très faible taux d'erreurs de séquençage. Ces résultats sont encourageants, mais certaines optimisations restent néanmoins à réaliser concernant les étapes de criblage des fosmidés contenant des régions PBS et d'assemblage des "reads" pour reconstituer les régions génomiques complètes à partir de mélange de fosmidés. Il conviendrait notamment d'effectuer le séquençage avec un taux de multiplexage plus faible que ce qui a été utilisé au cours de notre première étude (mélange de 25 fosmidés d'intérêt), ainsi que la réalisation de banques "mate-pairs" de 3 à 4 Kbp au lieu de banques "paired-end" de 350 pb. Nous ne pouvons cependant pas exclure que la présence de séquences chimériques puisse constituer une limitation pour l'étape d'assemblage. Ces améliorations pourront être appliquées aux autres banques de fosmidés que j'ai fabriqués et criblés au cours de cette thèse, et qui proviennent de différents environnements marins (Méditerranée, Atlantique NE, Mer du Nord). Ces échantillons permettront d'élargir les connaissances sur la diversité pigmentaire. Face aux limitations de la WGA, le développement de méthodologies ne nécessitant pas d'étapes d'amplification apparaît pour le futur comme une alternative pertinente. Ainsi, le développement des NGS de troisième génération devrait permettre d'accéder à la totalité de l'information génétique à partir d'une seule cellule, en évitant tous ces biais (Schadt et al., 2010).

Ces travaux permettront de mieux comprendre la diversité génique de cette région et de dessiner des amorces spécifiques de chacun des types pigmentaires. La quantification de ces derniers pourra alors être réalisée précisément dans le milieu naturel par PCR quantitative. Nous pouvons également espérer mettre en évidence de nouveaux gènes et/ou types pigmentaires au sein de ce genre picocyanobactérien, directement à partir du milieu naturel.

## PHYCOBILINES LYASES ET MÉCANISMES DE L'ACCLIMATATION CHROMATIQUE DE TYPE IV

Une part importante de ma thèse a été consacrée à l'obtention et l'analyse de 25 nouveaux génomes de *Synechococcus* qui ont été séquencés au Genoscope. Si l'analyse comparative des gènes impliqués dans la synthèse des phycobilisomes n'a pas révélé de nouvelles PBPs, elle a permis d'aller plus loin dans la compréhension de l'AC4. Ce processus est trouvé chez le type pigmentaire le plus sophistiqué des *Synechococcus* marins, lui permettant de modifier sa pigmentation en fonction de la couleur de la lumière incidente. Depuis sa découverte en 2001, l'AC4 a été beaucoup moins étudié que les processus d'AC observés chez les cyanobactéries d'eau douce. Une collaboration récente de mon équipe d'accueil avec celle de David Kehoe (Indiana University) a permis l'identification des sites de fixation affectés par ce phénomène chez *Synechococcus* sp. RS9916, dont un se trouve sur la PEI (Cys  $\alpha$ -139) et deux sur la PEII (Cys  $\alpha$ -83 et Cys  $\alpha$ -140; Shukla et al., 2012). Cet article a également mis en évidence le rôle de MpeZ, une PEB lyase-isomérase, qui agit spécifiquement sur la Cys  $\alpha$ -83 de la PEII, et dont le gène est beaucoup plus fortement exprimé en bleu qu'en vert. Cette tendance a été confirmée au cours de cette thèse sur d'autres souches capables d'AC4 et possédant le gène *mpeZ* (Chapitre III).

Les comparaisons génomiques effectuées sur les nouveaux génomes de *Synechococcus*, notamment de souches capables d'AC4, ont également révélé l'absence de *mpeZ* au sein des génomes d'un certain nombre de souches pourtant capables d'ajuster leur composition pigmentaire en fonction de la lumière incidente. Au sein des génomes, une autre phycobiline lyase, que nous avons nommée MpeW, participerait à ce phénomène, mais au contraire de MpeZ, le gène codant pour cette protéine est plus fortement exprimé en vert qu'en bleu. Ce résultat traduit ainsi la complexité du processus d'AC4 que l'on commence tout juste à appréhender. MpeZ et MpeW sont en effet homologues, mais l'expression différentielle des gènes codant ces protéines suggère que ces deux phycobilines lyases ont une fonction différente. Il est en effet possible que MpeW soit une phycobiline lyase dépourvue d'activité isomérase. Cette hypothèse est renforcée par l'absence du cluster AC4-B, qui inclut le gène *mpeW*, chez les souches du type pigmentaire 3c, qui à l'exception de cet îlot génomique, possède une région PBS très similaire. La caractérisation de MpeW paraît donc nécessaire afin de mieux comprendre ce nouveau type d'AC4. D'autres questions restent également en suspens. Il a en effet été démontré que MpeZ n'était pas responsable de la chromophorylation différentielle des résidus Cys-140 de l' $\alpha$ -PEII et Cys-139 de l' $\alpha$ -PEI entre bleu et vert (Shukla et al., 2012). De ce fait, on peut se demander quelle(s) lyase-isomérase(s) est (sont) responsable(s) de la

chromophorylation de ces résidus au sein des souches AC4-A. Parmi les différentes hypothèses avancées, l'équipe de Wendy Schluchter a proposé que MpeZ pourrait également être responsable de la chromophorylation de la Cys- $\alpha$ 140 de la PEII, mais que cette activité pourrait n'être effective qu'après la chromophorylation de la Cys- $\alpha$ 83 de la même sous-unité (Biswas, 2011). L'hypothèse la plus vraisemblable reste cependant l'implication de nouvelles phycobilines lyases non encore caractérisées et qui pourraient être structurellement différentes de celles actuellement décrites dans la littérature (voir l'onglet "fonctions" de Cyanolyase ([cyanolyase.genouest.org/](http://cyanolyase.genouest.org/)) et l'Annexe B). Un certain nombre d'études restent donc nécessaires pour compléter les connaissances actuelles sur la biosynthèse et la régulation des PBS chez les *Synechococcus* marins, parmi lesquelles :

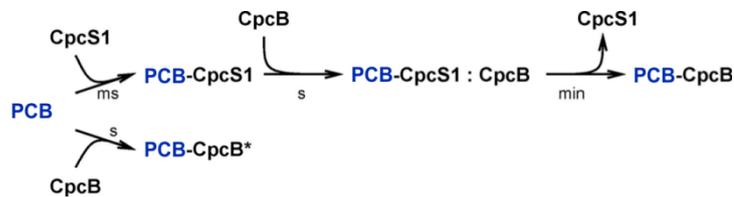
- 1) la caractérisation biochimique des phycobilines-lyases putatives dont les gènes ont été identifiés dans le génome des différentes souches de *Synechococcus* (CpeU, MpeU, MpeW et MpeY) par une approche d'expression hétérologue chez *E. coli* (Blot et al., 2009; Shukla et al., 2012),
- 2) la création de mutants des nombreux gènes inconnus de la région PBS (unk1-13) qui pourraient avoir un rôle dans la pigmentation des souches ou la régulation de l'expression des gènes (Six et al., 2007b) et
- 3) l'utilisation d'une approche de mutagenèse non ciblée (ex: par interposon) pour créer de manière aléatoire des mutants affectés dans leur pigmentation de par l'inactivation de gènes non caractérisés à ce jour.

D'autre part, les études actuelles menées sur la compréhension des mécanismes d'AC4 utilisent systématiquement le proxy  $Exc_{495:545}$  pour évaluer le contenu en PUB des PBS. Or cette relation n'est pas linéaire, comme cela a pu être démontré avec un mutant *mpeZ*, chez qui le changement d'un seul chromophore sur les trois affectés par ce processus, entraîne une baisse substantielle de ce rapport (Shukla et al., 2012). Des analyses structurales de la PEII semblent ainsi nécessaires afin d'appréhender les transferts d'énergie associés et d'établir d'éventuelles corrélations avec le proxy actuellement utilisé.

Au cours de cette thèse nous avons également observé que le temps nécessaire à un changement total de pigmentation pour une souche capable d'AC4 après modification de la lumière ambiante (bleue à verte ou vice versa) se situait entre 4 et 7 jours en conditions de faible intensité lumineuse, confirmant ainsi des études précédentes (Palenik, 2001; Everroad et al., 2006; Shukla et al., 2012). Ces études ont conduit à supposer qu'une action directe des phycobilines lyases sur les PBPs assemblées était à exclure, vraisemblablement en raison de contraintes stériques dues à l'assemblage en trimères ou hexamères. Cette vision est également renforcée par le fait que les cellules ont besoin du même temps d'acclimatation pour modifier leur contenu pigmentaire lors d'un passage de BL à GL ou de GL à BL. En effet, dans l'état actuel des connaissances, il semble que le

changement de conformation chimique de PEB en PUB ne soit pas réversible. La durée nécessaire au changement de pigmentation effectif de BL vers GL devrait donc être plus long si MpeZ (ou MpeW) était capable d'agir directement sur les PBS assemblés. En accord avec cette hypothèse, nous avons pu mettre en évidence que le processus d'AC4 était stoppé par un ajout de lincomycine ou lorsque les cellules étaient placées dans le noir. Il semble donc que le processus d'AC4 implique la synthèse *de novo* d'holoprotéines, leur chromophorylation avec de la PEB, qui serait ensuite isomérisées en PUB. Ces unités remplaceraient progressivement les PBPs existantes et seraient utilisées pour la fabrication de nouveaux PBS.

Le mécanisme d'action des phycobilines lyases reste à ce jour relativement peu étudié. Les études existantes s'accordent néanmoins sur le fait que ces enzymes fonctionneraient tel des chaperonnes (Böhm et al., 2007; Tu et al., 2008; Schluchter et al., 2010), un mécanisme d'action préalablement décrit pour la cytochrome *c* lyase, une autre enzyme liant également des substrats tétrapyrroliques (Turkarlan et al., 2006). Le chromophore est d'abord lié à la phycobiline-lyase, puis transféré au niveau de l'apoprotéine. Ce schéma est valable pour CpcS1 (Fig. 60; Kupka et al., 2009), CpeS1 (Tu et al., 2008) et l'hétéroduplexe PecE/F, bien que ce mécanisme soit beaucoup plus complexe chez ce dernier couple de lyases (Zhao et al., 2005a).



**Figure 60 : Mécanisme d'action de CpcS1.** Le complexe PCB-CpcB\* correspond à un attachement incorrect du chromophore PCB sur CpcB, résultant d'une addition autocatalytique. Modifié d'après Kupka et al. (2009).

Certaines questions restent à élucider sur le mécanisme d'action des phycobilines lyases, notamment en ce qui concerne la spécificité des interactions. Les motifs E/Z-HEAT, caractéristiques du clan E/F, semblent indispensables à la capacité de liaison des chromophores. Cependant, un domaine, est incapable de fixer un chromophore, montrant que la seule présence des domaines E/Z-HEAT n'est pas suffisante pour conférer cette capacité (Tu, 2008).

La transformation de la conformation chimique de la PEB en PUB est assurée par des phycobilines-lyases portant une fonction isomérase. Cependant, on ne sait pas si cette modification chimique est réalisée lors de la liaison de la PEB à la lyase-isomérase, ou après la liaison covalente du chromophore à la PBP. Il est intéressant de noter que le motif d'isomérisation NHCQGN, mis en évidence au sein de PecF ou RpcG (Storf et al., 2001; Zhao et al., 2005a; Blot et al., 2009), n'est présent ni chez MpeZ, ni chez ses homologues MpeW ou MpeY. Les comparaisons de séquences

effectuées au cours de notre étude (Chapitre III) ont permis d'identifier un autre motif très conservé, L-HRDE-(R)QY, au sein de ces trois familles protéiques. Des travaux de mutagenèse dirigée permettraient de valider ou non, l'implication de ce motif dans l'activité lyase ou l'activité isomérase chez les *Synechococcus* marins.

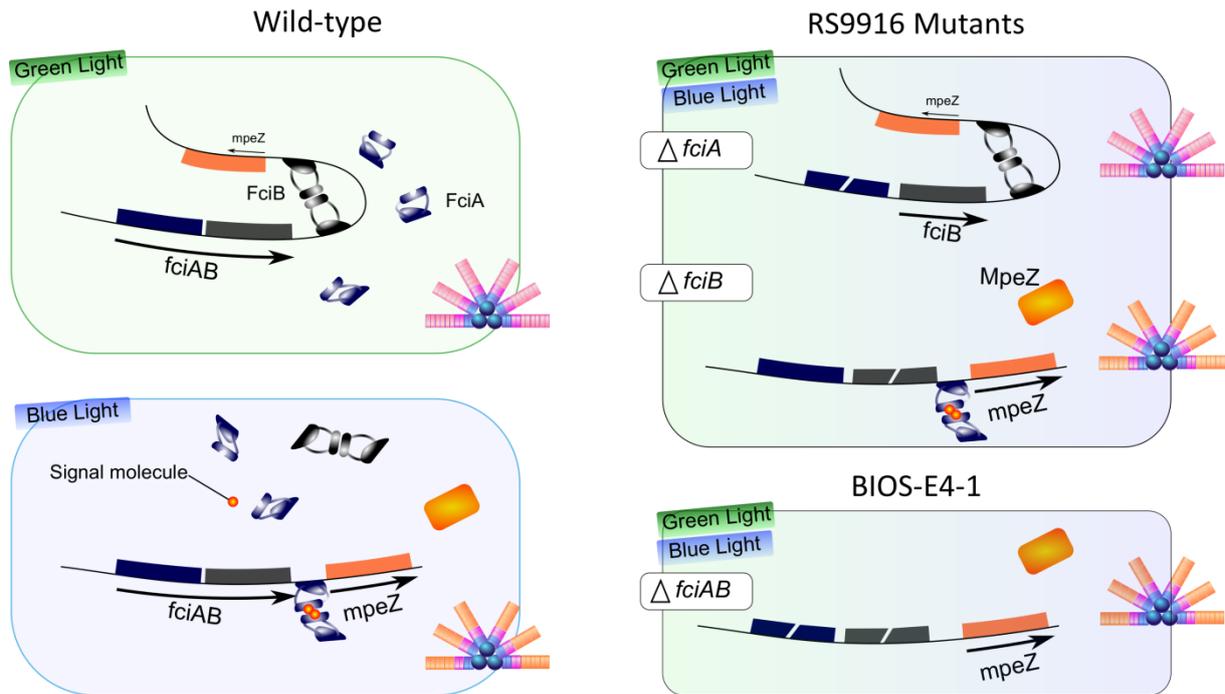
## RÉGULATION DE L'AC4

L'absence de protéines de type cyanobactériochrome au sein des génomes de *Synechococcus*, notamment des souches capables d'AC4, ont depuis longtemps suggéré l'existence d'un mécanisme de régulation original de ce phénomène par rapport à l'AC2 et AC3. Notre étude a permis de révéler la présence de gènes codant pour des protéines comportant des domaines de liaison à l'ADN, de type AraC/XylS, que nous avons nommés FciA et FciB (Chapitre III; Gallegos et al., 1997). Ces protéines seraient impliquées dans la régulation de l'AC4 car elles sont (i) systématiquement présentes chez toutes les souches AC4, (ii) retrouvées à proximité d'une enzyme caractérisée et jouant un rôle clé dans ce processus (MpeZ) et (iii) leur absence chez la souche BIOS-E4-1 entraîne l'incapacité de ces cellules à modifier leur chromophorylation. Le niveau d'expression de ces deux gènes semble également faible, comme pour la majorité de ce type de régulateurs (Gallegos et al., 1997; Pomposiello et Demple, 2001; Schleif, 2010).

Les régulateurs de type AraC/XylS possèdent des mécanismes de régulation distincts, et leurs mécanismes d'action varient largement suivant les processus qu'ils contrôlent. Cependant des caractères similaires peuvent être répertoriés. Les domaines de liaison à l'ADN sont systématiquement présents à l'extrémité C-terminale, alors que les domaines responsables de la capture du signal, de l'homo ou hétérodimérisation, sont quant à eux retrouvés au niveau de l'extrémité C-term (Gallegos et al., 1997; Schleif, 2000; Pomposiello et Demple, 2001; Wickstrum et al., 2007). Il est intéressant de noter que l'analyse structurale de FciA en utilisant Phyre2 (Kelley et Sternberg, 2009) indique que l'extrémité N-terminale présente une homologie structurale faible avec le domaine phosphorécepteur CheY-like de la diguanylate cyclase WspR de *Pseudomonas aeruginosa* (De et al., 2008).

Les différents membres de la famille AraC/XylS possèdent des mécanismes de régulation distincts, et leurs modes d'action varient très largement suivant les processus qu'ils contrôlent. Cependant des caractères similaires ont été répertoriés. Les domaines de liaison à l'ADN sont systématiquement présents à l'extrémité C-terminale, alors que les domaines responsables de la capture du signal, de l'homo ou hétérodimérisation, sont quant à eux retrouvés au niveau de

l'extrémité N-terminale (Gallegos et al., 1997; Schleif, 2000; Pomposiello et Demple, 2001; Wickstrum et al., 2007). Récemment, des mutants de *fciA* et *fciB*, 2 gènes codant pour des membres de la famille AraC/XylS, ont été obtenus chez *Synechococcus* sp. RS9916 par l'équipe de David Kehoe, permettant de proposer des hypothèses quant au rôle de chacune de ces protéines par homologie aux mécanismes décrits dans la littérature pour des régulateurs similaires (Fig. 61; Kehoe D. et Shukla A., comm. pers.). L'inactivation de FciB entraîne un phénotype similaire à celui des cellules placées en BL. Par homologie à la régulation de l'opéron arabinose ou rhamnose, il est donc possible que cette protéine intervienne dans la répression de l'expression du gène de la phycobiline lyase MpeZ en lumière verte (état de base), en formant un homoduplexe (Schleif, 2010; Wickstrum et al., 2010). L'hypothèse de la formation d'un hétéroduplexe régulateur FciA/FciB semble peu vraisemblable car un mutant de FciA, donc ne possédant que FciB, est capable de réprimer efficacement l'expression de MpeZ. Par homologie à AraC, nous proposons qu'un homodimère FciB/FciB réprime efficacement la transcription du gène en mettant en contact deux sites distants au niveau de l'ADN, entraînant la formation d'une boucle (Schleif, 2010). Les hypothèses sur le rôle de FciA sont plus difficiles à établir, mais cette protéine pourrait participer à l'activation de l'expression suite à un signal dont la nature reste à établir. Dans ce contexte, il est intéressant de noter que l'analyse structurale de FciA en utilisant Phyre2 (Kelley et Sternberg, 2009) indique que son extrémité N-terminale présente une homologie structurale faible avec le domaine phosphorécepteur CheY-like de la diguanylate cyclase WspR de *Pseudomonas aeruginosa* (De et al., 2008). Nous proposons donc qu'une molécule signal encore non identifiée induirait un changement conformationnel de FciA, qui serait alors capable de déstabiliser directement l'homodimère FciB/FciB, ou d'entrer en compétition avec ce dernier, provoquant ainsi la déstabilisation de la boucle d'ADN (Fig. 61; Wickstrum et al., 2010). FciA pourrait être actif à l'état monomère ou homodimère. Ces hypothèses restent largement spéculatives et le rôle de ces protéines est actuellement étudié dans l'équipe de David Kehoe. Il s'agira également de comprendre le rôle de FciC, présentant un domaine impliqué dans la répression transcriptionnelle, mais qui n'est pas retrouvé chez les souches possédant un cluster AC4 de type B.



**Figure 61** : Schéma hypothétique de la régulation de l'acclimation chromatique de type IV chez les *Synechococcus* marins en se basant sur le mutant naturel BIOS-E4-1 et mutants de RS9916 étudiés à ce jour. Les phycobilisomes dont les bras apparaissent en orange sont riches en PUB, alors que ceux représentés en rose sont riches en PEB. Les mutants  $\Delta fciA$  et  $\Delta fciB$  de RS9916 sont disponibles dans l'équipe de David Kehoe et actuellement en cours de caractérisation.

## SIGNIFICATION ÉCOLOGIQUE DE L'AC4

Plus d'un quart des souches de *Synechococcus* en culture sont capables de modifier leur composition pigmentaire, ce qui suggère une importance écologique de l'AC4 jusqu'alors insoupçonnée. Bien que l'idée que ce phénomène pourrait procurer un avantage compétitif vis-à-vis d'autres espèces phytoplanctoniques, notamment celles dont la pigmentation est fixe, soit couramment admise par la communauté scientifique, peu de travaux se sont attachés à démontrer cette hypothèse et l'impact de ce processus sur l'écologie et la biogéographie de ces organismes.

Une expérience en chémostat menée sur des souches d'eau douce en utilisant notamment *Tolypothrix*, capable d'AC3, et deux souches de *Synechococcus* possédant respectivement de la PC et de la PE, a permis de démontrer que *Tolypothrix* était ainsi capable de coexister avec celles-ci en lumière blanche, en utilisant la région du spectre lumineux non utilisé par ses compétiteurs (Stomp et al., 2004). L'avantage écologique de l'AC en condition lumineuse de qualité variable (changement entre lumière verte et rouge) a également été confirmé par des expériences en chémostat avec deux souches de pigmentation fixe, une riche en PC, l'autre riche en PE. La souche *Pseudanabaena*, capable de plasticité phénotypique, se révèle capable d'exclusion compétitive dans les deux conditions lumineuses testées (Stomp et al., 2008). De manière intéressante, cette étude reste la

seule à avoir démontré que cette plasticité phénotypique conférait un avantage compétitif d'autant plus efficace que la fréquence des variations lumineuses était faible (Stomp et al., 2008). A l'inverse, ce processus prend du temps et nécessite une grande quantité d'énergie, ce qui pourrait constituer un désavantage lorsque les variations de qualité de lumière sont rapides ou dans des conditions dites de stress, comme le suggèrent certains modèles mathématiques (Gabriel, 2005). A ce jour ce type d'expériences n'a pas été appliqué au phénomène d'AC4. Les expériences de compétition en chémostat se révèlent cependant difficiles à mettre en œuvre d'un point de vue expérimental. Les souches utilisées peuvent cependant présenter des optima différents de croissance au niveau température, besoins nutritionnels ou intensité lumineuse qui pourraient compliquer l'interprétation des résultats (Passarge et al., 2006; Stomp, 2008).

Au cours de cette thèse, une étude très préliminaire sur la distribution des cellules capables d'AC4 a également été menée, en recherchant la présence et la distribution des gènes codant pour les deux phycobilines lyase, MpeZ et MpeW, dans les métagénomés marins actuellement disponibles dans les bases de données publiques. Ces résultats sont toutefois à interpréter avec précautions car (i) certaines souches possédant ces gènes ne sont pas capable d'AC4 comme nous l'avons démontré avec les souches en culture et (ii) le nombre de séquences retrouvées dans les métagénomés utilisés est bien trop faible pour tirer des conclusions statistiquement significatives. Cependant ces résultats suggèrent que les AC4 seraient retrouvés à la fois dans les environnements côtiers et dans l'océan ouvert, où les conditions hydrologiques et lumineuses diffèrent largement. Les cellules capables de plasticité phénotypique seraient ainsi favorisées lors des évènements de mélange des masses d'eau, en étant capables de mieux absorber les longueurs d'onde bleues lorsque les cellules se retrouvent plus en profondeur dans la couche photique où ces longueurs d'onde dominant (Kirk, 1994). Ces phénomènes de transport verticaux chez *Synechococcus* sont une conséquence de processus purement physiques de mélange, mais sont aussi observés à la suite d'adsorption cellulaire à de larges particules (Vilibić et Šantić, 2008; Gao et al., 2011; Sohrin et al., 2011). L'AC4 serait également bénéfique lors de mouvements d'advection, qui mélangent des eaux vertes côtières à des eaux bleues pélagiques. Une autre conséquence du brassage des masses d'eau est l'enrichissement en nutriments, qui peut entraîner des modifications de la structure des communautés phytoplanctoniques, et ainsi modifier le spectre lumineux disponible (Huisman et al., 2004). Les efflorescences de diatomées ou de flagellés pigmentés, par exemple, absorbent fortement les longueurs d'onde dans la partie bleu-violette du spectre lumineux, provoquant ainsi un désavantage pour les cellules riches en PUB (Shikata et al., 2009).

---

RÉFÉRENCES

BIBLIOGRAPHIQUES

---



---

---

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

### A

- Abate, A.R., Hung, T., Mary, P., Agresti, J.J., and Weitz, D.A.** (2010) High-throughput injection with microfluidics using picoinjectors. *Proceedings of the National Academy of Sciences of the United States of America* **107**(45): 19163-19166.
- Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W.Q., Chang, S.H., Chang, H.W., Watson, D., Brodie, E.L., Hazen, T.C., and Keller, M.** (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Applied and Environmental Microbiology* **72**(5): 3291-3301.
- Acinas, S.G., Haverkamp, T.H.A., Huisman, J., and Stal, L.J.** (2008) Phenotypic and genetic diversification of *Pseudanabaena* spp.(cyanobacteria). *The ISME Journal* **3**(1): 31-46.
- Adams, D.G.** (2001) How do cyanobacteria glide? *Microbiology Today* **28**: 131-133.
- Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.-C., Marquez, M., Klibanov, A.M., Griffiths, A.D., and Weitz, D.A.** (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**(9): 4004-4009.
- Ahlgren, N.A., and Rocap, G.** (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Applied and Environmental Microbiology* **72**(11): 7193-7204.
- Ahlgren, N.A., Rocap, G., and Chisholm, S.W.** (2006) Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology* **8**(3): 441-454.
- Ahlgren, N.A., and Rocap, G.** (2012) Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Frontiers in Microbiology* **3**(213): doi: 10.3389/fmicb.2012.00213.
- Ajlani, G., and Vernotte, C.** (1998) Deletion of the PB-loop in the L<sub>CM</sub> subunit does not affect phycobilisome assembly or energy transfer functions in the cyanobacterium *Synechocystis* sp. PCC6714. *European Journal of Biochemistry* **257**(1): 154-159.
- Alberte, R.S., Wood, A.M., Kursar, T.A., and Guillard, R.R.L.** (1984) Novel phycoerythrins in marine *Synechococcus* spp. Characterization and evolutionary and ecological implications. *Plant Physiology* **75**(3): 732-739.
- Alberti, M., Burke, D.H., and Hearst, J.E.** (1995) Structure and sequence of the photosynthesis gene cluster. In *Anoxygenic Photosynthesis Bacteria*. Blankenship, R.E., Madigan, M., and Bauer, C.E. (eds). Dordrecht, The Netherlands: Kluwer, pp. 1083-1106.
- Allgaier, M., Uphoff, H., Felske, A., and Wagner-Döbler, I.** (2003) Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. *Applied and Environmental Microbiology* **69**(9): 5051-5059.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17): 3389-3402.
- Alvey, R.M., Karty, J.A., Roos, E., Reilly, J.P., and Kehoe, D.M.** (2003) Lesions in phycoerythrin chromophore biosynthesis in *Fremyella diplosiphon* reveal coordinated light regulation of apoprotein and pigment biosynthetic enzyme gene expression. *The Plant Cell Online* **15**(10): 2448-2463.
- Alvey, R.M., Bezy, R.P., Frankenberg-Dinkel, N., and Kehoe, D.M.** (2007) A light regulated OmpR-class promoter element co-ordinates light-harvesting protein and chromophore biosynthetic enzyme gene expression. *Molecular Microbiology* **64**(2): 319-332.
- Amann, R.I., Ludwig, W., and Schleifer, K.-H.** (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**(1): 143-169.

- Anderson, L.K., Rayner, M.C., Sweet, R.M., and Eiserling, F.A.** (1983) Regulation of *Nostoc* sp. phycobilisome structure by light and temperature. *Journal of Bacteriology* **155**(3): 1407-1416.
- Anderson, L.K., and Toole, C.M.** (1998) A model for early events in the assembly pathway of cyanobacterial phycobilisomes. *Molecular Microbiology* **30**(3): 467-474.
- Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W., and Bork, P.** (2001) Comparison of ARM and HEAT protein repeats. *Journal of Molecular Biology* **309**(B1): 1-18.
- Apple, J.K., Strom, S.L., Palenik, B., and Brahamsha, B.** (2011) Variability in protist grazing and growth on different marine *Synechococcus* isolates. *Applied and Environmental Microbiology* **77**(9): 3074-3084.
- Apt, K., Hoffman, N., and Grossman, A.** (1993) The gamma subunit of R-phycoerythrin and its possible mode of transport into the plastid of red algae. *Journal of Biological Chemistry* **268**(22): 16208-16215.
- Apt, K.E., Collier, J.L., and Grossman, A.R.** (1995) Evolution of phycobiliproteins. *Journal of Molecular Biology* **248**(1): 79-96.
- Arciero, D., Dallas, J., and Glazer, A.** (1988) *In vitro* attachment of bilins to apophycocyanin. II. Determination of the structures of tryptic bilin peptides derived from the phycocyanobilin adduct. *Journal of Biological Chemistry* **263**(34): 18350-18357.
- Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., and Bechner, M.** (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**(5693): 79-86.
- Arrigo, K.R.** (2004) Marine microorganisms and global nutrient cycles. *Nature* **437**(7057): 349-355.
- Arteni, A.A., Ajlani, G., and Boekema, E.J.** (2009) Structural organisation of phycobilisomes from *Synechocystis* sp. strain PCC6803 and their interaction with the membrane. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1787**(4): 272-279.
- Asada, K.** (1999) The water-water cycle in chloroplasts: scavenging of active oxygens and dissipation of excess photons. *Annual Review of Plant Biology* **50**(1): 601-639.
- Ashkin, A., and Dziedzic, J.** (1987) Optical trapping and manipulation of viruses and bacteria. *Science* **235**(4795): 1517-1520.
- Ashkin, A., Dziedzic, J., and Yamane, T.** (1987) Optical trapping and manipulation of single cells using infrared laser beams. *Nature* **330**(6150): 769-771.
- Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D.** (2011) Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**(7353): 604-608.
- Awramik, S.M.** (1992) The oldest records of photosynthesis. *Photosynthesis Research* **33**(2): 75-89.
- Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., and Thingstad, F.** (1983) The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series* **10**(3): 257-263.
- Azam, F., and Malfatti, F.** (2007) Microbial structuring of marine ecosystems. *Nature Reviews Microbiology* **5**(10): 782-791.

## B

- Balzano, S., Marie, D., Gourvil, P., and Vaultot, D.** (2012) Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *The ISME Journal* **6**(8): 1480-1498.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., and Prjibelski, A.D.** (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5): 455-477.
- Barber, J., Morris, E., and Büchel, C.** (2000) Revealing the structure of the photosystem II chlorophyll binding proteins, CP43 and CP47. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1459**(2): 239-247.
- Barberán, A., Bates, S.T., Casamayor, E.O., and Fierer, N.** (2011) Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal* **6**(2): 343-351.
- Barlow, R., and Alberte, R.** (1985) Photosynthetic characteristics of phycoerythrin-containing marine *Synechococcus* spp. *Marine Biology* **86**(1): 63-74.
- Battistuzzi, F.U., Feijao, A., and Hedges, S.B.** (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evolutionary Biology* **4**(1): 44.
- Baumann, L., Baumann, P., Mandel, M., and Allen, R.D.** (1972) Taxonomy of aerobic marine eubacteria. *Journal of Bacteriology* **110**(1): 402-429.
- Baumdicker, F., Hess, W., and Pfaffelhuber, P.** (2010) The diversity of a distributed genome in bacterial populations. *The Annals of Applied Probability* **20**(5): 1567-1606.
- Baumdicker, F., Hess, W.R., and Pfaffelhuber, P.** (2012) The infinitely many genes model for the distributed genome of bacteria. *Genome biology and evolution* **4**(4): 443-456.

- Beale, S.I. (1993) Biosynthesis of phycobilins. *Chemical Reviews* **93**(2): 785-802.
- Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R.R., and Stoeck, T. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology* **13**(2): 340-349.
- Behrenfeld, M.J., O'Malley, R.T., Siegel, D.A., McClain, C.R., Sarmiento, J.L., Feldman, G.C., Milligan, A.J., Falkowski, P.G., Letelier, R.M., and Boss, E.S. (2006) Climate-driven trends in contemporary ocean productivity. *Nature* **444**(7120): 752-755.
- Béja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., and Spudich, J.L. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**(5486): 1902-1906.
- Béja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., and Jovanovich, S.B. (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology* **2**(5): 516-529.
- Béja, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411**: 786-789.
- Béja, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., Eisen, J.A., Fraser, C.M., and DeLong, E.F. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**(6872): 630-633.
- Bekker, A., Holland, H.D., Wang, P.L., Rumble, D., Stein, H.J., Hannah, J.L., Coetzee, L.L., and Beukes, N.J. (2004) Dating the rise of atmospheric oxygen. *Nature* **427**(6970): 117-120.
- Bendall, D.S., and Manasse, R.S. (1995) Cyclic photophosphorylation and electron transport. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1229**(1): 23-38.
- Bennett, A., and Bogorad, L. (1973) Complementary chromatic adaptation in a filamentous blue-green alga. *The Journal of Cell Biology* **58**(2): 419-435.
- Bernard, C., Etienne, A.L., and Thomas, J.C. (1996) Synthesis and binding of phycoerythrin and its associated linkers to the phycobilisome in *Rhodospira rubra* (Rhodophyta) : compared effects of high-light and translation inhibitors. *Journal of Phycology* **32**(2): 265-271.
- Berns, D.S. (1967) Immunochemistry of biliproteins. *Plant Physiology* **42**(11): 1569-1586.
- Bezy, R.P., and Kehoe, D.M. (2010) Functional characterization of a cyanobacterial OmpR/PhoB class transcription factor binding site controlling light color responses. *Journal of Bacteriology* **192**(22): 5923-5933.
- Bezy, R.P., Wiltbank, L., and Kehoe, D.M. (2011) Light-dependent attenuation of phycoerythrin gene expression reveals convergent evolution of green light sensing in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* **108**(45): 18542-18547.
- Bhattacharya, D., Price, D.C., Bicep, C., Baptiste, E., Sarwade, M., Rajah, V.D., and Yoon, H.S. (2013) Identification of a marine cyanophage in a protist single-cell metagenome assembly. *Journal of Phycology* **49**(1): 207-212.
- Bianchi, C.N. (2007) Biodiversity issues for the forthcoming tropical Mediterranean Sea. *Hydrobiologia* **580**: 7-21.
- Bibby, T.S., Nield, J., and Barber, J. (2001a) Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**(6848): 743-745.
- Bibby, T.S., Nield, J., and Barber, J. (2001b) Three-dimensional model and characterization of the iron stress-induced CP43'-photosystem I supercomplex isolated from the cyanobacterium *Synechocystis* PCC 6803. *Journal of Biological Chemistry* **276**(46): 43246-43252.
- Bibby, T.S., Nield, J., Partensky, F., and Barber, J. (2001c) Oxyphotobacteria - Antenna ring around photosystem I. *Nature* **413**(6856).
- Bibby, T.S., Mary, I., Nield, J., Partensky, F., and Barber, J. (2003a) Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**(6952): 1051-1054.
- Bibby, T.S., Nield, J., Chen, M., Larkum, A.W.D., and Barber, J. (2003b) Structure of a photosystem II supercomplex isolated from *Prochlorion didemni* retaining its chlorophyll *a/b* light-harvesting system. *Proceedings of the National Academy of Sciences of the United States of America* **100**(15): 9050-9054.
- Biddanda, B., and Benner, R. (1997) Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. *Limnology and Oceanography* **42**(3): 506-518.
- Biers, E.J., Sun, S.L., and Howard, E.C. (2009) Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *Applied and Environmental Microbiology* **75**(7): 2221-2229.
- Bishop, J.E., Rapoport, H., Klotz, A.V., Chan, C.F., Glazer, A.N., Fueglistaller, P., and Zuber, H. (1987) Chromopeptides from phycoerythrocyanin. Structure and linkage of the three bilin groups. *Journal of the American Chemical Society* **109**(3): 875-881.

- Biswas, A.** (2011) Identification and characterization of enzymes involved in the biosynthesis of different phycobiliproteins in cyanobacteria. In: University of New Orleans, p. 211.
- Biswas, A., Boutaghou, M.N., Alvey, R.M., Kronfel, C.M., Cole, R.B., Bryant, D.A., and Schluchter, W.M.** (2011) Characterization of the activities of the CpeY, CpeZ, and CpeS bilin lyases in phycoerythrin biosynthesis in *Fremyella diplosiphon* strain UTEX 481. *Journal of Biological Chemistry* **286**(41): 35509-35521.
- Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., and Quake, S.R.** (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* **6**(2): e16626.
- Blainey, P.C., and Quake, S.R.** (2011) Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Research* **39**(4): e1910.1093/nar/gkq1074.
- Blainey, P.C.** (2013) The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews* **37**(3): 407-427.
- Blanco, L., and Salas, M.** (1984) Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication. *Proceedings of the National Academy of Sciences of the United States of America* **81**(17): 5325-5329.
- Blanco, L., Bernad, A., Lázaro, J.M., Martin, G., Garmendia, C., and Salas, M.** (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *Journal of Biological Chemistry* **264**(15): 8935-8940.
- Blank, C.E.** (2004) Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. *Geobiology* **2**(1): 1-20.
- Blank, C.E., and 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): 1-23.
- Bloch, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W., and Stetter, K.O.** (1997) *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**(1): 14-21.
- Blot, N., Wu, X.J., Thomas, J.C., Zhang, J., Garczarek, L., Bohm, S., Tu, J.M., Zhou, M., Ploscher, M., Eichacker, L., Partensky, F., Scheer, H., and Zhao, K.H.** (2009) Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *Journal of Biological Chemistry* **284**(14): 9290-9298.
- Boekema, E., Hifney, A., Yakushevskaya, A., Piotrowski, M., Keegstra, W., Berry, S., Michel, K.P., Pistorius, E., and Kruij, J.** (2001) A giant chlorophyll–protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**(6848): 745-748.
- Boenigk, J., Matz, C., Jürgens, K., and Arndt, H.** (2001) The influence of preculture conditions and food quality on the ingestion and digestion process of three species of heterotrophic nanoflagellates. *Microbial Ecology* **42**(2): 168-176.
- Böhm, S., Endres, S., Scheer, H., and Zhao, K.-H.** (2007) Biliprotein chromophore attachment: Chaperone-like function of the PecE subunit of  $\alpha$ -phycoerythrocyanin lyase. *Journal of Biological Chemistry* **282**(35): 25357-25366.
- Bolhar, R., and Van Kranendonk, M.J.** (2007) A non-marine depositional setting for the northern Fortescue Group, Pilbara Craton, inferred from trace element geochemistry of stromatolitic carbonates. *Precambrian Research* **155**(3-4): 229-250.
- Bond, P.L., Smriga, S.P., and Banfield, J.F.** (2000) Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Applied and Environmental Microbiology* **66**(9): 3842-3849.
- Boresch, K.** (1922) Die komplementäre chromatische Adaptation. *Archiv für Protistenkunde* **44**: 1-70.
- Bouman, H.A., Ulloa, O., Scanlan, D.J., Zwirgmaier, K., Li, W.K.W., Platt, T., Stuart, V., Barlow, R., Leth, O., Clementson, L., Lutz, V., Fukasawa, M., Watanabe, S., and Sathyendranath, S.** (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**(5775): 918-921.
- Brahamsha, B.** (1996) An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine cyanobacterium *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America* **93**(13): 6504-6509.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., and Rohwer, F.** (2002) Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America* **99**(22): 14250-14255.
- Brejč, K., Ficner, R., Huber, R., and Steinbacher, S.** (1995) Isolation, crystallization, crystal structure analysis and refinement of allophycocyanin from the cyanobacterium *Spirulina platensis* at 2.3 Å resolution. *Journal of Molecular Biology* **249**(2): 424-440.
- Breitaudeau, A., Coste, F., Humily, F., Garczarek, L., Le Corguillé, G., Six, C., Ratin, M., Collin, O., Schluchter, W.M., and Partensky, F.** (2012) CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. *Nucleic Acids Research* **41**(Database issue): 396-401.
- Brettel, K.** (1997) Electron transfer and arrangement of the redox cofactors in photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1318**(3): 322-373.

- Brettel, K., and Leibl, W.** (2001) Electron transfer in photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1507**(1): 100-114.
- Bricker, T.M.** (1990) The structure and function of CPa-1 and CPa-2 in photosystem II. *Photosynthesis Research* **24**(1): 1-13.
- Bricker, T.M., and Frankel, L.K.** (2002) The structure and function of CP47 and CP43 in photosystem II. *Photosynthesis Research* **72**(2): 131-146.
- Britschgi, T.B., and Giovannoni, S.J.** (1991) Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. *Applied and Environmental Microbiology* **57**(6): 1707-1713.
- Brown, M.V., and Fuhrman, J.A.** (2005) Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquatic Microbial Ecology* **41**(1): 15-23.
- Brown, M.V., Schwabach, M.S., Hewson, I., and Fuhrman, J.A.** (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environmental Microbiology* **7**(9): 1466-1479.
- Brown, M.V., Philip, G.K., Bunge, J.A., Smith, M.C., Bissett, A., Lauro, F.M., Fuhrman, J.A., and Donachie, S.P.** (2009) Microbial community structure in the North Pacific ocean. *The ISME Journal* **3**(12): 1374-1386.
- Bruyant, F., Babin, M., Genty, B., Prasil, O., Behrenfeld, M.J., Claustre, H., Bricaud, A., Garczarek, L., Holtendorff, J., and Koblizek, M.** (2005) Diel variations in the photosynthetic parameters of *Prochlorococcus* strain PCC 9511: Combined effects of light and cell cycle. *Limnology and Oceanography* **50**(3): 850-863.
- Bryant, D.A., Guglielmi, G., Marsac, N.T., Castets, A.-M., and Cohen-Bazire, G.** (1979) The structure of cyanobacterial phycobilisomes: a model. *Archives of Microbiology* **123**(2): 113-127.
- Bryant, D.A.** (1981) The photoregulated expression of multiple phycocyanin species. *European Journal of Biochemistry* **119**(2): 425-429.
- Bryant, D.A.** (1988) Genetic analysis of phycobilisome biosynthesis, assembly, structure, and function in the cyanobacterium *Synechococcus* sp. PCC 7002. In *Light-energy transduction in photosynthesis : higher plants and bacterial models*. Stevens, S.E.J., and Bryant, D.A. (eds). Rockville: American Society of Plant Physiologists, pp. 62-90.
- Buck, K.R., Chavez, F.P., and Campbell, L.** (1996) Basin-wide distributions of living carbon components and the inverted trophic pyramid of the central gyre of the North Atlantic Ocean, summer 1993. *Aquatic Microbial Ecology* **10**(3): 283-298.
- Buick, R.** (2008) When did oxygenic photosynthesis evolve? *Philosophical Transactions of the Royal Society of London B, Biological Sciences* **363**(1504): 2731-2743.
- Buitenhuis, E., Li, W., Lomas, M., Karl, D., Landry, M., Partensky, F., Karl, D.M., Ulloa, O., Jacquet, S., Lantoine, F., Chavez, F., Macias, D., Gosselin, M., and McManus, G.B.** (2012) Bacterial biomass distribution in the global ocean. *Earth System Science Data Discussions* **5**: 301-315.
- Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, H.-P., Fraser, C.M., Smith, H.O., Woese, C.R., and Venter, J.C.** (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**(5278): 1058-1073.
- Burger-Wiersma, T., Veenhuis, M., Korthals, H.J., Van de Wiel, C., and Mur, L.** (1986) A new prokaryote containing chlorophylls *a* and *b*. *Nature* **320**(6059): 262-264.
- Busch, A., Reijerse, E., Lubitz, W., Hofmann, E., and Frankenberg-Dinkel, N.** (2011) Radical mechanism of cyanophage phycoerythrobilin synthase (PebS). *Biochemical Journal* **433**: 469-476.
- Butcher, R.** (1952) Contributions to our knowledge of the smaller marine algae. *Journal of the Marine Biological Association UK* **31**: 175-191.

## C

**Cadoret, J.C., Demoulière, R., Lavaud, J., van Gorkom, H.J., Houmard, J., and Etienne, A.L.** (2004) Dissipation of excess energy triggered by blue light in cyanobacteria with CP43'(*isiA*). *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1659**(1): 100-104.

**Cai, H.Y., Wang, K., Huang, S.J., Jiao, N.Z., and Chen, F.** (2010) Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Applied and Environmental Microbiology* **76**(9): 2955-2960.

**Calbet, A., Trepas, I., Almeda, R., Salo, V., Saiz, E., Movilla, J.I., Alcaraz, M., Yebra, L., and Simó, R.** (2008) Impact of micro- and nanograzers on phytoplankton assessed by standard and size-fractionated dilution grazing experiments. *Aquatic Microbial Ecology* **50**(2): 145-156.

**Campbell, L., and Vaulot, D.** (1993) Photosynthetic picoplankton community structure in the subtropical

- north Pacific ocean near Hawaii (Station ALOHA). *Deep-Sea Research Part I-Oceanographic Research Papers* **40**(10): 2043-2060.
- Campbell, L., Nola, H.A., and Vaultot, D.** (1994) The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnology and Oceanography* **39**(4): 954-961.
- Campbell, L., Liu, H., Nolla, H.A., and Vaultot, D.** (1997) Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991-1994 ENSO event. *Deep Sea Research Part I: Oceanographic Research Papers* **44**(2): 167-192.
- Campbell, L., Landry, M.R., Constantinou, J., Nolla, H.A., Brown, S.L., Liu, H., and Caron, D.A.** (1998) Response of microbial community structure to environmental forcing in the Arabian Sea. *Deep-Sea Research Part II: Topical Studies in Oceanography* **45**(10-11): 2301-2325.
- Canfield, D.E., Habicht, K.S., and Thamdrup, B.** (2000) The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* **288**(5466): 658-661.
- Capuano, V., Braux, A.S., de Marsac, N.T., and Houmard, J.** (1991) The " anchor polypeptide" of cyanobacterial phycobilisomes. Molecular characterization of the *Synechococcus* sp. PCC 6301 *apcE* gene. *Journal of Biological Chemistry* **266**(11): 7239-7247.
- Capuano, V., Thomas, J., de Marsac, N.T., and Houmard, J.** (1993) An *in vivo* approach to define the role of the L<sub>CM</sub>, the key polypeptide of cyanobacterial phycobilisomes. *Journal of Biological Chemistry* **268**(11): 8277-8283.
- Carr, M.-E., Friedrichs, M.A., Schmeltz, M., Noguchi Aita, M., Antoine, D., Arrigo, K.R., Asanuma, I., Aumont, O., Barber, R., and Behrenfeld, M.** (2006) A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography* **53**(5): 741-770.
- Castenholz, R.W., and Waterbury, J.B.** (1989) Group I. Cyanobacteria. In *Bergey's Manual of Systematic Bacteriology*. JT Staley, M.B., N Pfennig, JG Holt, eds (ed). Baltimore: William & Wilkins, pp. 1710-1727.
- Certes, A.** (1884) On the culture, free from known sources of contamination, from waters and from sediments brought back by the expeditions of the Travailleur and the Talisman; 1882-1883. *Seances Acad Sci* **98**: 690-693.
- Charpy, L., and Blanchot, J.** (1998) Photosynthetic picoplankton in French Polynesian atoll lagoons: estimation of taxa contribution to biomass and production by flow cytometry. *Marine Ecology Progress Series* **162**: 57-70.
- Chen, F., Wang, K., Kan, J.J., Bachoon, D.S., Lu, J.R., Lau, S., and Campbell, L.** (2004) Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcl*) sequences. *Aquatic Microbial Ecology* **36**(2): 153-164.
- Chen, F., Wang, K., Kan, J., Suzuki, M.T., and Wommack, K.E.** (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Applied and Environmental Microbiology* **72**(3): 2239-2243.
- Chen, M., and Blankenship, R.E.** (2011) Expanding the solar spectrum used by photosynthesis. *Trends in Plant Science* **16**(8): 427-431.
- Chen, Y., Dumont, M.G., Neufeld, J.D., Bodrossy, L., Stralis-Pavese, N., McNamara, N.P., Ostle, N., Briones, M.J.I., and Murrell, J.C.** (2008) Revealing the uncultivated majority: combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated *Methylocystis* in acidic peatlands. *Environmental Microbiology* **10**(10): 2609-2622.
- Cheung, M.K., Au, C.H., Chu, K.H., Kwan, H.S., and Wong, C.K.** (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The ISME Journal* **4**(8): 1053-1059.
- Chiang, G.G., Schaefer, M.R., and Grossman, A.R.** (1992) Complementation of a red-light-indifferent cyanobacterial mutant. *Proceedings of the National Academy of Sciences of the United States of America* **89**(20): 9415-9419.
- Chisholm, S.W., Olson, R.J., and Yentsch, C.M.** (1988) Flow cytometry in oceanography : status and prospects. *EOS* **69**: 562-572.
- Chisholm, S.W., Frankel, S.L., Goericke, R., Olson, R.J., Palenik, B., Waterbury, J.B., Westjohnsrud, L., and Zettler, E.R.** (1992) *Prochlorococcus marinus* nov genov sp - an oxyphototrophic marine prokaryote containing divinyl chlorophyll-*a* and chlorophyll-*b*. *Archives of Microbiology* **157**(3): 297-300.
- Chitsaz, H., Yee-Greenbaum, J.L., Tesler, G., Lombardo, M.-J., Dupont, C.L., Badger, J.H., Novotny, M., Rusch, D.B., Fraser, L.J., and Gormley, N.A.** (2011) Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nature Biotechnology* **29**(10): 915-921.
- Cho, J.C., Stapels, M.D., Morris, R.M., Vergin, K.L., Schwalbach, M.S., Givan, S.A., Barofsky, D.F., and Giovannoni, S.J.** (2007) Polyphyletic photosynthetic reaction centre genes in oligotrophic marine Gammaproteobacteria. *Environmental Microbiology* **9**(6): 1456-1463.
- Choi, D.H., and Noh, J.H.** (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiology Ecology* **69**(3): 439-448.

- Choi, D.H., Noh, J.H., Hahm, M.S., and Lee, C.M.** (2011) Picocyanobacterial abundances and diversity in surface water of the northwestern Pacific Ocean. *Ocean Science Journal* **46**(4): 265-271.
- Choi, D.H.** (2012) Picocyanobacterial diversity and distribution during summer in the northern East China Sea. *Ocean and Polar Research* **34**(1): 19-28.
- Christaki, U., Jacquet, S., Dolan, J.R., Vault, D., and Rassoulzadegan, F.** (1999) Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnology and Oceanography* **44**(1): 52-61.
- Christaki, U., Giannakourou, A., Van Wambeke, F., and Grégori, G.** (2001) Nanoflagellate predation on auto- and heterotrophic picoplankton in the oligotrophic Mediterranean Sea. *Journal of Plankton Research* **23**(11): 1297-1310.
- Christaki, U., Courties, C., Karayanni, H., Giannakourou, A., Maravelias, C., Kormas, K.A., and Lebaron, P.** (2002) Dynamic characteristics of *Prochlorococcus* and *Synechococcus* consumption by bacterivorous nanoflagellates. *Microbial Ecology* **43**(3): 341-352.
- Christaki, U., Vázquez-Domínguez, E., Courties, C., and Lebaron, P.** (2005) Grazing impact of different heterotrophic nanoflagellates on eukaryotic (*Ostreococcus tauri*) and prokaryotic picoautotrophs (*Prochlorococcus* and *Synechococcus*). *Environmental Microbiology* **7**(8): 1200-1210.
- Chung, C.-C., Chang, J., Gong, G.-C., Hsu, S.-C., Chiang, K.-P., and Liao, C.-W.** (2011) Effects of Asian dust storms on *Synechococcus* populations in the subtropical Kuroshio current. *Marine Biotechnology* **13**(4): 751-763.
- Claustre, H., Bricaud, A., Babin, M., Bruyant, F., Guillou, L., Le Gall, F., Marie, D., and Partensky, F.** (2002) Diel variations in *Prochlorococcus* optical properties. *Limnology and Oceanography*: 1637-1647.
- Cline, J., Braman, J.C., and Hogrefe, H.H.** (1996) PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research* **24**(18): 3546-3551.
- Coale, K.H., Johnson, K.S., Fitzwater, S.E., Michael Gordon, R., Tanner, S., Chavez, F.P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., Steinberger, P., Nightingale, P., Cooper, D., Cochlan, W.P., Landry, M.R., Constantinou, J., Rollwagen, G., Trasvina, A., and Kudela, R.** (1996) A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**: 495-501.
- Cobley, J.G., Clark, A.C., Weerasurya, S., Quesada, F.A., Xiao, J.Y., Bandrapali, N., D'Silva, I., Thounaojam, M., Oda, J.F., and Sumiyoshi, T.** (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*). *Molecular Microbiology* **44**(6): 1517-1531.
- Cohan, F.M.** (2006) Toward a conceptual and operational union of bacterial systematics, ecology, and evolution. *Proceedings of the Royal Society of London Series B* **361**: 1985-1996.
- Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W.** (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**(5768): 1768-1770.
- Coleman, M.L., and Chisholm, S.W.** (2007) Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends in Microbiology* **15**(9): 398-407.
- Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Lasram, F.B., Aguzzi, J., Ballesteros, E., Bianchi, C.N., Corbera, J., Dailianis, T., Danovaro, R., Estrada, M., Frogli, C., Galil, B.S., Gasol, J.M., Gertwagen, R., Gil, J., Guilhaumon, F., Kesner-Reyes, K., Kitsos, M.S., Koukouras, A., Lampadariou, N., Laxamana, E., de la Cuadra, C., Lotze, H.K., Martin, D., Mouillot, D., Oro, D., Raicevich, S., Rius-Barile, J., Saiz-Salinas, J.I., San Vicente, C., Somot, S., Templado, J., Turon, X., Vafidis, D., Villanueva, R., and Voultsiadou, E.** (2010) The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. *PLoS One* **5**(8): e11842.
- Conley, P.B., Lemaux, P.G., and Grossman, A.R.** (1985) Cyanobacterial light-harvesting complex subunits encoded in two red light-induced transcripts. *Science* **230**(4725): 550-553.
- Conley, P.B., Lemaux, P.G., Lomax, T.L., and Grossman, A.R.** (1986) Genes encoding major light-harvesting polypeptides are clustered on the genome of the cyanobacterium *Fremyella diplosiphon*. *Proceedings of the National Academy of Sciences of the United States of America* **83**(11): 3924-3928.
- Conley, P.B., Lemaux, P.G., and Grossman, A.** (1988) Molecular characterization and evolution of sequences encoding light-harvesting components in the chromatically adapting cyanobacterium *Fremyella diplosiphon*. *Journal of Molecular Biology* **199**(3): 447-465.
- Connon, S.A., and Giovannoni, S.J.** (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied and Environmental Microbiology* **68**(8): 3878-3885.
- Cornejo, J., and Beale, S.I.** (1997) Phycobilin biosynthetic reactions in extracts of cyanobacteria. *Photosynthesis Research* **51**(3): 223-230.
- Cornejo, J., Willows, R.D., and Beale, S.I.** (1998) Phycobilin biosynthesis: cloning and expression of a gene encoding soluble ferredoxin-dependent heme oxygenase from *Synechocystis* sp. PCC 6803. *The Plant Journal* **15**(1): 99-107.
- Cotner, J., Ammerman, J., Peele, E., and Bentzen, E.** (1997) Phosphorus-limited bacterioplankton growth in the Sargasso Sea. *Aquatic Microbial Ecology* **13**(2): 141-149.

- Cracraft, J., and Donoghue, M.J. (2004) *Assembling the tree of life*: Oxford University Press, USA.
- Cuvelier, M.L., Ortiz, A., Kim, E., Moehlig, H., Richardson, D.E., Heidelberg, J.F., Archibald, J.M., and Worden, A.Z. (2008) Widespread distribution of a unique marine protistan lineage. *Environmental Microbiology* **10**(6): 1621-1634.
- Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messié, M., Tringe, S.G., Woyke, T., Welsh, R.M., Ishoey, T., and Lee, J.-H. (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the National Academy of Sciences of the United States of America* **107**(33): 14679-14684.
- 
- D**
- Dammeyer, T., and Frankenberg-Dinkel, N. (2006) Insights into phycoerythrobilin biosynthesis point toward metabolic channeling. *Journal of Biological Chemistry* **281**(37): 27081-27089.
- Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W., and Frankenberg-Dinkel, N. (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Current Biology* **18**(6): 442-448.
- Daniel, R. (2005) The metagenomics of soil. *Nature Reviews Microbiology* **3**(6): 470-478.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**(8): 772-772.
- Darwin, C. (1859) *On the origin of species*. Albemarle Street. London.: John Murray.
- Davis, S.J., Kurepa, J., and Vierstra, R.D. (1999) The *Arabidopsis thaliana* HY1 locus, required for phytochrome-chromophore biosynthesis, encodes a protein related to heme oxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **96**(11): 6541-6546.
- De La Torre, J.R., Christianson, L.M., Béjà, O., Suzuki, M.T., Karl, D.M., Heidelberg, J., and DeLong, E.F. (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proceedings of the National Academy of Sciences of the United States of America* **100**(22): 12830-12835.
- De Las Rivas, J., Balsera, M., and Barber, J. (2004) Evolution of oxygenic photosynthesis: genome-wide analysis of the OEC extrinsic proteins. *Trends in Plant Science* **9**(1): 18-25.
- de Lorimier, R.M., Bryant, D., Porter, R., Liu, W., Jay, E., and Stevens Jr, S. (1984) Genes for the  $\alpha$  and  $\beta$  subunits of phycocyanin. *Proceedings of the National Academy of Sciences of the United States of America* **81**(24): 7946-7950.
- de Lorimier, R.M., Bryant, D.A., and Stevens, S.E. (1990) Genetic analysis of a 9 kDa phycocyanin-associated linker polypeptide. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1019**(1): 29-41.
- de Lorimier, R.M., Smith, R.L., and Stevens, S.E. (1992) Regulation of phycobilisome structure and gene expression by light intensity. *Plant Physiology* **98**(3): 1003-1010.
- de Lorimier, R.M., Wilbanks, S.M., and Glazer, A.N. (1993) Genes of the R-phycocyanin II locus of marine *Synechococcus* spp., and comparison of protein-chromophore interactions in phycocyanins differing in bilin composition. *Plant Molecular Biology* **21**(2): 225-237.
- De, N., Pirruccello, M., Krasteva, P.V., Bae, N., Raghavan, R.V., and Sondermann, H. (2008) Phosphorylation-independent regulation of the diguanylate cyclase WsprR. *PLoS Biology* **6**(3): e67.
- Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research* **11**(6): 1095-1099.
- Dean, F.B., Hosono, S., Fang, L.H., Wu, X.H., Faruqi, A.F., Bray-Ward, P., Sun, Z.Y., Zong, Q.L., Du, Y.F., Du, J., Driscoll, M., Song, W.M., Kingsmore, S.F., Egholm, M., and Lasken, R.S. (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America* **99**(8): 5261-5266.
- Deeming-Adams, B., and Adams, W. (1993) The xanthophylls cycles. In *Antioxydants in higher plants*. Alscher, R., and Hess, J. (eds), pp. 91-110.
- del Giorgio, P.A., Gasol, J.M., Vaque, D., Mura, P., Agusti, S., and Duarte, C.M. (1996) Bacterioplankton community structure: protists control net production and the proportion of active bacteria in a coastal marine community. *Limnology and Oceanography*: 1169-1179.
- Delmont, T.O., Robe, P., Cecillon, S., Clark, I.M., Constancias, F., Simonet, P., Hirsch, P.R., and Vogel, T.M. (2010) Accessing microbial diversity for soil metagenomic studies. *Applied and Environmental Microbiology*.
- Delmont, T.O., Robe, P., Clark, I., Simonet, P., and Vogel, T.M. (2011) Metagenomic comparison of direct and indirect soil DNA extraction approaches. *Journal of Microbiological Methods* **86**(3): 397-400.
- DeLong, E.F. (1992) Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America* **89**(12): 5685-5689.
- DeLong, E.F. (2009) The microbial ocean from genomes to biomes. *Nature* **459**(7244): 200-206.

- DeLong, E.F., and Béjà, O.** (2010) The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biology* **8**(4): e1000359.
- Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H.** (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* **36**(16): e105-e105.
- Donia, M.S., Fricke, W.F., Partensky, F., Cox, J., Elshahawi, S.I., White, J.R., Phillippy, A.M., Schatz, M.C., Piel, J., and Haygood, M.G.** (2011) Complex microbiome underlying secondary and primary metabolism in the tunicate-*Prochloron* symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **108**(51): E1423-E1432.
- Doust, A.B., Marai, C.N.J., Harrop, S.J., Wilk, K.E., Curmi, P.M.G., and Scholes, G.D.** (2004) Developing a structure–function model for the cryptophyte phycoerythrin 545 using ultrahigh resolution crystallography and ultrafast laser spectroscopy. *Journal of Molecular Biology* **344**(1): 135-153.
- Dron, A., Rabouille, S., Claquin, P., Le Roy, B., Talec, A., and Sciandra, A.** (2012) Light–dark (12: 12) cycle of carbon and nitrogen metabolism in *Crocospaera watsonii* WH8501: relation to the cell cycle. *Environmental Microbiology* **14**(4): 967-981.
- Ducret, A., Sidler, W., Frank, G., and Zuber, H.** (1994) The complete amino acid sequence of R-phycoerythrin  $\alpha$  and  $\beta$  subunits from the red alga *Porphyridium cruentum*. *European Journal of Biochemistry* **221**(1): 563-580.
- Ducret, A., Müller, S.A., Goldie, K.N., Hefti, A., Sidler, W.A., Zuber, H., and Engel, A.** (1998) Reconstitution, characterisation and mass analysis of the pentacylindrical allophycoerythrin core complex from the cyanobacterium *Anabaena* sp. PCC 7120. *Journal of Molecular Biology* **278**(2): 369-388.
- Duerring, M., Huber, R., Bode, W., Ruembeli, R., and Zuber, H.** (1990) Refined three-dimensional structure of phycoerythrocyanin from the cyanobacterium *Mastigocladus laminosus* at 2.7 Å. *Journal of Molecular Biology* **211**(3): 633-644.
- Duerring, M., Schmidt, G.B., and Huber, R.** (1991) Isolation, crystallization, crystal structure analysis and refinement of constitutive C-phycoerythrin from the chromatically adapting cyanobacterium *Fremyella diplosiphon* at 1.66 Å resolution. *Journal of Molecular Biology* **217**(3): 577-592.
- Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E.V., and Le Gall, F.** (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences of the United States of America* **100**(17): 10020-10025.
- Dufresne, A., Garczarek, L., and Partensky, F.** (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biology* **6**(2): 10.
- Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A.F., Hess, W.R., and Partensky, F.** (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biology* **9**(5): 16.
- Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.-J., Richter, R.A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J.D., and Haft, D.H.** (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME Journal* **6**(6): 1186-1199.
- DuRand, M.D., Olson, R.J., and Chisholm, S.W.** (2001) Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Research Part II: Topical Studies in Oceanography* **48**(8-9): 1983-2003.
- Durrant, J.R., Klug, D.R., Kwa, S., Van Grondelle, R., Porter, G., and Dekker, J.P.** (1995) A multimer model for P680, the primary electron donor of photosystem II. *Proceedings of the National Academy of Sciences of the United States of America* **92**(11): 4798-4802.
- Duxbury, Z., Schliep, M., Ritchie, R.J., Larkum, A.W.D., and Chen, M.** (2009) Chromatic photoacclimation extends utilisable photosynthetically active radiation in the chlorophyll *d*-containing cyanobacterium, *Acaryochloris marina*. *Photosynthesis Research* **101**(1): 69-75.

## E

- Edwards, M., and Richardson, A.J.** (2004) Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* **430**(7002): 881-884.
- Edwards, R.A., and Rohwer, F.** (2005) Viral metagenomics. *Nature Reviews Microbiology* **3**(6): 504-510.
- Egelhoff, T., and Grossman, A.** (1983) Cytoplasmic and chloroplast synthesis of phycobilisome polypeptides. *Proceedings of the National Academy of Sciences of the United States of America* **80**(11): 3339-3343.
- Ehlers, K., and Oster, G.** (2012) On the mysterious propulsion of *Synechococcus*. *PLoS One* **7**(5): e36081.
- Eiler, A., Heinrich, F., and Bertilsson, S.** (2011) Coherent dynamics and association networks among lake bacterioplankton taxa. *The ISME Journal* **6**(2): 330-342.
- Engelmann, T.W.** (1883a) Farbe und Assimilation. III. Weitere Folgerungen. **41**: 17-29.
- Engelmann, T.W.** (1883b) Farbe und Assimilation. Assimilation findet nur in den farbstoffhaltigen Plasmathielchen statt. II. Näherer Zusammenhang zwischen Lichtabsorption und Assimilation. **41**: 1-13.

- Eppley, R.W., and Sloan, P.R.** (1966) Growth rates of marine phytoplankton: correlation with light absorption by cell chlorophyll *a*. *Physiologia Plantarum* **19**(1): 47-59.
- Eppley, R.W.** (1972) Temperature and phytoplankton growth in the sea. *Fishery Bulletin* **70**(4): 1063-1085.
- Esenbeck, N.** (1836) Ueber einen blau-rothen Farbstoff, der sich bei der Zersetzung von Oscillatorien bildet. *Justus Liebigs Annalen der Chemie* **VXII**: 75-82.
- Everroad, C.R., Six, C., Partensky, F., Thomas, J.C., Holtzendorff, J., and Wood, A.M.** (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp. *Journal of Bacteriology* **188**(9): 3345-3356.
- Everroad, C.R., and Wood, M.A.** (2006) Comparative molecular evolution of newly discovered picocyanobacterial strains reveals a phylogenetically informative variable region a  $\beta$ -phycoerythrin. *Journal of Phycology* **42**(6): 1300-1311.
- Everroad, C.R., and Wood, M.A.** (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Molecular Phylogenetics and Evolution* **64**(3): 381-392.
- 
- F**
- Fairchild, C.D., Zhao, J., Zhou, J., Colson, S.E., Bryant, D.A., and Glazer, A.N.** (1992) Phycocyanin  $\alpha$ - subunit phycocyanobilin lyase. *Proceedings of the National Academy of Sciences of the United States of America* **89**(15): 7017-7021.
- Fairchild, C.D., and Glazer, A.N.** (1994) Nonenzymatic bilin addition to the  $\alpha$ - subunit of an apophytoerythrin. *Journal of Biological Chemistry* **269**(46): 28988-28996.
- Falk, H.** (1989) *The chemistry of linear oligopyrroles and bile pigments*. Verlag Wien - New-York: Springer.
- Falkowski, P.** (1994) The role of phytoplankton photosynthesis in global biogeochemical cycles. *Photosynthesis Research* **39**(3): 235-258.
- Falkowski, P.G.** (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature* **387**(6630): 272-275.
- Falkowski, P.G.** (2000) Rationalizing elemental ratios in unicellular algae. *Journal of Phycology* **36**(1): 3-6.
- Farquhar, J., Bao, H., and Thieme, M.** (2000) Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**(5480): 756-759.
- Farquhar, J., Peters, M., Johnston, D.T., Strauss, H., Masterson, A., Wiechert, U., and Kaufman, A.J.** (2007) Isotopic evidence for Mesoarchaeoan anoxia and changing atmospheric sulphur chemistry. *Nature* **449**(7163): 706-709.
- Federspiel, N.A., and Grossman, A.** (1990) Characterization of the light-regulated operon encoding the phycoerythrin-associated linker proteins from the cyanobacterium *Fremyella diplosiphon*. *Journal of Bacteriology* **172**(7): 4072-4081.
- Federspiel, N.A., and Scott, L.** (1992) Characterization of a light-regulated gene encoding a new phycoerythrin-associated linker protein from the cyanobacterium *Fremyella diplosiphon*. *Journal of Bacteriology* **174**(18): 5994-5998.
- Feng, X., Oropeza, R., and Kenney, L.J.** (2003) Dual regulation by phospho-OmpR of *ssrA/B* gene expression in *Salmonella* pathogenicity island 2. *Molecular Microbiology* **48**(4): 1131-1143.
- Ferguson, R., and Rublee, P.** (1976) Contribution of bacteria to standing crop of coastal plankton. *Limnology and Oceanography* **21**(1): 141-145.
- Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., and Iwata, S.** (2004) Architecture of the photosynthetic oxygen-evolving center. *Science* **303**(5665): 1831-1838.
- Ferris, M.J., and Palenik, B.** (1998) Niche adaptation in ocean cyanobacteria. *Nature* **396**(6708): 226-228.
- Ficner, R., Lobeck, K., Schmidt, G., and Huber, R.** (1992) Isolation, crystallization, crystal structure analysis and refinement of B-phycoerythrin from the red alga *Porphyridium sordidum* at 2.2 Å resolution. *Journal of Molecular Biology* **228**(3): 935-950.
- Ficner, R., and Huber, R.** (1993) Refined crystal structure of phycoerythrin from *Porphyridium cruentum* at 0.23-nm resolution and localization of the  $\gamma$  subunit. *European Journal of Biochemistry* **218**(1): 103-106.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P.** (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**(5374): 237.
- Finkel, Z.V., Sebbo, J., Feist-Burkhardt, S., Irwin, A., Katz, M., Schofield, O., Young, J., and Falkowski, P.** (2007) A universal driver of macroevolutionary change in the size of marine phytoplankton over the Cenozoic. *Proceedings of the National Academy of Sciences of the United States of America* **104**(51): 20416-20420.
- Finkel, Z.V., Beardall, J., Flynn, K.J., Quigg, A., Rees, T.A.V., and Raven, J.A.** (2010) Phytoplankton in a changing world: cell size and elemental stoichiometry. *Journal of Plankton Research* **32**(1): 119-137.
- Fisher, R.G., Woods, N.E., Fuchs, H.E., and Sweet, R.** (1980) Three-dimensional structures of C-phycoerythrin and B-phycoerythrin at 5-Å resolution. *Journal of Biological Chemistry* **255**(11): 5082-5089.
- Fitch, W.M., and Margoliash, E.** (1967) Construction of phylogenetic trees. *Science* **155**(760): 279-284.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F.,**

- Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* *Science* **269**(5223): 496-512.
- Fouilland, E., Descolas-Gros, C., Courties, C., and Pons, V. (1999) Autotrophic carbon assimilation and biomass from size-fractionated phytoplankton in the surface waters across the subtropical frontal zone (Indian Ocean). *Polar Biology* **21**(2): 90-96.
- Fox, G., Stackebrandt, E., Hespell, R., Gibson, J., Maniloff, J., Dyer, T., Wolfe, R., Balch, W., Tanner, R., and Magrum, L. (1980) The phylogeny of prokaryotes. *Science* **209**(4455): 457-463.
- Frank, H.A., and Cogdell, R.J. (1996) Carotenoids in photosynthesis. *Photochemistry and Photobiology* **63**(3): 257-264.
- Frankenberg, N., Mukougawa, K., Kohchi, T., and Lagarias, J.C. (2001) Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *The Plant Cell Online* **13**(4): 965-978.
- Frankenberg, N., and Lagarias, J.C. (2003) Phycocyanobilin: ferredoxin oxidoreductase of *Anabaena* sp. PCC 7120. *Journal of Biological Chemistry* **278**(11): 9219-9226.
- Frankenberg, N., and Lagarias, I.C. (2006) Biosynthesis and biological function of bilins. In *The Porphyrin Handbook, Volumes 11-20: Chlorophylls and Bilins: Biosynthesis, Synthesis and Degradation*. Smith, K.M., and Guillard, R. (eds). Amsterdam: Academic Press, pp. 211-236.
- Frankland, P., and Frankland, G.C.T. (1894) *Microorganisms in water: their significance, identification and removal, together with an account of the bacteriological methods employed in their investigation, specially designed for the use of those connected with the sanitary aspects of water-supply*: Longmans, Green.
- Fraser, C., Hanage, W.P., and Spratt, B.G. (2007) Recombination and the nature of bacterial speciation. *Science* **315**(5811): 476-480.
- Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and DeLong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America* **105**(10): 3805-3810.
- Frigaard, N.-U., Martinez, A., Mincer, T.J., and DeLong, E.F. (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439**(7078): 847-850.
- Fu, F.X., Warner, M.E., Zhang, Y., Feng, Y., and Hutchins, D.A. (2007) Effects of increased temperature and CO<sub>2</sub> on photosynthesis, growth, and elemental ratios in marine *Synechococcus* and *Prochlorococcus* (Cyanobacteria). *Journal of Phycology* **43**(3): 485-496.
- Fuchs, B.M., Spring, S., Teeling, H., Quast, C., Wulf, J., Schattener, M., Yan, S., Ferrera, S., Johnson, J., and Glöckner, F.O. (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**(8): 2891-2896.
- Fuhrman, J., McCallum, K., and Davis, A. (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. *Applied and Environmental Microbiology* **59**(5): 1294-1302.
- Fuhrman, J.A. (1992) Novel major archaeobacterial group from marine plankton. *Nature* **356**(6365): 148-149.
- Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* **399**(6736): 541-548.
- Fuhrman, J.A., and Schwalbach, M. (2003) Viral influence on aquatic bacterial communities. *The Biological Bulletin* **204**(2): 192-195.
- Fujita, Y., and Shimura, S. (1974) Phycoerythrin of the marine blue-green alga *Trichodesmium thiebautii*. *Plant and Cell Physiology* **15**(5): 939-942.
- Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Applied and Environmental Microbiology* **69**(5): 2430-2443.
- Fuller, N.J., Tarran, G.A., Yallop, M., Orcutt, K.M., and Scanlan, D.J. (2006) Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnology and Oceanography* **51**(6): 2515-2526.

## G

- Gabriel, W. (2005) How stress selects for reversible phenotypic plasticity. *Journal of Evolutionary Biology* **18**(4): 873-883.
- Gadberry, M.D., Malcomber, S.T., Doust, A.N., and Kellogg, E.A. (2005) Primaclade - a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* **21**(7): 1263-1264.
- Gaidukov, N. (1902) Über den Einfluss farbigen Lichts auf die Färbung lebender Oscillarien. *Königl Akademie der Wissenschaften* **V**: 1-36.

- Gaidukov, N.** (1903) Die Farbveränderung bei den prozessen der komplementären chromatischen adaptation. *Berichte der Deutschen Botanischen Gesellschaft* **21**: 517-522.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C.** (2009a) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences of the United States of America* **106**(52): 22427-22432.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., Potvin, M., and Lovejoy, C.** (2009b) Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *The ISME Journal* **3**(7): 860-869.
- Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J.L.** (1997) Arac/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews* **61**(4): 393-410.
- Gantt, E., and Conti, S.** (1966a) Granules associated with the chloroplast lamellae of *Porphyridium cruentum*. *The Journal of Cell Biology* **29**(3): 423-434.
- Gantt, E., and Conti, S.** (1966b) Phycobiliprotein localization in algae. *Brookhaven Symposia in Biology* **19**: 393-405.
- Gantt, E.** (1980) Structure and function of phycobilisomes: light harvesting pigment complexes in red and blue-green algae. *International Review of Cytology* **66**(4.5): 45-80.
- Gao, W., Shi, X., Wu, J., Jin, Y., Zhang, W., and Meldrum, D.R.** (2011) Phylogenetic and gene expression analysis of cyanobacteria and diatoms in the twilight waters of the temperate northeast pacific ocean. *Microbial Ecology* **62**(4): 765-775.
- Gao, X., Wei, T.-D., Zhang, N., Xie, B.-B., Su, H.-N., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Wang, Z.-X., Wu, J.-W., and Zhang, Y.-Z.** (2012) Molecular insights into the terminal energy acceptor in cyanobacterial phycobilisome. *Molecular Microbiology* **85**(5): 907-915.
- Garcia-Pichel, F., Belnap, J., Neuer, S., and Schanz, F.** (2003) Estimates of global cyanobacterial biomass and its distribution. *Algalogical Studies* **109**: 213-228.
- Garcia, S.L., McMahon, K.D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., Woyke, T., and Warnecke, F.** (2013) Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME Journal* **7**(1): 137-147.
- Garczarek, L., Hess, W.R., Holtzendorff, J., van der Staay, G.W.M., and Partensky, F.** (2000) Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote. *Proceedings of the National Academy of Sciences of the United States of America* **97**(8): 4098-4101.
- Garczarek, L., Partensky, F., Irlbacher, H., Holtzendorff, J., Babin, M., Mary, I., Thomas, J.C., and Hess, W.R.** (2001) Differential expression of antenna and core genes in *Prochlorococcus* PCC 9511 (Oxyphotobacteria) grown under a modulated light-dark cycle. *Environmental Microbiology* **3**(3): 168-175.
- Garczarek, L., Dufresne, A., Rousvoal, S., West, N.J., Mazard, S., Marie, D., Claustre, H., Raimbault, P., Post, A.F., Scanlan, D.J., and Partensky, F.** (2007) High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiology Ecology* **60**(2): 189-206.
- Garczarek, L., Dufresne, A., Blot, N., Cockshutt, A.M., Peyrat, A., Campbell, D.A., Joubin, L., and Six, C.** (2008) Function and evolution of the *psbA* gene family in marine *Synechococcus*: *Synechococcus* sp. WH7803 as a case study. *The ISME Journal* **2**(9): 937-953.
- Garg, H.P., and Prakash, J.** (2000) *Solar energy: fundamentals and applications*. New Delhi: Tata McGraw-Hill Education.
- Gause, G.F.** (1934) *The struggle for existence*. Baltimore, M.D.: Williams and Wilkins.
- Geider, R.J., and Roche, J.L.** (2002) Redfield revisited: variability of C : N : P in marine microalgae and its biochemical basis. *European Journal of Phycology* **37**(01): 1-17.
- Georg, J., Voß, B., Scholz, I., Mitschke, J., Wilde, A., and Hess, W.R.** (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Molecular Systems Biology* **5**(1): doi: 10.1038/msb.2009.1063.
- Ghai, R., Pašić, L., Fernández, A.B., Martin-Cuadrado, A.-B., Mizuno, C.M., McMahon, K.D., Papke, R.T., Stepanauskas, R., Rodriguez-Brito, B., and Rohwer, F.** (2011) New abundant microbial groups in aquatic hypersaline environments. *Scientific reports* **1**(135): doi: 10.1038/srep00135.
- Ghanotakis, D.F., and Yocum, C.F.** (1990) Photosystem II and the oxygen-evolving complex. *Annual Review of Plant Biology* **41**(1): 255-276.
- Ghosh, K., and Schnitzer, M.** (1980) Macromolecular structures of humic substances. *Soil Science* **129**(5): 266-276.
- Gilbert, J.** (2010) Aquatic metagenome library (archive; expression) generation and analysis. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, K. (ed). Berlin: Springer, pp. 4347-4352.
- Gilbert, J.A., and Dupont, C.L.** (2011) Microbial metagenomics: beyond the genome. *Annual Review of Marine Science* **3**: 347-371.
- Gillbro, T., Sharkov, A.V., Kryukov, I.V., Khoroshilov, E.V., Kruykov, P.G., Fischer, R., and Scheer, H.** (1993) Förster energy transfer between neighbouring chromophores in C-phycocyanin trimers. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1140**: 321-326.

- Gindt, Y.M., Zhou, J., Bryant, D.A., and Sauer, K.** (1994) Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores: Assignment of excited state energies to the  $L_{cm}$ ,  $\beta^{18}$  and  $\alpha^{AP-B}$  chromophores. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1186**(3): 153-162.
- Giovannoni, S., and Stingl, U.** (2007) The importance of culturing bacterioplankton in the 'omics' age. *Nature Reviews Microbiology* **5**(10): 820-826.
- Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G.** (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**(6270): 60-63.
- Giovannoni, S.J., Bibbs, L., Cho, J.-C., Stapels, M.D., Desiderio, R., Vergin, K.L., Rappé, M.S., Laney, S., Wilhelm, L.J., and Tripp, H.J.** (2005a) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**(7064): 82-85.
- Giovannoni, S.J., and Stingl, U.** (2005) Molecular diversity and ecology of microbial plankton. *Nature* **437**(7057): 343-348.
- Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., and Noordewier, M.** (2005b) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**(5738): 1242-1245.
- Glauser, M., Bryant, D.A., Frank, G., Wehrli, E., Rusconi, S.S., Sidler, W., and Zuber, H.** (1992a) Phycobilisome structure in the cyanobacteria *Mastigocladus laminosus* and *Anabaena* sp. PCC 7120. *European Journal of Biochemistry* **205**(3): 907-915.
- Glauser, M., Sidler, W.A., Graham, K.W., Bryant, D.A., Frank, G., Wehrli, E., and Zuber, H.** (1992b) Three C-phycoerythrin-associated linker polypeptides in the phycobilisome of green-light-grown *Calothrix* sp. PCC 7601 (cyanobacteria). *FEBS Letters* **297**(1): 19-23.
- Glauser, M., Stirewalt, V.L., Bryant, D.A., Sidler, W., and Zuber, H.** (1992c) Structure of the genes encoding the rod-core linker polypeptides of *Mastigocladus laminosus* phycobilisomes and functional aspects of the phycobiliprotein/linker-polypeptide interactions. *European Journal of Biochemistry* **205**(3): 927-937.
- Glazer, A., and Hixson, C.S.** (1977) Subunit structure and chromophore composition of rhodophyten phycoerythrins. *Porphyridium cruentum* B-phycoerythrin and b-phycoerythrin. *Journal of Biological Chemistry* **252**(1): 32-42.
- Glazer, A.** (1982) Phycobilisomes: structure and dynamics. *Annual Reviews in Microbiology* **36**(1): 173-198.
- Glazer, A.** (1985) Light harvesting by phycobilisomes. *Annual Review of Biophysics and Biophysical Chemistry* **1**(1): 47-77.
- Glazer, A.** (1989a) Light guides. Directional energy transfer in a photosynthetic antenna. *Journal of Biological Chemistry* **264**(1): 1-4.
- Glazer, A.N., West, J.A., and Chan, C.** (1982) Phycoerythrins as chemotaxonomic markers in red algae: a survey. *Biochemical Systematics and Ecology* **10**(3): 203-215.
- Glazer, A.N.** (1989b) Directional energy transfer in a photosynthetic antenna. *Journal of Biological Chemistry* **264**(1): 1-4.
- Glazer, A.N.** (1994) Adaptive variations in phycobilisome structure. *Advances in Molecular and Cell Biology* **10**: 119-149.
- Glenn, T.C.** (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**(5): 759-769.
- Glöckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Beck, A., Borzym, K., and Heitmann, K.** (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proceedings of the National Academy of Sciences of the United States of America* **100**(14): 8298-8303.
- Goberville, E., Beaugrand, G., Sautour, B., Tréguer, P., and Team, S.** (2010) Climate-driven changes in coastal marine systems of western Europe. *Marine Ecology Progress Series* **408**: 129-148.
- Goericke, R., and Repeta, D.J.** (1992) The pigments of *Prochlorococcus* marinus: The presence of divinyl chlorophyll *a* and *b* in a marine prokaryote. *Limnology and Oceanography* **37**: 425-433.
- Golecki, J.R.** (1977) Studies on ultrastructure and composition of cell walls of the cyanobacterium *Anacystis nidulans*. *Archives of Microbiology* **114**(1): 35-41.
- Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L., González, J.M., and Pinhassi, J.** (2010) Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biology* **8**(4): e1000358.
- Gonzalez, J.M., Sherr, E.B., and Sherr, B.F.** (1990) Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. *Applied and Environmental Microbiology* **56**(3): 583-589.
- Gonzalez, J.M., Sherr, E., and Sherr, B.F.** (1993) Differential feeding by marine flagellates on growing versus starving, and on motile versus nonmotile, bacterial prey. *Marine Ecology Progress Series* **102**: 257-267.
- Gorbon, H.R., and Morel, A.** (1983) *Remote assessment of ocean color for interpretation of satellite visible imagery: A review.* New-York (USA): Springer-Verlag.
- Gorsky, G., Chretiennot-Dinet, M.J., Blanchot, J., and Palazzoli, I.** (1999) Picoplankton and nanoplankton aggregation by appendicularians: Fecal pellet contents of *Megalocercus huxleyi* in the equatorial Pacific. *Journal of Geophysical Research: Oceans* **104**(C2): 3381-3390.

- Gradinger, R., and Lenz, J.** (1989) Picocyanobacteria in the high Arctic. *Marine Ecology Progress Series* **52**(1): 99-101.
- Gradinger, R., and Lenz, J.** (1995) Seasonal occurrence of picocyanobacteria in the Greenland Sea and central Arctic Ocean. *Polar Biology* **15**(6): 447-452.
- Green, B., and Durnford, D.** (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annual Review of Plant Biology* **47**(1): 685-714.
- Grindberg, R.V., Ishoey, T., Brinza, D., Esquenazi, E., Coates, R.C., Liu, W.-t., Gerwick, L., Dorrestein, P.C., Pevzner, P., and Lasken, R.** (2011) Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One* **6**(4): e18565.
- Grossman, A.R., Lemaux, P.G., Conley, P.B., Bruns, B.U., and Anderson, L.K.** (1988) Characterization of phycobiliprotein and linker polypeptide genes in *Fremyella diplosiphon* and their regulated expression during complementary chromatic adaptation. *Photosynthesis Research* **17**(1): 23-56.
- Grossman, A.R., Schaefer, M.R., Chiang, G.G., and Collier, J.L.** (1993) The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiological Reviews* **57**(3): 725-749.
- Grossman, A.R.** (2003) A molecular understanding of complementary chromatic adaptation. *Photosynthesis Research* **76**(1): 207-215.
- Guard-Friar, D., Eisenberg, B.L., Edwards, M.R., and MacColl, R.** (1986) Immunochemistry on cryptomonad biliproteins. *Plant Physiology* **80**(1): 38-42.
- Guillou, L., Jacquet, S., Chrétiennot-Dinet, M.-J., and Vaulot, D.** (2001) Grazing impact of two small heterotrophic flagellates on *Prochlorococcus* and *Synechococcus*. *Aquatic Microbial Ecology* **26**(2): 201-207.
- Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V., and Gascuel, O.** (2009) PhyML: Fast and Accurate Phylogeny Reconstruction by Maximum Likelihood. *Infection Genetics and Evolution* **9**(3): 384-385.
- Guiry, M.D.** (2012) *Synechococcales*. In *AlgaeBase World-wide electronic publication World Register of Marine Species* at <http://www.marinespecies.org/aphi.php?p=taxdetails&id=345514>. Guiry, M.D., and Guiry, G.M. (eds): National University of Ireland, Galway.
- Guo, C., Jing, H., Kong, L., and Liu, H.** (2013) Effect of East Asian aerosol enrichment on microbial community composition in the South China Sea. *Journal of Plankton Research*: doi: 10.1093/plankt/fbt1002.
- Gupta, R.S., and Mathews, D.W.** (2011) Signature proteins for the major clades of Cyanobacteria. *BMC Evolutionary Biology* **10**(24): doi:10.1186/1471-2148-1110-1124.
- Gutu, A., and Kehoe, D.M.** (2012) Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. *Molecular Plant* **5**(1): 1-13.

---

## H

- Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., and Sodergren, E.** (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* **21**(3): 494-504.
- Haeckel, E.** (1866) *Generelle Morphologie der Organismen : Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch vue von Charles Darwin reformirte Descendenz-Theorie*. Berlin: Georg Reimer.
- Hagemann, M., Richter, S., and Mikkat, S.** (1997) The *ggtA* gene encodes a subunit of the transport system for the osmoprotective compound glucosylglycerol in *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* **179**(3): 714-720.
- Hagiwara, Y., Sugishima, M., Takahashi, Y., and Fukuyama, K.** (2006) Crystal structure of phycocyanobilin: ferredoxin oxidoreductase in complex with biliverdin IX $\alpha$ , a key enzyme in the biosynthesis of phycocyanobilin. *Proceedings of the National Academy of Sciences of the United States of America* **103**(1): 27-32.
- Hallam, S.J., Girguis, P.R., Preston, C.M., Richardson, P.M., and DeLong, E.F.** (2003) Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Applied and Environmental Microbiology* **69**(9): 5483-5491.
- Hallegraeff, G.M.** (2010) Ocean climate change, phytoplankton community responses, and harmful algal blooms : a formidable predictive challenge. *Journal of Phycology* **46**(2): 220-235.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M.** (1998) Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry & Biology* **5**(10): R245-R249.
- Handelsman, J.** (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68**(4): 669-685.
- Hardin, G.** (1960) The competitive exclusion principle. *Science* **131**(3409): 1292-1297.
- Harrington, E.D., Singh, A.H., Doerks, T., Letunic, I., von Mering, C., Jensen, L.J., Raes, J., and Bork, P.** (2007) Quantitative assessment of protein function prediction from metagenomics shotgun sequences. *Proceedings of the National Academy of Sciences of the United States of America* **104**(35): 13913-13918.

- Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., and Stal, L.J. (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environmental Microbiology* **10**(1): 174-188.
- Hedges, J.I., Baldock, J.A., Gelinas, Y., Lee, C., Peterson, M., and Wakeham, S.G. (2001) Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* **409**(6822): 801-804.
- Herdman, M., Castenholz, R.W., Waterbury, J.B., and Rippka, R. (2001) *Form-genus XII. Synechococcus*. In DR Boone, RW Castenholz, eds, *Bergey's Manual of Systematic Bacteriology. Ed 2d, Ed Vol 1*. New-York: Springer-Verlag.
- Hess, W.R., Partensky, F., Van Der Staay, G., Garcia-Fernandez, J.M., Börner, T., and Vaultot, D. (1996) Coexistence of phycoerythrin and a chlorophyll *a/b* antenna in a marine prokaryote. *Proceedings of the National Academy of Sciences of the United States of America* **93**(20): 11126-11130.
- Heywood, J.L., Zubkov, M.V., Tarran, G.A., Fuchs, B.M., and Holligan, P.M. (2006) Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: Evaluation of inter-annual variability. *Deep-Sea Research Part II-Topical Studies in Oceanography* **53**(14-16): 1530-1547.
- Hirose, M., Katano, T., and Nakano, S.-I. (2008a) Growth and grazing mortality rates of *Prochlorococcus*, *Synechococcus* and eukaryotic picophytoplankton in a bay of the Uwa Sea, Japan. *Journal of Plankton Research* **30**(3): 241-250.
- Hirose, Y., Shimada, T., Narikawa, R., Katayama, M., and Ikeuchi, M. (2008b) Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein. *Proceedings of the National Academy of Sciences of the United States of America* **105**(28): 9528-9533.
- Hirose, Y., Narikawa, R., Katayama, M., and Ikeuchi, M. (2010) Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in *Nostoc punctiforme*, a group II chromatic adapter. *Proceedings of the National Academy of Sciences of the United States of America* **107**(19): 8854-8859.
- Hobbie, J.E., Daley, R.J., and Jasper, S. (1977) Use of nuclepore filters for counting bacteria by fluorescence microscopy. *Applied and Environmental Microbiology* **33**(5): 1225-1228.
- Hoff, K. (2009) The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics* **10**(1): 520.
- Hohmann-Marriott, M.F., and Blankenship, R.E. (2011) Evolution of photosynthesis. *Annual Review of Plant Biology* **62**: 515-548.
- Hoiczuk, E., and Hansel, A. (2000) Cyanobacterial cell walls: news from an unusual prokaryotic envelope. *Journal of Bacteriology* **182**(5): 1191-1199.
- Holland, H.D. (2006) The oxygenation of the atmosphere and oceans. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* **361**(1470): 903-915.
- Holtzendorff, J., Partensky, F., Jacquet, S., Bruyant, F., Marie, D., Garczarek, L., Mary, I., Vaultot, D., and Hess, W.R. (2001) Diel expression of cell cycle-related genes in synchronized cultures of *Prochlorococcus* sp. strain PCC 9511. *Journal of Bacteriology* **183**(3): 915-920.
- Holtzendorff, J., Partensky, F., Mella, D., Lennon, J.-F., Hess, W.R., and Garczarek, L. (2008) Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511. *Journal of Biological Rhythms* **23**(3): 187-199.
- Holzwarth, A.R. (1986) Fluorescence lifetimes in photosynthetic systems. *Photochemistry and Photobiology* **43**(6): 707-725.
- Honda, D., Yokota, A., and Sugiyama, J. (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *Journal of Molecular Evolution* **48**(6): 723-739.
- Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Taylor, T.D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M., and Ohkuma, M. (2008) Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proceedings of the National Academy of Sciences of the United States of America* **105**(14): 5555-5560.
- Hosono, S., Faruqi, A.F., Dean, F.B., Du, Y., Sun, Z., Wu, X., Du, J., Kingsmore, S.F., Egholm, M., and Lasken, R.S. (2003) Unbiased whole-genome amplification directly from clinical samples. *Genome Research* **13**(5): 954-964.
- Houmard, J., Mazel, D., Moguet, C., Bryant, D., and Tandeau de Marsac, N. (1986) Organization and nucleotide sequence of genes encoding core components of the phycobilisomes from *Synechococcus* 6301. *Molecular and General Genetics* **205**(3): 404-410.
- Houmard, J., Capuano, V., Coursin, T., and de Marsac, N.T. (1988) Genes encoding core components of the phycobilisome in the cyanobacterium *Calothrix* sp. strain PCC 7601: occurrence of a multigene family. *Journal of Bacteriology* **170**(12): 5512-5521.
- Huang, S., Wilhelm, S.W., Rodger Harvey, H., Taylor, K., Jiao, N., and Chen, F. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *The ISME Journal* **6**(2): 285-297.
- Huber, J.A., Welch, D.B.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007)

Microbial population structures in the deep marine biosphere. *Science* **318**(5847): 97-100.

**Huelsenbeck, J.P., and Ronquist, F.** (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8): 754-755.

**Hugenholtz, P., Goebel, B.M., and Pace, N.R.** (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology* **180**(18): 4765-4774.

**Huisman, J., Sharples, J., Stroom, J.M., Visser, P.M., Kardinaal, W.E.A., Verspagen, J.M.H., and Sommeijer, B.** (2004) Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology* **85**(11): 2960-2970.

**Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M.** (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**(7): R143.

**Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A., and Sogin, M.L.** (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics* **4**(11): e1000255.

**Huse, S.M., Welch, D.M., Morrison, H.G., and Sogin, M.L.** (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* **12**(7): 1889-1898.

**Hutchins, D., Fu, F.X., Zhang, Y., Warner, M., Feng, Y., Portune, K., Bernhardt, P., and Mulholland, M.** (2007) CO<sub>2</sub> control of *Trichodesmium* N<sub>2</sub> fixation, photosynthesis, growth rates, and elemental ratios: implications for past, present, and future ocean biogeochemistry. *Limnology and Oceanography* **52**(4): 1293-1304.

**Hutchinson, G.E.** (1961) The paradox of the plankton. *The American Naturalist* **95**(882): 137-145.

---

## I

**Iancu, C.V., Ding, H.J., Morris, D.M., Dias, D.P., Gonzales, A.D., Martino, A., and Jensen, G.J.** (2007) The structure of isolated *Synechococcus* strain WH8102 carboxysomes as revealed by electron cryotomography. *Journal of Molecular Biology* **372**(3): 764-773.

**Ikeuchi, M., and Ishizuka, T.** (2008) Cyanobacteriochromes: a new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. *Photochemical and Photobiological Sciences* **7**(10): 1159-1167.

**Irwin, A.J., Finkel, Z.V., Schofield, O.M.E., and Falkowski, P.G.** (2006) Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *Journal of Plankton Research* **28**(5): 459-471.

**Irwin, A.J., and Oliver, M.J.** (2009) Are ocean deserts getting larger? *Geophysical Research Letters* **36**(18): L18609.

**Ishoev, T., Kvist, T., Westermann, P., and Ahring, B.K.** (2006) An improved method for single cell isolation of prokaryotes from meso-, thermo- and hyperthermophilic environments using micromanipulation. *Applied Microbiology and Biotechnology* **69**(5): 510-514.

**Ito, H., and Tanaka, A.** (2011) Evolution of a divinyl chlorophyll-based photosystem in *Prochlorococcus*. *Proceedings of the National Academy of Sciences of the United States of America* **108**(44): 18014-18019.

**Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V.** (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**(6068): 587-590.

---

## J

**Jacquet, S.p., Partensky, F.d.r., Marie, D., Casotti, R., and Vaulot, D.** (2001) Cell cycle regulation by light in *Prochlorococcus* strains. *Applied and Environmental Microbiology* **67**(2): 782-790.

**Jameson, E., Joint, I., Mann, N.H., and Mühling, M.** (2010) Detailed analysis of the microdiversity of *Prochlorococcus* populations along a North-South Atlantic Ocean transect. *Environmental Microbiology* **12**(1): 156-171.

**Jeffrey, S.W.** (1976) A report of green algal pigments in the Central North Pacific Ocean. *Marine Biology* **37**(1): 33-37.

**Jenkins, B.D., Zehr, J.P., Gibson, A., and Campbell, L.** (2006) Cyanobacterial assimilatory nitrate reductase gene diversity in coastal and oligotrophic marine environments. *Environmental Microbiology* **8**(12): 2083-2095.

**Jerlov, N.G.** (1976) *Marine Optics*. Amsterdam: Elsevier Oceanography Series.

**Jezberová, J., and Komárková, J.** (2007) Morphological transformation in a freshwater *Cyanobium* sp. induced by grazers. *Environmental Microbiology* **9**(7): 1858-1862.

**Jiang, T., Zhang, J., and Liang, D.** (1999) Structure and function of chromophores in R-phycoerythrin at 1.9 Å resolution. *Proteins: Structure, Function, and Bioinformatics* **34**(2): 224-231.

**Jiang, T., Zhang, J., Chang, W., and Liang, D.** (2001) Crystal structure of R-phycoerythrin and possible energy transfer pathways in the phycobilisome. *Biophysical Journal* **81**(2): 1171-1179.

## K

- Jiao, N., Herndl, G.J., Hansell, D.A., Benner, R., Kattner, G., Wilhelm, S.W., Kirchman, D.L., Weinbauer, M.G., Luo, T., Chen, F., and Azam, F. (2010) Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews Microbiology* **8**(8): 593-599.
- Jing, H., Liu, H., and Suzuki, K. (2009a) Phylogenetic diversity of marine *Synechococcus* spp. in the Sea of Okhotsk. *Aquatic Microbial Ecology* **56**(1): 55-63.
- Jing, H., and Liu, H. (2011) Phylogenetic composition of *Prochlorococcus* and *Synechococcus* in cold eddies of the South China Sea. *Aquatic Microbial Ecology* **65**(3): 207.
- Jing, H.M., Zhang, R., Pointing, S.B., Liu, H.B., and Qian, P.Y. (2009b) Genetic diversity and temporal variation of the marine *Synechococcus* community in the subtropical coastal waters of Hong Kong. *Canadian Journal of Microbiology* **55**(3): 311-318.
- Johnson, P.W., and Sieburth, J.M. (1979) Chroococcoid cyanobacteria in the sea - ubiquitous and diverse phototropic biomass. *Limnology and Oceanography* **24**(5): 928-935.
- Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**(5768): 1737-1740.
- Joliot, P., and Joliot, A. (2002) Cyclic electron transfer in plant leaf. *Proceedings of the National Academy of Sciences of the United States of America* **99**(15): 10209-10214.
- Jung, L.J., Chan, C.F., and Glazer, A.N. (1995) Candidate genes for the phycoerythrocyanin subunit lyase. *Journal of Biological Chemistry* **270**(21): 12877-12884.
- Jürgens, K., Pernthaler, J., Schalla, S., and Amann, R. (1999) Morphological and compositional changes in a planktonic bacterial community in response to enhanced protozoan grazing. *Applied and Environmental Microbiology* **65**(3): 1241-1250.
- Jürgens, K., and Matz, C. (2002) Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie van Leeuwenhoek* **81**(1): 413-434.
- Jürgens, U.J., Drews, G., and Weckesser, J. (1983) Primary structure of the peptidoglycan from the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6714. *Journal of Bacteriology* **154**(1): 471-478.
- Jürgens, U.J., and Weckesser, J. (1986) Polysaccharide covalently linked to the peptidoglycan of the cyanobacterium *Synechocystis* sp. strain PCC6714. *Journal of Bacteriology* **168**(2): 568-573.
- Jürgens, U.J., and Burger-Wiersma, T. (1989) Peptidoglycan-polysaccharide complex in the cell wall of the filamentous prochlorophyte *Prochlorothrix hollandica*. *Journal of Bacteriology* **171B**(1): 498-502.
- Kahn, K., Mazel, D., Houmard, J., Tandeau de Marsac, N., and Schaefer, M.R. (1997) A role for *cpeYZ* in cyanobacterial phycoerythrin biosynthesis. *Journal of Bacteriology* **179**(4): 998-1006.
- Kana, T.M., and Glibert, P.M. (1987) Effect of irradiances up to 2000  $\mu\text{E m}^{-2}\cdot\text{s}^{-1}$  on marine *Synechococcus* WH7803-I. Growth, pigmentation, and cell composition. *Deep Sea Research Part A Oceanographic Research Papers* **34**(4): 479-495.
- Kana, T.M., Glibert, P.M., Goericke, R., and Welschmeyer, N.A. (1988) Zeaxanthin and  $\beta$ -carotene in *Synechococcus* WH7803 respond differently to irradiance. *Limnology and Oceanography* **33**(6): 1623-1627.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirose, M., Sugiura, M., and Sasamoto, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research* **3**(3): 109-136.
- Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., and Kimura, T. (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Research* **8**(5): 205-213.
- Karl, D., Bidigare, R., and Letelier, R. (2001) Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep Sea Research Part II: Topical Studies in Oceanography* **48**(8): 1449-1470.
- Karl, D.M., and Lukas, R. (1996) The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Research Part II: Topical Studies in Oceanography* **43**(2): 129-156.
- Karner, M.B., DeLong, E.F., and Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507-510.
- Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J., Sullivan, M., Arendt, D., Benzon, F., and Claverie, J.-M. (2011) A holistic approach to marine eco-systems biology. *PLoS Biology* **9**(10): e1001177.
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**(2): 511-518.

- Kawamura, M., Mimuro, M., and Fujita, Y. (1979) Quantitative relationship between two reaction centers in the photosynthetic system of blue-green algae. *Plant and Cell Physiology* **20**(4): 697-705.
- Kehoe, D.M., and Grossman, A.R. (1996) Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* **273**(5280): 1409-1412.
- Kehoe, D.M., and Grossman, A.R. (1997) New classes of mutants in complementary chromatic adaptation provide evidence for a novel four-step phosphorelay system. *Journal of Bacteriology* **179**(12): 3914-3921.
- Kehoe, D.M., and Gutu, A. (2006) Responding to color: The regulation of complementary chromatic adaptation. In *Annual Review of Plant Biology*. Palo Alto: Annual Reviews, pp. 127-150.
- Kelley, L.A., and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**(3): 363-371.
- Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M., and Yeates, T.O. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* **309**(5736): 936-938.
- Kerfeld, C.A., Heinhorst, S., and Cannon, G.C. (2010) Bacterial microcompartments. *Annual Review of Microbiology* **64**: 391-408.
- Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G.M., Richardson, P., and Chisholm, S.W. (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genetics* **3**(12): 2515-2528.
- Kiang, N.Y., Siefert, J., and Blankenship, R.E. (2007) Spectral signatures of photosynthesis. I. Review of earth organisms. *Astrobiology* **7**(1): 222-251.
- Kikuchi, H., Wako, H., Yura, K., Go, M., and Mimuro, M. (2000) Significance of a two-domain structure in subunits of phycobiliproteins revealed by the normal mode analysis. *Biophysical Journal* **79**(3): 1587-1600.
- Kim, U.-J., Shizuya, H., De Jong, P.J., Birren, B., and Simon, M.I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Research* **20**(5): 1083-1085.
- Kirchman, D.L. (2008) *Microbial ecology of the oceans*. Hoboken, New Jersey: Wiley & sons.
- Kirchman, D.L., Cottrell, M.T., and Lovejoy, C. (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology* **12**(5): 1132-1143.
- Kirk, J.T.O. (1994) *Light and photosynthesis in aquatic ecosystems*. Cambridge, UK: Cambridge University Press.
- Kirschvink, J.L., Gaidos, E.J., Bertani, L.E., Beukes, N.J., Gutzmer, J., Maepa, L.N., and Steinberger, R.E. (2000) Paleoproterozoic snowball Earth: extreme climatic and geochemical global change and its biological consequences. *Proceedings of the National Academy of Sciences of the United States of America* **97**(4): 1400.
- Klotz, A.V., and Glazer, A. (1985) Characterization of the bilin attachment sites in R-phycoerythrin. *Journal of Biological Chemistry* **260**(8): 4856-4863.
- Knoll, A.H. (1992) The early evolution of eukaryotes: a geological perspective. *Science* **256**(5057): 622-627.
- Kodani, S., Ishida, K., and Murakami, M. (1999) Occurrence and identification of UDP-N-acetylmuramyl-pentapeptide from the cyanobacterium *Anabaena cylindrica*. *FEMS Microbiology Letters* **176**(2): 321-325.
- Kohchi, T., Mukougawa, K., Frankenberg, N., Masuda, M., Yokota, A., and Lagarias, J.C. (2001) The *Arabidopsis* HY2 gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. *The Plant Cell Online* **13**(2): 425-436.
- Kok, B., and Gott, W. (1960) Activation spectra of 700 nm absorption change in photosynthesis. *Plant Physiology* **35**(6): 802.
- Kolber, Z., Van Dover, C., Niederman, R., and Falkowski, P. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**(6801): 177-179.
- Kolber, Z.S., Gerald, F., Lang, A.S., Beatty, J.T., Blankenship, R.E., VanDover, C.L., Vetriani, C., Koblizek, M., Rathgeber, C., and Falkowski, P.G. (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* **292**(5526): 2492-2495.
- Kondo, K., Geng, X.X., Katayama, M., and Ikeuchi, M. (2005) Distinct roles of CpcG1 and CpcG2 in phycobilisome assembly in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynthesis Research* **84**(1): 269-273.
- Kondo, K., Mullineaux, C.W., and Ikeuchi, M. (2009) Distinct roles of CpcG1-phycobilisome and CpcG2-phycobilisome in state transitions in a cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynthesis Research* **99**(3): 217-225.
- Krause, L., Diaz, N.N., Bartels, D., Edwards, R.A., Pühler, A., Rohwer, F., Meyer, F., and Stoye, J. (2006) Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics* **22**(14): e281-e289.
- Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F., Edwards, R.A., and Stoye, J. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research* **36**(7): 2230-2239.
- Krauss, N., Hinrichs, W., Witt, I., Fromme, P., Pritzkow, W., Dauter, Z., Betzel, C., Wilson, K.S., Witt, H.T., and Saenger, W. (1993) Three-dimensional structure of

- system I of photosynthesis at 6 Å resolution. *Nature* **361**(6410): 326-331.
- Kristensen, D.M., Mushegian, A.R., Dolja, V.V., and Koonin, E.V.** (2010) New dimensions of the virus world discovered through metagenomics. *Trends in Microbiology* **18**(1): 11-19.
- Krom, M., Kress, N., Brenner, S., and Gordon, L.** (1991) Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. *Limnology and Oceanography* **36**(3): 424-432.
- Kühlbrandt, W., Wang, D.N., and Y., F.** (1994) Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**: 614-621.
- Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P.** (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**(1): 118-123.
- Kupka, M., Zhang, J., Fu, W.L., Tu, J.M., Böhm, S., Su, P., Chen, Y., Zhou, M., Scheer, H., and Zhao, K.H.** (2009) Catalytic mechanism of S-type phycobiliprotein lyase. Chaperone-like action and functional amino acid residues. *Journal of Biological Chemistry* **284**(52): 36405-36414.
- Küpper, H., Šetlík, I., Seibert, S., Prášil, O., Šetlikova, E., Strittmatter, M., Levitan, O., Lohscheider, J., Adamska, I., and Berman-Frank, I.** (2008) Iron limitation in the marine cyanobacterium *Trichodesmium* reveals new insights into regulation of photosynthesis and nitrogen fixation. *New Phytologist* **179**(3): 784-798.
- Kützing, F.T.** (1843) *Phycologia generalis, oder Anatomie, Physiologie und Systemkunde der Tange*. Leipzig.
- Kvist, T., Ahring, B.K., Lasken, R.S., and Westermann, P.** (2007) Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Applied Microbiology and Biotechnology* **74**(4): 926-935.
- 
- L**
- Labrie, S.J., Frois-Moniz, K., Osburne, M.S., Kelly, L., Roggensack, S.E., Sullivan, M.B., Gearin, G., Zeng, Q., Fitzgerald, M., Henn, M.R., and Chisholm, S.W.** (2013) Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environmental Microbiology*: doi: 10.1111/1462-2920.12053.
- Lami, R., Cottrell, M.T., Campbell, B.J., and Kirchman, D.L.** (2009) Light-dependent growth and proteorhodopsin expression by Flavobacteria and SAR11 in experiments with Delaware coastal waters. *Environmental Microbiology* **11**(12): 3201-3209.
- Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R.** (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* **82**(20): 6955-6959.
- LaRoche, J., vanderStaa, G.W.M., Partensky, F., Ducret, A., Aebersold, R., Li, R., Golden, S.S., Hiller, R.G., Wrench, P.M., Larkum, A.W.D., and Green, B.R.** (1996) Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. *Proceedings of the National Academy of Sciences of the United States of America* **93**(26): 15244-15248.
- Larsson, J., Nylander, J.A., and Bergman, B.** (2011) Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evolutionary Biology* **11**(1): 187.
- Lasken, R.S., and Stockwell, T.B.** (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnology* **7**: 19.
- Lasken, R.S.** (2012) Genomic sequencing of uncultured microorganisms from single cells. *Nature Reviews Microbiology* **10**(9): 631-640.
- Laudenbach, D.E., and Straus, N.A.** (1988) Characterization of a cyanobacterial iron stress-induced gene similar to *psbC*. *Journal of Bacteriology* **170**(11): 5018-5026.
- Lavin, P., Gonzalez, B., Santibanez, J.F., Scanlan, D.J., and Ulloa, O.** (2010) Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environmental Microbiology Reports* **2**(6): 728-738.
- Lawry, N.H., and Jensen, T.E.** (1979) Deposition of condensed phosphate as an effect of varying sulfur deficiency in the cyanobacterium *Synechococcus* sp. (*Anacystis nidulans*). *Archives of Microbiology* **120**(1): 1-7.
- Lawry, N.H., and Jensen, T.E.** (1986) Condensed phosphate deposition, sulfur amino acid use, and unidirectional transsulfuration in *Synechococcus leopoliensis*. *Archives of Microbiology* **144**(4): 317-323.
- Laws, E.A.** (1991) Photosynthetic quotients, new production and net community production in the open ocean. *Deep Sea Research Part A Oceanographic Research Papers* **38**(1): 143-167.
- Lejeusne, C., Chevaldonne, P., Pergent-Martini, C., Boudouresque, C.F., and Perez, T.** (2010) Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. *Trends in Ecology & Evolution* **25**(4): 250-260.
- Leliaert, F., Verbruggen, H., and Zechman, F.W.** (2011) Into the deep: New discoveries at the base of the green plant phylogeny. *Bioessays* **33**(9): 683-692.
- Lemberg, R.** (1928) Die chromoproteide der rootalgen. I. *Justus Liebigs Annalen der Chemie* **461**(1): 46-89.
- Lemberg, R.** (1930) Die chromoproteide der rootalgen. II. Spaltung mit pepsin un säuren. Isolierung eines

- pyrrolfarbstoffs. *Justus Liebigs Annalen der Chemie* **477**: 195-145.
- Lemberg, R., and Bader, G.** (1933) Die phycobiline der rotalgen. Überführung in mesobilirubin and dehydro - mesobilirubin. *Justus Liebigs Annalen der Chemie* **505**: 151-177.
- Leonhardt, K., and Straus, N.A.** (1992) An iron stress operon involved in photosynthetic electron transport in the marine cyanobacterium *Synechococcus* sp. PCC 7002. *Journal of General Microbiology* **138**(8): 1613-1621.
- Leonhardt, K., and Straus, N.A.** (1994) Photosystem II genes *isiA*, *psbDI* and *psbC* in *Anabaena* sp. PCC 7120: cloning, sequencing and the transcriptional regulation in iron-stressed and iron-repleted cells. *Plant Molecular Biology* **24**(1): 63-73.
- Lepère, C., Demura, M., Kawachi, M., Romac, S., Probert, I., and Vaultot, D.** (2011) Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiology Ecology* **76**(3): 513-523.
- Leung, K., Zahn, H., Leaver, T., Konwar, K.M., Hanson, N.W., Pagé, A.P., Lo, C.-C., Chain, P.S., Hallam, S.J., and Hansen, C.L.** (2012) A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proceedings of the National Academy of Sciences of the United States of America* **109**(20): 7665-7670.
- Lewin, R., and Withers, N.W.** (1975) Extraordinary pigment composition of a prokaryotic alga. *Nature* **256**: 735-737.
- Lewin, R.A.** (1976) Prochlorophyta as a proposed new division of algae. *Nature* **261**(5562): 697-698.
- Li, L., and Kehoe, D.M.** (2005) *In vivo* analysis of the roles of conserved aspartate and histidine residues within a complex response regulator. *Molecular Microbiology* **55**(5): 1538-1552.
- Li, L., Alvey, R.M., Bezy, R.P., and Kehoe, D.M.** (2008) Inverse transcriptional activities during complementary chromatic adaptation are controlled by the response regulator RcaC binding to red and green light-responsive promoters. *Molecular Microbiology* **68**(2): 286-297.
- Li, W.** (1994) Primary productivity of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnology and Oceanography* **39**(1): 169-175.
- Li, W.** (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology and Oceanography* **43**(1): 1746-1753.
- Li, W.** (2002) Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature* **419**(6903): 154-157.
- Li, W.K.W., Rao, D.V.S., Harrison, W.G., Smith, J.C., Cullen, J.J., Irwin, B., and Platt, T.** (1983) Autotrophic picoplankton in the tropical ocean. *Science* **219**(4582): 292-295.
- Lichtlé, C., Thomas, J.C., Spilar, A., and Partensky, F.** (1995) Immunological and ultrastructural characterization of the photosynthetic complexes of the prochlorophyte *Prochlorococcus* (Oxychlorobacteria). *Journal of Phycology* **31**(6): 934-941.
- Liles, M.R., Manske, B.F., Bintrim, S.B., Handelsman, J., and Goodman, R.M.** (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Applied and Environmental Microbiology* **69**(5): 2684-2691.
- Lindell, D., and Post, A.F.** (1995) Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat), Red Sea. *Limnology and Oceanography* **40**(6): 1130-1141.
- Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W.** (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proceedings of the National Academy of Sciences of the United States of America* **101**(30): 11013-11018.
- Lindell, D., Penno, S., Al-Qutob, M., David, E., Rivlin, T., Lazar, B., and Post, A.F.** (2005) Expression of the nitrogen stress response gene *ntcA* reveals nitrogen-sufficient *Synechococcus* populations in the oligotrophic northern Red Sea. *Limnology and Oceanography* **50**(6): 1932-1944.
- Link, D.R., Grasland-Mongrain, E., Duri, A., Sarrazin, F., Cheng, Z., Cristobal, G., Marquez, M., and Weitz, D.A.** (2006) Electric control of droplets in microfluidic devices. *Angewandte Chemie International Edition* **45**(16): 2556-2560.
- Liu, L.N., Chen, X.L., Zhang, Y.Z., and Zhou, B.C.** (2005) Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: an overview. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1708**(2): 133-142.
- Lohmann, H.** (1908) Untersuchungen zur Feststellung des vollständigen Gehaltes des Meeres an Plankton. *Wiss Meeresunters Abt Kiel* **10**: 129-370.
- Lohmann, H.** (1911) Über das nannoplankton und die Zentrifugierung kleinster Wasserproben zur Gewinnung desselben in lebendem Zustande. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* **4**: 1-38.
- Lomax, T., Conley, P., Schilling, J., and Grossman, A.** (1987) Isolation and characterization of light-regulated phycobilisome linker polypeptide genes and their transcription as a polycistronic mRNA. *Journal of Bacteriology* **169**(6): 2675-2684.
- López-García, P., Brochier, C., Moreira, D., and Rodríguez-Valera, F.** (2004) Comparative analysis of a

- genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environmental Microbiology* **6**(1): 19-34.
- Lorenzen, C.J.** (1963) Diurnal variation in photosynthetic activity of natural phytoplankton populations. *Limnology and Oceanography* **8**(1): 56-62.
- Lourenço, S.O., Barbarino, E., Marquez, U.M.L., and Aidar, E.** (1998) Distribution of intracellular nitrogen in marine microalgae : basis for the calculation of specific nitrogen-to-protein conversion factors. *Journal of Phycology* **34**(5): 798-811.
- Lu, J., Chen, F., and Hodson, R.E.** (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Applied and Environmental Microbiology* **67**(7): 3285-3290.
- Ludwig, M., and Gibbs, S.P.** (1985) DNA is present in the nucleomorph of cryptomonads: further evidence that the chloroplast evolved from a eukaryotic endosymbiont. *Protoplasts* **127**(1): 9-20.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., Lai, T., Steppi, S., Jobb, G., and Förster, W.** (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* **32**(4): 1363-1371.
- Lundell, D., and Glazer, A.N.** (1980) Linker polypeptides of cyanobacterial phycobilisomes. *Federation Proceedings* **39**(1676).
- Lundell, D., Williams, R., and Glazer, A.** (1981a) Molecular architecture of a light-harvesting antenna. *In vitro* assembly of the rod substructures of *Synechococcus* 6301 phycobilisomes. *Journal of Biological Chemistry* **256**(7): 3580-3592.
- Lundell, D.J., Yamanaka, G., and Glazer, A.N.** (1981b) A terminal energy acceptor of the phycobilisome: the 75,000-dalton polypeptide of *Synechococcus* 6301 phycobilisomes - a new biliprotein. *The Journal of Cell Biology* **91**(1): 315-319.
- Luo, C., Tsementzi, D., Kyrpides, N.C., and Konstantinidis, K.T.** (2011a) Individual genome assembly from complex community short-read metagenomic datasets. *The ISME Journal* **6**(4): 898-901.
- Luo, H., Friedman, R., Tang, J., and Hughes, A.L.** (2011b) Genome reduction by deletion of paralogs in the marine cyanobacterium *Prochlorococcus*. *Molecular Biology and Evolution* **28**(10): 2751-2760.
- MacColl, R.** (2004) Allophycocyanin and energy transfer. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1657**(2-3): 73-81.
- Macdonald, K.C., Becker, K., Spiess, F.N., and Ballard, R.D.** (1980) Hydrothermal heat-flux of the black smoker vents on the East-Pacific rise. *Earth and Planetary Science Letters* **48**(1): 1-7.
- MacNeil, I., Tiong, C., Minor, C., August, P., Grossman, T., Loiacono, K., Lynch, B., Phillips, T., Narula, S., and Sundaramoorthi, R.** (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *Journal of Molecular Microbiology and Biotechnology* **3**(2): 301-308.
- Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W.** (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *The ISME Journal* **4**(10): 1252-1264.
- Malmstrom, R.R., Rodrigue, S., Huang, K.H., Kelly, L., Kern, S.E., Thompson, A., Roggensack, S., Berube, P.M., Henn, M.R., and Chisholm, S.W.** (2013) Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *The ISME Journal* **7**: 184-198.
- Mann, D., and Droop, S.** (1996) Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* **336**(1): 19-32.
- Mann, E.L., and Chisholm, S.W.** (2000) Iron limits the cell division rate of *Prochlorococcus* in the eastern equatorial Pacific. *Limnology and Oceanography* **45**(5): 1067-1076.
- Mann, E.L., Ahlgren, N., Moffett, J.W., and Chisholm, S.W.** (2002) Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnology and Oceanography* **47**(4): 976-988.
- Marcy, Y., Ishoey, T., Lasken, R.S., Stockwell, T.B., Walenz, B.P., Halpern, A.L., Beeson, K.Y., Goldberg, S.M., and Quake, S.R.** (2007) Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genetics* **3**(9): e155.
- Mardis, E.R.** (2008a) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**(3): 133-141.
- Mardis, E.R.** (2008b) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* **9**: 387-402.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., and Chen, Z.** (2005) Genome sequencing in microfabricated high-density picoliter reactors. *Nature* **437**(7057): 376-380.
- Marie, D., Brussaard, C., Partensky, F., Vaulot, D., and Wiley, J.** (1999) Flow cytometric analysis of

## M

- MacArthur, R., and Levins, R.** (1967) The limiting similarity, convergence, and divergence of coexisting species. *American Naturalist*: 377-385.
- MacColl, R.** (1998) Cyanobacterial phycobilisomes. *Journal of Structural Biology* **124**(2-3): 311-334.

- phytoplankton, bacteria and viruses. *Current Protocols in Cytometry Supplement 10*: 11.11.11-11.11.15.
- Marin, B., and Melkonian, M.** (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov (Chlorophyta) based on sequence comparisons of the nuclear- and lastid-encoded rRNA operons. *Protist* **161**(2): 304-336.
- Marquardt, J., Senger, H., Miyashita, H., Miyachi, S., and Mörschel, E.** (1997) Isolation and characterization of biliprotein aggregates from *Acaryochloris marina*, a *Prochloron*-like prokaryote containing mainly chlorophyll *d*. *FEBS Letters* **410**(2): 428-432.
- Marshall, I.P., Blainey, P.C., Spormann, A.M., and Quake, S.R.** (2012) A single-cell genome for *Thiovulum* sp. *Applied and Environmental Microbiology* **78**(24): 8555-8563.
- Marston, M.F., and Sallee, J.L.** (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Applied and Environmental Microbiology* **69**(8): 4639-4647.
- Martel, C.M.** (2009) Conceptual bases for prey biorecognition and feeding selectivity in the microplanktonic marine phagotroph *Oxyrrhis marina*. *Microbial Ecology* **57**(4): 589-597.
- Martin, J.H., Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S.J., Hunter, C.N., Elrod, V.A., Nowicki, J.L., Coley, T.L., Barber, R.T., Lindley, S., Watson, A.J., Van Scoy, K., Law, C.S., Liddicoat, M.I., Ling, R., Stanton, T., Stockel, J., Collins, C., Anderson, A., Bidigare, R., Ondrusek, M., Latasa, M., Millero, F.J., Lee, K., Yao, W., Zhang, J.Z., Friederich, G., Sakamoto, C., Chavez, F., Buck, K., Kolber, Z., Greene, R., Falkowski, P., Chisholm, S.W., Hoge, F., Swift, R., Yungel, J., Turner, S., Nightingale, P., Hatton, A., Liss, P., and Tindale, N.W.** (1994) Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**(6493): 123-129.
- Martinez-Garcia, M., Brazel, D., Poulton, N.J., Swan, B.K., Gomez, M.L., Masland, D., Sieracki, M.E., and Stepanauskas, R.** (2011a) Unveiling *in situ* interactions between marine protists and bacteria through single cell sequencing. *The ISME Journal* **6**(3): 703-707.
- Martinez-Garcia, M., Swan, B.K., Poulton, N.J., Gomez, M.L., Masland, D., Sieracki, M.E., and Stepanauskas, R.** (2011b) High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *The ISME Journal* **6**(1): 113-123.
- Martínez, J.M., Poulton, N.J., Stepanauskas, R., Sieracki, M.E., and Wilson, W.H.** (2011) Targeted sorting of single virus-infected cells of the coccolithophore *Emiliania huxleyi*. *PLoS One* **6**(7): e22520.
- Martiny, A.C., Kathuria, S., and Berube, P.M.** (2009a) Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proceedings of the National Academy of Sciences of the United States of America* **106**(26): 10787-10792.
- Martiny, A.C., Tai, A.P.K., Veneziano, D., Primeau, F., and Chisholm, S.W.** (2009b) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environmental Microbiology* **11**(4): 823-832.
- Mary, I., Garczarek, L., Tarran, G.A., Kolowrat, C., Terry, M.J., Scanlan, D.J., Burkill, P.H., and Zubkov, M.V.** (2008) Diel rhythmicity in amino acid uptake by *Prochlorococcus*. *Environmental Microbiology* **10**(8): 2124-2131.
- Mass, T., Kline, D., Roopin, M., Veal, C., Cohen, S., Iluz, D., and Levy, O.** (2010) The spectral quality of light is a key driver of photosynthesis and photoadaptation in *Stylophora pistillata* colonies from different depths in the Red Sea. *The Journal of Experimental Biology* **213**(23): 4084-4091.
- Mazard, S.** (2008) From ecology to genomics : Phosphorus acquisition and regulatory mechanisms in marine cyanobacteria mediate niche adaptation. In: Coventry, United-Kingdom: University of Warwick, p. 273.
- Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J.** (2012a) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environmental Microbiology* **14**(2): 372-386.
- Mazard, S., Wilson, W.H., and Scanlan, D.J.** (2012b) Dissection the physiological response to phosphorus stress in marine *Synechococcus* isolates (Cyanophyceae). *Journal of Phycology* **48**(1): 94-105.
- Mazel, D., Guglielmi, G., Houmard, J., Sidler, W., Bryant, D.A., and De Marsac, N.T.** (1986) Green light induces transcription of the phycoerythrin operon in the cyanobacterium *Calothrix* 7601. *Nucleic Acids Research* **14**(21): 8279-8290.
- Mazel, D., Houmard, J., and Marsac, N.T.** (1988) A multigene family in *Calothrix* sp. PCC 7601 encodes phycocyanin, the major component of the cyanobacterial light-harvesting antenna. *Molecular and General Genetics MGG* **211**(2): 296-304.
- Mazel, D., and Marliere, P.** (1989) Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature* **341**(6239): 245-248.
- McCarren, J., Heuser, J., Roth, R., Yamada, N., Martone, M., and Brahmasha, B.** (2005) Inactivation of *swmA* results in the loss of an outer cell layer in a swimming *Synechococcus* strain. *Journal of Bacteriology* **187**(1): 224-230.
- McCarren, J., and Brahmasha, B.** (2007) SwmB, a 1.12-megadalton protein that is required for nonflagellar swimming motility in *Synechococcus*. *Journal of Bacteriology* **189**(3): 1158-1162.
- McCarren, J., and DeLong, E.F.** (2007) Proteorhodopsin photosystem gene clusters exhibit co-evolutionary

- trends and shared ancestry among diverse marine microbial phyla. *Environmental Microbiology* **9**(4): 846-858.
- McCarren, J., and Brahamsha, B.** (2009) Swimming motility mutants of marine *Synechococcus* affected in production and localization of the S-Layer protein SwmA. *Journal of Bacteriology* **191**(3): 1111-1114.
- Melezhik, V.A., Fallick, A., Hanski, E.J., Kump, L.R., Lepland, A., Prave, A.R., and Strauss, H.** (2005) Emergence of the aerobic biosphere during the Archean-Proterozoic transition: Challenges of future research. *Gsa Today* **15**(11): 4-11.
- Mella-Flores, D.** (2011) Diversité génétique et fonctionnelle des cyanobactéries picoplanctoniques marines et adaptation aux stress environnementaux. In: Université Pierre et Marie Curie (Paris VI), p. 314 pages.
- Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahe, F., Bariat, L., Courties, C., Marie, D., Ras, J., Mauriac, R., Jeanthon, C., Bendif, E.M., Ostrowski, M., Scanlan, D.J., and Garczarek, L.** (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences* **8**(9): 2785-2804.
- Metzker, M.L.** (2009) Sequencing technologies—the next generation. *Nature Reviews Genetics* **11**(1): 31-46.
- Mills, M.M., Ridame, C., Davey, M., La Roche, J., and Geider, R.J.** (2004) Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* **429**(6989): 292-294.
- Mimuro, M., Füglistaller, P., Rümbeili, R., and Zuber, H.** (1986) Functional assignment of chromophores and energy transfer in C phycocyanin isolated from the thermophilic cyanobacterium *Mastigocladus laminosus*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **848**(2): 155-166.
- Miyashita, H., Ikemoto, H., and Kurano, N.** (1996) Chlorophyll *d* as a major pigment. *Nature* **383**: 402.
- Miyashita, H., Ikemoto, H., Kurano, N., Miyachi, S., and Chihara, M.** (2003) *Acaryochloris marina* gen. et sp. nov. (Cyanobacteria), an oxygenic photosynthetic prokaryote containing chl *d* as a major pigment. *Journal of Phycology* **39**(6): 1247-1253.
- Moisan, J.R., Moisan, T.A., and Abbott, M.R.** (2002) Modelling the effect of temperature on the maximum growth rates of phytoplankton populations. *Ecological Modelling* **153**(3): 197-215.
- Montgomery, B.L., Casey, E.S., Grossman, A.R., and Kehoe, D.M.** (2004) Apla, a member of a new class of phycobiliproteins lacking a traditional role in photosynthetic light harvesting. *Journal of Bacteriology* **186**(21): 7420-7428.
- Moon-van der Staay, S.Y., De Wachter, R., and Vaultot, D.** (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**(6820): 607-610.
- Moore, L.R., Goericke, R., and Chisholm, S.W.** (1995) Comparative physiology of *Synechococcus* and *Prochlorococcus* - influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Marine Ecology-Progress Series* **116**(1-3): 259-275.
- Moore, L.R., Rocap, G., and Chisholm, S.W.** (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**(6684): 464-467.
- Moore, L.R., Coe, A., Zinser, E.R., Saito, M.A., Sullivan, M.B., Lindell, D., Frois-Moniz, K., Waterbury, J., and Chisholm, S.W.** (2007) Culturing the marine cyanobacterium *Prochlorococcus*. *Limnology and Oceanography-Methods* **5**: 353-362.
- Moran, M.A., Buchan, A., González, J.M., Heidelberg, J.F., Whitman, W.B., Kiene, R.P., Henriksen, J.R., King, G.M., Belas, R., and Fuqua, C.** (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**(7019): 910-913.
- Moreira, D., Rodríguez-Valera, F., and López-García, P.** (2004) Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environmental Microbiology* **6**(9): 959-969.
- Morel, A., and Prieur, L.** (1977) Analysis of variations in ocean color. *Limnology and Oceanography* **22**(4): 709-722.
- Morel, A.** (1988) Optical modeling of the upper ocean in relation to its biogenous matter content (case-I Waters). *Journal of Geophysical Research: Oceans* **93**(C9): 10749-10768.
- Morel, A., Ahn, Y.H., Partensky, F., Vaultot, D., and Claustre, H.** (1993) *Prochlorococcus* and *Synechococcus* - A comparative study of their optical properties in relation to their size and pigmentation. *Journal of Marine Research* **51**(3): 617-649.
- Morel, A., and Maritorea, S.** (2001) Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research: Oceans* **106**(C4): 7163-7180.
- Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., and Tieche, F.** (2007) Optical properties of the "clearest" natural waters. *Limnology and Oceanography* **52**(1): 217-229.
- Morel, F., and Price, N.** (2003) The biogeochemical cycles of trace metals in the oceans. *Science* **300**(5621): 944-947.
- Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J.** (2002) SAR 11 clade dominates ocean surface

bacterioplankton communities. *Nature* **420**(6917): 806-810.

**Mühling, M., Fuller, N.J., Millard, A., Somerfield, P.J., Marie, D., Wilson, W.H., Scanlan, D.J., Post, A.F., Joint, I., and Mann, N.H.** (2005) Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environmental Microbiology* **7**(4): 499-508.

**Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H.** (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor Symposia on Quantitative Biology*, pp. 263-273.

**Muramoto, T., Kohchi, T., Yokota, A., Hwang, I., and Goodman, H.M.** (1999) The *Arabidopsis* photomorphogenic mutant *hy1* is deficient in phytochrome chromophore biosynthesis as a result of a mutation in a plastid heme oxygenase. *The Plant Cell Online* **11**(3): 335-347.

**Mußmann, M., Hu, F.Z., Richter, M., De Beer, D., Preisler, A., Jørgensen, B.B., Huntemann, M., Glöckner, F.O., Amann, R., and Koopman, W.J.** (2007) Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biology* **5**(9): e230.

**Myers, J.** (1980) On the algae: thoughts about physiology and measurements of efficiency. In *Primary Productivity in the Sea*. Falkowski, P. (ed). New-York: Plenum Press, pp. 1-16.

## N

**Nägeli, C.** (1849) *Gattungen einzelliger Algen physiologisch und systematisch bearbeitet*. Zürich.

**Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C.** (2008) Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environmental Microbiology* **10**(6): 1526-1535.

**Neuschaefler-Rube, O., Böger, P., and Ernst, A.** (2002) Interference of an *apcA* insertion with complementary chromatic adaptation in the diazotrophic *Synechocystis* sp. strain BO 8402. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1553**(3): 279-295.

**Nield, J., Rizkallah, P.J., Barber, J., and Chayen, N.E.** (2003) The 1.45 Å three-dimensional structure of C-phycoyanin from the thermophilic cyanobacterium *Synechococcus elongatus*. *Journal of Structural Biology* **141**(2): 149-155.

**Niyogi, K.K.** (2000) Safety valves for photosynthesis. *Current Opinion in Plant Biology* **3**(6): 455-460.

**Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwälder, B., J., B., and Schlötterer, C.** (2010) Contrasting seasonal niche separation between rare

and abundant taxa conceals the extent of protist diversity. *Molecular Ecology* **19**(14): 2908-2915.

**Not, F., Valentin, K., Romari, K., Lovejoy, C., Massana, R., Töbe, K., Vaultot, D., and Medlin, L.K.** (2007) Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315**(5809): 253-255.

**Nugent, J.H.A., Rich, A.M., and Evans, M.C.W.** (2001) Photosynthetic water oxidation: towards a mechanism. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1503**(1): 138-146.

## O

**Oesterhelt, D., and Stoeckenius, W.** (1971) Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature* **233**(39): 149-152.

**Ohki, K., Gantt, E., Lipschultz, C.A., and Ernst, M.C.** (1985) Constant phycobilisome size in chromatically adapted acilus cells of the cyanobacterium *Tolypothrix tenuis*, and variation in *Nostoc* sp. *Plant Physiology* **79**(4): 943-948.

**Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., and Stahl, D.A.** (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annual Reviews in Microbiology* **40**(1): 337-365.

**Olson, J.M., and Blankenship, R.E.** (2004) Thinking about the evolution of photosynthesis. *Photosynthesis Research* **80**(1): 373-386.

**Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V.** (1988) Analysis of *Synechococcus* pigment types in the sea using single and dual beam flow-cytometry. *Deep-Sea Research Part A-Oceanographic Research Papers* **35**(3): 425-440.

**Olson, R.J., Chisholm, S.W., Zettler, E.R., Altabet, M.A., and Dusenberry, J.A.** (1990a) Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep Sea Research Part A Oceanographic Research Papers* **37**(6): 1033-1051.

**Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V.** (1990b) Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnology and Oceanography* **35**(1): 45-58.

**Ong, L.J., and Glazer, A.N.** (1987) R-Phycocyanin-II, a new phycocyanin occurring in marine *Synechococcus* species - Identification of the terminal energy acceptor bilin in phycocyanins. *Journal of Biological Chemistry* **262**(13): 6323-6327.

**Ong, L.J., and Glazer, A.N.** (1988) Structural studies of phycobiliproteins in unicellular marine cyanobacteria. In *Light-energy transduction in photosynthesis: higher plants and bacterial models*. Stevens, S.E.J., and Bryant, D.A. (eds). Rockville: American Society of Plant Physiologists, pp. 102-121.

- Öquist, G. (1974) Light-induced changes in pigment composition of photosynthetic lamellae and cell-free extracts from the blue-green alga *Anacystis nidulans*. *Physiologia Plantarum* **30**(1): 45-48.
- Ostrowski, M., Mazard, S., Tetu, S.G., Phillippy, K., Johnson, A., Palenik, B., Paulsen, I.T., and Scanlan, D.J. (2010) PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*. *The ISME Journal* **4**(7): 908-921.
- Ottesen, E.A., Hong, J.W., Quake, S.R., and Leadbetter, J.R. (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314**(5804): 1464-1467.
- Oz, A., Sabehi, G., Koblížek, M., Massana, R., and Béjà, O. (2005) *Roseobacter*-like bacteria in Red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Applied and Environmental Microbiology* **71**(1): 344-353.
- 
- P
- Pace, N., Stahl, D., Lane, D., and Olsen, G. (1986) The analysis of natural microbial populations by ribosomal RNA sequences. *Advances in Microbial Ecology* **9**: 1-55.
- Paerl, R., Johnson, K.S., Welsh, R.M., Worden, A.Z., Chavez, F.P., and Zher, J.P. (2011a) Differential distributions of *Synechococcus* subgroups across the California current system. *Frontiers in Microbiology* **2**(59): 1-22.
- Paerl, R., Turk, K.A., Beinart, R.A., Chavez, F.P., and Zher, J.P. (2011b) Seasonal change in the abundance of *Synechococcus* and multiple distinct phylotypes in Monterey Bay determined by *rbcl* and *narB* quantitative PCR. *Environmental Microbiology* **14**(3): 580-593.
- Paerl, R.W., Foster, R.A., Jenkins, B.D., Montoya, J.P., and Zehr, J.P. (2008) Phylogenetic diversity of cyanobacterial *narB* genes from various marine habitats. *Environmental Microbiology* **10**(12): 3377-3387.
- Pakrasi, H.B. (1995) Genetic analysis of the form and function of photosystem I and photosystem II. *Annual Review of Genetics* **29**(1): 755-776.
- Palenik, B. (1994) Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. *Applied and Environmental Microbiology* **60**(9): 3212-3219.
- Palenik, B. (2001) Chromatic adaptation in marine *Synechococcus* strains. *Applied and Environmental Microbiology* **67**(2): 991-994.
- Palenik, B., Brahamsha, B., Larimer, F., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E., and McCarren, J. (2003) The genome of a motile marine *Synechococcus*. *Nature* **424**(6952): 1037-1042.
- Palenik, B., Ren, Q.H., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., Durkin, A.S., Daugherty, S.C., Sullivan, S.A., Khouri, H., Mohamoud, Y., Halpin, R., and Paulsen, I.T. (2006) Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proceedings of the National Academy of Sciences of the United States of America* **103**(36): 13555-13559.
- Palenik, B., Ren, Q., Tai, V., and Paulsen, I.T. (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environmental Microbiology* **11**(2): 349-359.
- Palmer, J.D., and Delwiche, C.F. (1996) Second-hand chloroplasts and the case of the disappearing nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **93**(15): 7432.
- Pamp, S.J., Harrington, E.D., Quake, S.R., Relman, D.A., and Blainey, P.C. (2012) Single-cell sequencing provides clues about the host interactions of segmented filamentous bacteria (SFB). *Genome Research* **22**(6): 1107-1119.
- Parbel, A., and Scheer, H. (2000) Model for the phycobilisome rod with interlocking disks based on domain-weighted linker-polypeptide sequence homologies of *Mastigocladus laminosus*. *International Journal of Photoenergy* **2**(1): 31-40.
- Park, Y.I., Sandström, S., Gustafsson, P., and Öquist, G. (1999) Expression of the *isiA* gene is essential for the survival of the cyanobacterium *Synechococcus* sp. PCC 7942 by protecting photosystem II from excess light under iron limitation. *Molecular Microbiology* **32**(1): 123-129.
- Partensky, F., Blanchot, J., Lantoine, F., Neveux, J., and Marie, D. (1996) Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers* **43**(8): 1191-1213.
- Partensky, F., Blanchot, G., and Vaultot, D. (1999a) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters : a review. *Bulletin de l'Institut océanographique* **19** (Marine Cyanobacteria): 457-475.
- Partensky, F., Hess, W.R., and Vaultot, D. (1999b) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews* **63**(1): 106-127.
- Partensky, F., and Garczarek, L. (2003) The photosynthetic apparatus of chlorophyll *b* and *d* containing oxychlorobacteria. In *Photosynthesis in algae*. Larkum, A.W.D., Douglas, S.E., and Raven, J.A. (eds). Dordrecht, The Netherlands: Kluwer Academic Publisher, pp. 29-62.

- Partensky, F., and Garczarek, L.** (2010) *Prochlorococcus*: advantages and limits of minimalism. In *Annual Review of Marine Science*. Palo Alto: Annual Reviews, pp. 305-331.
- Passarge, J., Hol, S., Escher, M., and Huisman, J.** (2006) Competition for nutrients and light: stable coexistence, alternative stable states, or competitive exclusion? *Ecological Monographs* **76**(1): 57-72.
- Pawlowski, J., Christen, R., Lecroq, B., Bachar, D., Shahbazkia, H.R., Amaral-Zettler, L., and Guillou, L.** (2011) Eukaryotic richness in the abyss: insights from pyrotag sequencing. *PLoS One* **6**(4): e18169.
- Pedrós-Alió, C.** (2006) Marine microbial diversity: can it be determined? *Trends in Microbiology* **14**(6): 257-263.
- Penno, S., Lindell, D., and Post, A.F.** (2006) Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environmental Microbiology* **8**(7): 1200-1211.
- Perkins, F.O., Haas, L.W., Phillips, D.E., and Webb, K.L.** (1981) Ultrastructure of a marine *Synechococcus* possessing spinae. *Canadian Journal of Microbiology* **27**(3): 318-329.
- Perry, M., Talbot, M., and Alberte, R.** (1981) Photoadaptation in marine phytoplankton: response of the photosynthetic unit. *Marine Biology* **62**(2): 91-101.
- Pham, V.D., Konstantinidis, K.T., Palden, T., and DeLong, E.F.** (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environmental Microbiology* **10**(9): 2313-2330.
- Pizarro, S.A., and Sauer, K.** (2001) Spectroscopic study of the light-harvesting protein C-Phycocyanin associated with colorless linker peptides. *Photochemistry and Photobiology* **73**(5): 556-563.
- Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L., and Keller, M.** (2007) Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Applied and Environmental Microbiology* **73**(10): 3205-3214.
- Polovina, J.J., Howell, E.A., and Abecassis, M.** (2008) Ocean's least productive waters are expanding. *Geophysical Research Letters* **35**(3): L03618.
- Pomeroy, L.R., William, P.J., Azam, F., and Hobbie, E.A.** (2007) The microbial loop. *Oceanography* **20**(2): 28-33.
- Pomposiello, P.J., and Demple, B.** (2001) Redox-operated genetic switches: the SoxR and OxyR transcription factors. *Trends in Biotechnology* **19**(3): 109-113.
- Post, A.F., Penno, S., Zandbank, K., Paytan, A., Huse, S.M., and Welch, D.M.** (2011) Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, Northern Red Sea. *Frontiers in Microbiology* **2**(131): doi:10.3389/fmicb.2011.00131.
- Postius, C., Neuschaefter-Rube, O., Haid, V., and Boger, P.** (2001) N<sub>2</sub>-fixation and complementary chromatic adaptation in non-heterocystous cyanobacteria from Lake Constance. *FEMS Microbiology Ecology* **37**(2): 117-125.
- 
- Q**
- Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., Eck, J., Sensen, C.W., and Schleper, C.** (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environmental Microbiology* **4**(10): 603-611.
- Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S.C., Treusch, A.H., Eck, J., and Schleper, C.** (2003) Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Molecular Microbiology* **50**(2): 563-575.
- Quevedo, M., and Anadón, R.** (2001) Protist control of phytoplankton growth in the subtropical north-east Atlantic. *Marine Ecology Progress Series* **221**: 29-38.
- 
- R**
- Raghunathan, A., Ferguson, H.R., Bornarth, C.J., Song, W.M., Driscoll, M., and Lasken, R.S.** (2005) Genomic DNA amplification from a single bacterium. *Applied and Environmental Microbiology* **71**(6): 3342-3347.
- Rappé, M.S., Vergin, K., and Giovannoni, S.J.** (2000) Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiology Ecology* **33**(3): 219-232.
- Rappé, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J.** (2002) Cultivation of the ubiquitous SAR 11 marine bacterioplankton clade. *Nature* **418**(6898): 630-633.
- Rappé, M.S., and Giovannoni, S.J.** (2003) The uncultured microbial majority. *Annual Review of Microbiology* **57**: 369-394.
- Raven, J.A.** (2009) Functional evolution of photochemical energy transformations in oxygen-producing organisms. *Functional Plant Biology* **36**(6): 505-515.
- Redfield, A.C.** (1934) On the proportions of organic derivations in sea water and their relation to the composition of plankton. In *James Johnstone Memorial Volume*. Daniel, R.J. (ed). Liverpool: University Press of Liverpool, pp. 177-192.
- Redfield, A.C.** (1963) The influence of organisms on the composition of sea-water. In *The Sea*. Hill, M. (ed). New-York: Interscience, pp. 26-77.

- Reuter, W., Wiegand, G., Huber, R., and Than, M.E. (1999) Structural analysis at 2.2 Å of orthorhombic crystals presents the asymmetry of the allophycocyanin-linker complex, APC-L<sub>C7</sub>, from phycobilisomes of *Mastigocladus laminosus*. *Proceedings of the National Academy of Sciences of the United States of America* **96**(4): 1363-1368.
- Rhee, K.H., Morris, E.P., Barber, J., and Kühlbrandt, W. (1998) Three-dimensional structure of the plant photosystem II reaction centre at 8 Å resolution. *Nature* **396**(6708): 283-286.
- Richardson, T.L., and Jackson, G.A. (2007) Small phytoplankton and carbon export from the surface ocean. *Science* **315**(5813): 838-840.
- Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annual Reviews of Genetics* **38**: 525-552.
- Rippka, R., Coursin, T., Hess, W., Lichtle, C., Scanlan, D.J., Palinska, K.A., Itean, I., Partensky, F., Houmard, J., and Herdman, M. (2000a) *Prochlorococcus marinus* Chisholm et al. 1992 subsp. *pastoris* subsp. nov strain PCC 9511, the first axenic chlorophyll *a*<sub>2</sub>/*b*<sub>2</sub>-containing cyanobacterium (Oxyphotobacteria). *International Journal of Systematic and Evolutionary Microbiology* **50**: 1833-1847.
- Rippka, R., Coursin, T., Hess, W., Lichtlé, C., Scanlan, D.J., Palinska, K.A., Itean, I., Partensky, F., Houmard, J., and Herdman, M. (2000b) *Prochlorococcus marinus* Chisholm et al. 1992 subsp. *pastoris* subsp. nov strain PCC 9511, the first axenic chlorophyll *a*<sub>2</sub>/*b*<sub>2</sub>-containing cyanobacterium (Oxyphotobacteria). *Intl J Syst Evol Microbiol* **50**: 1833-1847.
- Ritter, S., Hiller, R.G., Wrench, P.M., Welte, W., and Diederichs, K. (1999) Crystal structure of a phycourobilin-containing phycoerythrin at 1.90-Å resolution. *Journal of Structural Biology* **126**(2): 86-97.
- Robertson, B.R., Tezuka, N., and Watanabe, M.M. (2001) Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. *International Journal of Systematic and Evolutionary Microbiology* **51**: 861-871.
- Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology* **68**(3): 1180-1191.
- Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., Johnson, Z.I., Land, M., Lindell, D., Post, A.F., Regala, W., Shah, M., Shaw, S.L., Steglich, C., Sullivan, M.B., Ting, C.S., Tolonen, A.C., Webb, E.A., Zinser, E.R., and Chisholm, S.W. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**(6952): 1042-1047.
- Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R., and Chisholm, S.W. (2009) Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS One* **4**(9): e686.
- Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K., Kent, A.D., Daroub, S.H., Camargo, F.A., Farmerie, W.G., and Triplett, E.W. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal* **1**(4): 283-290.
- Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., and Minor, C. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* **66**(6): 2541-2547.
- Roose, J.L., Wegener, K.M., and Pakrasi, H.B. (2007) The extrinsic proteins of photosystem II. *Photosynthesis Research* **92**(3): 369-387.
- Rose, J.M., and Caron, D.A. (2007) Does low temperature constrain the growth rates of heterotrophic protists? Evidence and implications for algal blooms in cold waters. *Limnology and Oceanography* **52**(2): 886-895.
- Rosinski, J., Hainfeld, J., Rigbi, M., and Siegelman, H. (1981) Phycobilisome ultrastructure and chromatic adaptation in *Fremyella diplosiphon*. *Annals of Botany* **47**(1): 1-12.
- Rothschild, L.J., and Mancinelli, R.L. (2001) Life in extreme environments. *Nature* **409**(6823): 1092-1101.
- Roy, S., Llewellyn, C., Egeland, E.S., and Johnsen, G. (2011) *Phytoplankton pigments: characterization, chemotaxonomy and applications in oceanography*: Cambridge University Press.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., and Remington, K. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology* **5**(3): e77.
- Rusch, D.B., Martiny, A.C., Dupont, C.L., Halpern, A.L., and Venter, J.C. (2010) Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proceedings of the National Academy of Sciences of the United States of America* **107**(37): 16184-16189.
- Rusch, D.B., Lombardo, M.-J., Yee-Greenbaum, J., Novotny, M., Brinkac, L.M., Lasken, R.S., and Dupont, C.L. (2013) Draft genome sequence of a single cell of SAR86 clade subgroup IIIa. *Genome Announcements* **1**(1).
- Ryan-Keogh, T.J., Macey, A.I., Cockshutt, A.M., Moore, C.M., and Bibby, T.S. (2012) The cyanobacterial chlorophyll-binding protein IsiA acts to increase the in

vivo effecton absorption cross-section of PSI under iron limitation. *Journal of Phycology* **48**(1): 145-154.

**Ryther, J.H.** (1956) Photosynthesis in the ocean as a function of light intensity. *Limnology and Oceanography* **1**(1): 61-70.

---

## S

**Sabehi, G., Massana, R., Bielawski, J.P., Rosenberg, M., DeLong, E.F., and Bèjà, O.** (2003) Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environmental Microbiology* **5**(10): 842-849.

**Sabehi, G., Bèjà, O., Suzuki, M.T., Preston, C.M., and DeLong, E.F.** (2004) Different SAR86 subgroups harbour divergent proteorhodopsins. *Environmental Microbiology* **6**(9): 903-910.

**Saito, M.A., Rocap, G., and Moffett, J.W.** (2005) Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnology and Oceanography* **50**(1): 279-290.

**Samuel, A.D.T., Petersen, J.D., and Reese, T.S.** (2001) Envelope structure of *Synechococcus* sp. WH8113, a nonflagellated swimming cyanobacterium. *BMC Microbiology* **1**(1): 4.

**Sánchez-Baracaldo, P., Hayes, P.K., and Blank, C.E.** (2005) Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* **3**(3): 145-165.

**Sandström, S., Ivanov, A.G., Park, Y.I., Öquist, G., and Gustafsson, P.** (2002) Iron stress responses in the cyanobacterium *Synechococcus* sp. PCC7942. *Physiologia Plantarum* **116**(2): 255-263.

**Sanger, F., Nicklen, S., and Coulson, A.R.** (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**(12): 5463-5467.

**Sarcina, M., and Mullineaux, C.W.** (2004) Mobility of the IsiA chlorophyll-binding protein in cyanobacterial thylakoid membranes. *Journal of Biological Chemistry* **279**(35): 36514-36518.

**Sato, M., Yoshikawa, T., Takeda, S., and Furuya, K.** (2007) Application of the size-fractionation method to simultaneous estimation of clearance rates by heterotrophic flagellates and ciliates of pico- and nanophytoplankton. *Journal of Experimental Marine Biology and Ecology* **349**(2): 334-343.

**Satoh, S., and Tanaka, A.** (2006) Identification of chlorophyllide *a* oxygenase in the *Prochlorococcus* genome by a comparative genomic approach. *Plant and Cell Physiology* **47**(12): 1622-1629.

**Saunée, N.A., Williams, S.R., Bryant, D.A., and Schluchter, W.M.** (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. PCC7002. *Journal of Biological Chemistry* **283**(12): 7513-7522.

**Scanlan, D.J., and West, N.J.** (2002) Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiology Ecology* **40**: 1-12.

**Scanlan, D.J.** (2003) Physiological diversity and niche adaptation in marine *Synechococcus*. *Advances in Microbial Physiology Volume* **47**: 1-64.

**Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., Post, A.F., Hagemann, M., Paulsen, I., and Partensky, F.** (2009) Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews* **73**(2): 249-299.

**Schadt, E.E., Turner, S., and Kasarskis, A.** (2010) A window into third-generation sequencing. *Human molecular genetics* **19**(R2): R227-R240.

**Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J.** (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology* **11**(8): 2078-2093.

**Scheer, H., and Zhao, K.H.** (2008) Biliprotein maturation: the chromophore attachment. *Molecular Microbiology* **68**(2): 263-276.

**Scheller, H.V., Naver, H., and Møller, B.L.** (1997) Molecular aspects of photosystem I. *Physiologia Plantarum* **100**(4): 842-851.

**Schirmer, T., Bode, W., Huber, R., Sidler, W., and Zuber, H.** (1985) X-ray crystallographic structure of the light-harvesting biliprotein C-phycocyanin from the thermophilic cyanobacterium *Mastigocladus laminosus* and its resemblance to globin structures. *Journal of Molecular Biology* **184**(2): 257-277.

**Schirmer, T., Huber, R., Schneider, M., Bode, W., Miller, M., and Hackert, M.L.** (1986) Crystal structure analysis and refinement at 2.5 Å of hexameric C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*: The molecular model and its implications for light-harvesting. *Journal of Molecular Biology* **188**(4): 651-676.

**Schleif, R.** (2000) Regulation of the l-arabinose operon of *Escherichia coli*. *Trends in Genetics* **16**(12): 559-565.

**Schleif, R.** (2010) AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiology Reviews* **34**(5): 779-796.

**Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.-Y., and Swanson, R.V.** (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *Journal of Bacteriology* **180**(19): 5003-5009.

- Schloss, P.D., and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* **14**(3): 303-310.
- Schloss, P.D., and Handelsman, J. (2004) Status of the microbial census. *Microbiology and Molecular Biology Reviews* **68**(4): 686-691.
- Schloss, P.D., Gevers, D., and Westcott, S.L. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* **6**(12): e27310.
- Schluchter, W.M., and Glazer, A.N. (1999) Biosynthesis of phycobiliproteins in cyanobacteria. In *The phototrophic prokaryotes*. Peschek, G.A., Löffelhardt, W., and Schmeterer, G. (eds). New-York: Kluwer/Plenum Pres, pp. 83-95.
- Schluchter, W.M., Shen, G., Alvey, R.M., Biswas, A., Saunée, N.A., Williams, S.R., Mille, C.A., and Bryant, D.A. (2010) Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification. In *Recent Advances in Phototrophic Prokaryotes*: Springer New York, pp. 211-228.
- Schmidt, M., Patel, A., Zhao, Y., and Reuter, W. (2007) Structural basis for the photochemistry of  $\alpha$ -phycoerythrocyanin. *Biochemistry* **46**(2): 416-423.
- Schmidt, T.M., DeLong, E.F., and N.R., P. (1991) Analysis of a marine picoplankton community by 16S rRNA gene and sequencing. *Journal of Bacteriology* **173**(14): 4371-4378.
- Schmitz, O., Katayama, M., Williams, S.B., Kondo, T., and Golden, S.S. (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science Signalling* **289**(5480): 765.
- Schneider, S., and Jürgens, U.J. (1991) Cell wall and sheath constituents of the cyanobacterium *Gloeobacter violaceus*. *Archives of Microbiology* **156**(4): 312-318.
- Schnitzer, M. (1978) Some observations on chemistry of humic substances. *Agrochimica* **22**(3-4): 216-225.
- Schopf, J.W., and Packer, B.M. (1987) Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* **237**(4810): 70-73.
- Schopf, J.W. (1993) Microfossils of the early archean apex chert - new evidence of the antiquity of life. *Science* **260**(5108): 640-646.
- Schultze-Lam, S., and Beveridge, T.J. (1994) Physicochemical characteristics of the mineral-forming S-layer from the cyanobacterium *Synechococcus* strain GL24. *Canadian Journal of Microbiology* **40**(3): 216-223.
- Seenivasan, R., Sausen, N., Medlin, L.K., and Melkonian, M. (2013) *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as 'Picobiliphytes'. *PLoS one* **8**(3): e59565.
- Seib, L.O., and Kehoe, D.M. (2002) A turquoise mutant genetically separates expression of genes encoding phycoerythrin and its associated linker peptides. *Journal of Bacteriology* **184**(4): 962-970.
- Seidler, A. (1996) The extrinsic polypeptides of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1277**: 35-60.
- Shalapyonok, A., Olson, R.J., and Shalapyonok, L.S. (1998) Ultradian growth in *Prochlorococcus* spp. *Applied and Environmental Microbiology* **64**(3): 1066-1069.
- Shan, J., Jia, Y., Clokie, M.R.J., and Mann, N.H. (2008) Infection by the 'photosynthetic' phage S-PM2 induces increased synthesis of phycoerythrin in *Synechococcus* sp. WH7803. *FEMS Microbiology Letters* **283**(2): 154-161.
- Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., Polz, M.F., and Alm, E.J. (2012) Population genomics of early events in the ecological differentiation of bacteria. *science* **336**(6077): 48-51.
- Shapiro, L.P., and Haugen, E.M. (1988) Seasonal distribution and temperature tolerance of *Synechococcus* in Boothbay Harbor, Maine. *Estuarine, Coastal and Shelf Science* **26**(5): 517-525.
- Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R.Y., Partensky, F., Koonin, E.V., Wolf, Y.I., Nelson, N., and Beja, O. (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461**(7261): 258-262.
- Shen, G., Saunée, N.A., Williams, S.R., Gallo, E.F., Schluchter, W.M., and Bryant, D.A. (2006) Identification and characterization of a new class of bilin lyase the *cpcT* gene encodes a bilin lyase responsible for attachment of phycocyanobilin to Cys-153 on the  $\beta$ -subunit of phycocyanin in *Synechococcus* sp. PCC7002. *Journal of Biological Chemistry* **281**(26): 17768-17778.
- Shen, G., Schluchter, W.M., and Bryant, D.A. (2008) Biogenesis of Phycobiliproteins. I. *cpcS-I* and *cpcU* mutants of the cyanobacterium *Synechococcus* sp. PCC7002 define a heterodimeric phycocyanobilin lyase specific for  $\beta$ -phycocyanin and allophycocyanin subunits. *Journal of Biological Chemistry* **283**(12): 7503-7512.
- Sherr, E., and Sherr, B. (2002) Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* **81**(1): 293-308.
- Shi, F. (2011) The coevolution of phycobilisomes: molecular structure adapting to functional evolution. *Comparative and Functional Genomics* **2011**.
- Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., de Marsac, N.T., and Rippka, R.

- (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **110**(3): 1053-1058.
- Shikata, T., Nukata, A., Yoshikawa, S., Matsubara, T., Yamasaki, Y., Shimasaki, Y., Oshima, Y., and Honjo, T.** (2009) Effects of light quality on initiation and development of meroplanktonic diatom blooms in a eutrophic shallow sea. *Marine Biology* **156**(5): 875-889.
- Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M.** (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the United States of America* **89**(18): 8794-8797.
- Shokralla, S., Spall, J.L., Gibson, J.F., and Hajibabaei, M.** (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**(8): 1794-1805.
- Shukla, A., Biswas, A., Blot, N., Partensky, F., Karty, J.A., Hammad, L.A., Garczarek, L., Gutu, A., Schluchter, W.M., and Kehoe, D.M.** (2012) Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America*.
- Sidler, W., Nutt, H., Kumpf, B., Frank, B., Suter, F., Brenzel, A., Wermeyer, W., and Zuber, H.** (1990) The complete amino-acid sequence and the phylogenetic origin of phycocyanin-645 from the cryptophyten alga *Chroomonas* sp. *Biological Chemistry Hoppe-Seyler* **371**(2): 537-548.
- Sidler, W.A.** (1994) Phycobilisome and phycobiliprotein structures. In *The molecular biology of cyanobacteria*. Bryant, D.A. (ed). Netherlands: Kluwer Academic Publishers, pp. 139-216.
- Sieburth, J.M., Smetacek, V., and Lenz, J.** (1978) Pelagic ecosystem structure - heterotrophic compartments of plankton and their relationship to plankton size fractions. *Limnology and Oceanography* **23**(6): 1256-1263.
- Siegelman, H.W., and Kycia, J.H.** (1982) Molecular morphology of cyanobacterial phycobilisomes. *Plant Physiology* **70**(3): 887-897.
- Sieracki, M.E., Haugen, E.M., and Cucci, T.L.** (1995) Overestimation of heterotrophic bacteria in the Sargasso Sea: direct evidence by flow and imaging cytometry. *Deep Sea Research Part I: Oceanographic Research Papers* **42**(8): 1399-1409.
- Singh, A.K., and Sherman, L.A.** (2007) Reflections on the function of IsiA, a cyanobacterial stress-inducible, Chl-binding protein. *Photosynthesis Research* **93**(1): 17-25.
- Six, C., Thomas, J., Brahamsha, B., Lemoine, Y., and Partensky, F.** (2004) Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquatic Microbial Ecology* **35**(1): 17-29.
- Six, C., Thomas, J.C., Thion, L., Lemoine, Y., Zal, F., and Partensky, F.** (2005) Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. *Journal of Bacteriology* **187**(5): 1685-1694.
- Six, C., Joubin, L., Partensky, F., Holtzendorff, J., and Garczarek, L.** (2007a) UV-induced phycobilisome dismantling in the marine picocyanobacterium *Synechococcus* sp. WH8102. *Photosynthesis Research* **92**(1): 75-86.
- Six, C., Thomas, J.C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., Scanlan, D.J., and Partensky, F.** (2007b) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biology* **8**(12): R259.
- Šmarda, J., Šmajš, D., Komrska, J., and Krzyžánek, V.** (2002) S-layers on cell walls of cyanobacteria. *Micron* **33**(3): 257-277.
- Smit, E., Leeftang, P., and Wernars, K.** (2006) Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. *FEMS Microbiology Ecology* **23**(3): 249-261.
- Smith, R.C., and Baker, K.S.** (1981) Optical properties of the clearest natural waters (200-800 nm). *Applied Optics* **20**(2): 177-184.
- Snyder, D.S., Brahamsha, B., Azadi, P., and Palenik, B.** (2009) Structure of compositionally simple lipopolysaccharide from marine *Synechococcus*. *Journal of Bacteriology* **191**(17): 5499-5509.
- Sobczyk, A., Bely, A., Demarsac, N.T., and Houmard, J.** (1994) A phosphorylated DNA-binding protein is specific for the red-light signal during complementary chromatic adaptation in cyanobacteria. *Molecular Microbiology* **13**(5): 875-885.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J.** (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**(32): 12115-12120.
- Sohrin, R., Isaji, M., Obara, Y., Agostini, S., Suzuki, Y., Hiroe, Y., Ichikawa, T., and Hidaka, K.** (2011) Distribution of *Synechococcus* in the dark ocean. *Aquatic Microbial Ecology* **64**(1): 1.
- Somerville, C.C., Knight, I.T., Straube, W.L., and Colwell, R.R.** (1989) Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Applied and Environmental Microbiology* **55**(3): 548-554.

- Stahl, D.A., Lane, D.J., Olsen, G.J., and Pace, N.R.** (1984) Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science* **224**(4647): 409-411.
- Staley, J.T., and Konopka, A.** (1985) Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology* **39**(1): 321-346.
- Stanier, R.Y., Kunisawa, R., Mandel, M., and Cohenbaz, G.** (1971) Purification and properties of unicellular blue green algae (ordre Chroococcales). *Bacteriological Reviews* **35**(2): 171-205.
- Stanier, R.Y., and Cohen-Bazire, G.** (1977) Phototrophic prokaryotes: the cyanobacteria. *Annual Review of Microbiology* **31**: 225-274.
- Steglich, C., Mullineaux, C.W., Teuchner, K., Hess, W.R., and Lokstein, H.** (2003a) Photophysical properties of *Prochlorococcus marinus* SS120 divinyl chlorophylls and phycoerythrin *in vitro* and *in vivo*. *FEBS Letters* **553**(1): 79-84.
- Steglich, C., Post, A.F., and Hess, W.R.** (2003b) Analysis of natural populations of *Prochlorococcus* spp. in the northern Red Sea using phycoerythrin gene sequences. *Environmental Microbiology* **5**(8): 681-690.
- Steglich, C., Frankenberg-Dinkel, N., Penno, S., and Hess, W.R.** (2005) A green light-absorbing phycoerythrin is present in the high-light-adapted marine cyanobacterium *Prochlorococcus* sp. MED4. *Environmental Microbiology* **7**(10): 1611-1618.
- Steglich, C., Futschik, M.E., Lindell, D., Voss, B., Chisholm, S.W., and Hess, W.R.** (2008) The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genetics* **4**(8): e1000173.
- Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F.** (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* **178**(3): 591-599.
- Steinberg, D.K., Carlson, C.A., Bates, N.R., Johnson, R.J., Michaels, A.F., and Knap, A.H.** (2001) Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Research Part II: Topical Studies in Oceanography* **48**(8): 1405-1447.
- Stepanuskas, R., and Sieracki, M.E.** (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences of the United States of America* **104**(21): 9052-9057.
- Stepanuskas, R.** (2012) Single cell genomics: an individual look at microbes. *Current Opinion in Microbiology* **15**(5): 613-620.
- Stickforth, P., Steiger, S., Hess, W.R., and Sandmann, G.** (2003) A novel type of lycopene epsilon-cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. *Archives of Microbiology* **179**: 409-415.
- Stomp, M., Huisman, J., De Jongh, F., Veraart, A.J., Gerla, D., Rijkeboer, M., Ibelings, B.W., Wollenzien, U.I., and Stal, L.J.** (2004) Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* **432**(7013): 104-107.
- Stomp, M., Huisman, J., Stal, L.J., and Matthijs, H.C.P.** (2007) Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. *The ISME Journal* **1**(4): 271-282.
- Stomp, M.** (2008) Colourful coexistence : a new solution to the plankton paradox. In. University of Amsterdam, p. 137.
- Stomp, M., van Dijk, M.A., van Overzee, H.M.J., Wortel, M.T., Sigon, C.A.M., Egas, M., Hoogveld, H., Gons, H.J., and Huisman, J.** (2008) The timescale of phenotypic plasticity and its impact on competition in fluctuating environments. *The American Naturalist* **172**(5): E169-E185.
- Storf, M., Parbel, A., Meyer, M., Strohmam, B., Scheer, H., Deng, M.G., Zheng, M., Zhou, M., and Zhao, K.H.** (2001) Chromophore attachment to biliproteins: specificity of PecE/PecF, a lyase-isomerase for the photoactive 3(1)-Cys- $\alpha$ -84-phycoviolobilin chromophore of phycoerythrocyanin. *Biochemistry* **40**(41): 12444-12456.
- Stowe-Evans, E.L., Ford, J., and Kehoe, D.M.** (2004) Genomic DNA microarray analysis: identification of new genes regulated by light color in the cyanobacterium *Fremyella diplosiphon*. *Journal of Bacteriology* **186**(13): 4338-4349.
- Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W.** (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**(6952): 1047-1051.
- Sullivan, M.B., Coleman, M.L., Weigle, P., Rohwer, F., and Chisholm, S.W.** (2005) Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biology* **3**(5): e144.
- Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., Stocks, K., Allen, E.E., Ellisman, M., and Grethe, J.** (2011) Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Research* **39**(suppl 1): D546-D551.
- Suttle, C.A., Chan, A.M., and Cottrell, M.T.** (1990) Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **347**(6292): 467-469.
- Suttle, C.A., and Chan, A.M.** (1993) Marine cyanophages infecting oceanic and coastal strains of *Synechococcus* - abundance, morphology, cross-infectivity and growth characteristics. *Marine Ecology-Progress Series* **92**(1-2): 99-109.

- Suttle, C.A., and Chan, A.M.** (1994) Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Applied and Environmental Microbiology* **60**(9): 3167-3174.
- Suttle, C.A.** (2005) Viruses in the sea. *Nature* **437**(7057): 356-361.
- Suttle, C.A.** (2007) Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5**(10): 801-812.
- Suzuki, E., Ohkawa, H., Moriya, K., Matsubara, T., Nagaïke, Y., Iwasaki, I., Fujiwara, S., Tsuzuki, M., and Nakamura, Y.** (2010) Carbohydrate metabolism in mutants of the cyanobacterium *Synechococcus elongatus* PCC 7942 defective in glycogen synthesis. *Applied and Environmental Microbiology* **76**(10): 3153-3159.
- Swan, B.K., Martinez-Garcia, M., Preston, C.M., Szczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N.J., Masland, E.D.P., and Gomez, M.L.** (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**(6047): 1296-1300.
- Swanson, R., De Lorimier, R., and Glazer, A.** (1992) Genes encoding the phycobilisome rod substructure are clustered on the *Anabaena* chromosome: characterization of the phycoerythrocyanin operon. *Journal of Bacteriology* **174**(8): 2640-2647.
- Swanson, R.V., Ong, L.J., Wilbanks, S.M., and Glazer, A.N.** (1991) Phycoerythrins of marine unicellular cyanobacteria. II. Characterization of phycobiliproteins with unusually high phycourobilin content. *Journal of Biological Chemistry* **266**(15): 9528-9534.
- Swingley, W.D., Chen, M., Cheung, P.C., Conrad, A.L., Dejesa, L.C., Hao, J., Honchak, B.M., Karbach, L.E., Kurdoglu, A., and Lahiri, S.** (2008) Niche adaptation and genome expansion in the chlorophyll *d*-producing cyanobacterium *Acaryochloris marina*. *Proceedings of the National Academy of Sciences of the United States of America* **105**(6): 2005-2010.
- 
- T**
- Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R., and Phillips, R.** (2011) Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* **333**(6038): 58-62.
- Tai, V., and Palenik, B.** (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *The ISME Journal* **3**(8): 903-915.
- Tai, V., Burton, R.S., and Palenik, B.** (2011) Temporal and spatial distributions of marine *Synechococcus* in the Southern California Bight assessed by hybridization to bead-arrays. *Marine Ecology-Progress Series* **426**: 133-164.
- Takano, H., and Gusella, J.** (2002) The predominantly HEAT-like motif structure of huntingtin and its association and coincident nuclear entry with dorsal, an NF- $\kappa$ B/Rel/dorsal family transcription factor. *BMC Neuroscience* **3**(1): 15.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.** (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**(10): 2731-2739.
- Tandeau de Marsac, N.** (1977) Occurrence and nature of chromatic adaptation in Cyanobacteria. *Journal of Bacteriology* **130**(1): 82-91.
- Tandeau de Marsac, N., and Cohen-Bazire, G.** (1977) Molecular composition of cyanobacterial phycobilisomes. *Proceedings of the National Academy of Sciences of the United States of America* **74**(4): 1635-1639.
- Tandeau de Marsac, N., Castets, A.M., and Cohen-Bazire, G.** (1980) Wavelength modulation of phycoerythrin synthesis in *Synechocystis* sp. 6701. *Journal of Bacteriology* **142**(1): 310-314.
- Tandeau de Marsac, N.** (1983) Phycobilisomes and complementary chromatic adaptation in cyanobacteria. *Bulletin de L'Institut Pasteur* **81**(3): 201-254.
- Taylor, F., Hoppenrath, M., and Saldarriaga, J.** (2008) *Dinoflagellate diversity and distribution*: Springer Netherlands.
- Teale, F., and Dale, R.** (1970) Isolation and spectral characterization of phycobiliproteins. *Biochemical Journal* **116**(2): 161-169.
- Telenius, H., Carter, N.P., Bebb, C.E., Ponder, B.A., and Tunnacliffe, A.** (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**(3): 718-725.
- Telfer, A.** (2005) Too much light? How  $\beta$ -carotene protects the photosystem II reaction centre. *Photochemical & Photobiological Sciences* **4**(12): 950-956.
- Temperton, B., Field, D., Oliver, A., Tiwari, B., Mühling, M., Joint, I., and Gilbert, J.A.** (2009) Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *The ISME Journal* **3**(7): 792-796.
- Terauchi, K., Montgomery, B.L., Grossman, A.R., Lagarias, J.C., and Kehoe, D.M.** (2004) RcaE is a complementary chromatic adaptation photoreceptor required for green and red light responsiveness. *Molecular Microbiology* **51**(2): 567-577.
- Terry, M.J., McDowell, M.T., and Lagarias, J.C.** (1995) (3Z)- and (3E)-phytychromobilin are intermediates in the

- biosynthesis of the phytochrome chromophore. *Journal of Biological Chemistry* **270**(19): 11111-11118.
- Tettelin, H., Riley, D., Cattuto, C., and Medini, D.** (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology* **11**(5): 472-477.
- Tewhey, R., Warner, J.B., Nakano, M., Libby, B., Medkova, M., David, P.H., Kotsopoulos, S.K., Samuels, M.L., Hutchison, J.B., and Larson, J.W.** (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nature Biotechnology* **27**(11): 1025-1031.
- Thingstad, T.F., Krom, M.D., Mantoura, R.F.C., Flaten, G.A.F., Groom, S., Herut, B., Kress, N., Law, C.S., Pasternak, A., Pitta, P., Psarra, S., Rassoulzadegan, F., Tanaka, T., Tselepidis, A., Wassmann, P., Woodward, E.M.S., Riser, C.W., Zodiatis, G., and Zohary, T.** (2005) Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. *Science* **309**(5737): 1068-1071.
- Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaulot, D., Kuypers, M.M., and Zehr, J.P.** (2012) Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337**(6101): 1546-1550.
- Thorsen, T., Roberts, R.W., Arnold, F.H., and Quake, S.R.** (2001) Dynamic pattern formation in a vesicle-generating microfluidic device. *Physical Review Letters* **86**(18): 4163-4166.
- Ting, C.S., Rocap, G., King, J., and Chisholm, S.W.** (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends in Microbiology* **10**(3): 134-142.
- Ting, C.S., Hsieh, C., Sundararaman, S., Mannella, C., and Marko, M.** (2007) Cryo-electron tomography reveals the comparative three-dimensional architecture of *Prochlorococcus*, a globally important marine cyanobacterium. *Journal of Bacteriology* **189**(12): 4485-4493.
- Toledo, G., Palenik, B., and Brahmsha, B.** (1999) Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Applied and Environmental Microbiology* **65**(12): 5247-5251.
- Toledo, G., and Palenik, B.** (2003) A *Synechococcus* serotype is found preferentially in surface marine waters. *Limnology and Oceanography* **48**(5): 1744-1755.
- Toole, C.M., and Alnutt, F.C.T.** (2003) Red, cryptomonad and glaucocystophyte algal phycobiliproteins. In *Photosynthesis in algae*. Larkum, A.W.D., Douglas, S.E., and Raven, J.A. (eds). Dordrecht/Boston/London: Kluwer Academic.
- Treusch, A.H., Kletzin, A., Raddatz, G., Ochsenreiter, T., Quaiser, A., Meurer, G., Schuster, S.C., and Schleper, C.** (2004) Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environmental Microbiology* **6**(9): 970-980.
- Tringe, S.G., and Rubin, E.M.** (2005) Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics* **6**(11): 805-814.
- Tripp, H.J., Bench, S.R., Turk, K.A., Foster, R.A., Desany, B.A., Niazi, F., Affourtit, J.P., and Zehr, J.P.** (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464**(7285): 90-94.
- Tu, J.-M.** (2008) Phycobiliprotein lyases : Structure of reconstitution products and mechanistic studies. In. Munich: University of Munich, p. 183.
- Tu, J.M., Kupka, M., Böhm, S., Plöschner, M., Eichacker, L., Zhao, K.H., and Scheer, H.** (2008) Intermediate binding of phycocyanobilin to the lyase, CpeS1, and transfer to apoprotein. *Photosynthesis Research* **95**(2): 163-168.
- Tu, S.L., Gunn, A., Toney, M.D., Britt, R.D., and Lagarias, J.C.** (2004) Biliverdin reduction by cyanobacterial phycocyanobilin: ferredoxin oxidoreductase (PcyA) proceeds via linear tetrapyrrole radical intermediates. *Journal of the American Chemical Society* **126**(28): 8682-8693.
- Turkarlan, S., Sanders, C., and Daldal, F.** (2006) Extracytoplasmic prosthetic group ligation to apoproteins: maturation of c-type cytochromes. *Molecular Microbiology* **60**(3): 537-541.
- Turner, S., Pryer, K.M., Miao, V.P.W., and Palmer, J.D.** (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryotic Microbiology* **46**(4): 327-338.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F.** (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**(6978): 37-43.

U

- Umena, Y., Kawakami, K., Shen, J.-R., and Kamiya, N.** (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**(7345): 55-60.
- Urbach, E., Scanlan, D.J., Distel, D.L., Waterbury, J.B., and Chisholm, S.W.** (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *Journal of Molecular Evolution* **46**(2): 188-201.
- Ussery, D.W., and Hallin, P.F.** (2004) Genome Update: AT content in sequenced prokaryotic genomes. *Microbiology* **150**(4): 749-752.

V



- van Niel, C.B. (1955) The microbe as a whole. In *Perspectives and Horizons in Microbiology*. Waksman, S.A. (ed). New Brunswick: Rutgers University Press, p. 202.
- VanInsberghe, D., Hartmann, M., Stewart, G.R., and Mohn, W.W. (2013) Isolation of a substantial proportion of forest soil bacterial communities detected via pyrotag sequencing. *Applied and Environmental Microbiology* **79**(6): 2096-2098.
- Vass, I., Styring, S., Hundal, T., Koivuniemi, A., Aro, E., and Andersson, B. (1992) Reversible and irreversible intermediates during photoinhibition of photosystem II: stable reduced QA species promote chlorophyll triplet formation. *Proceedings of the National Academy of Sciences of the United States of America* **89**(4): 1408-1412.
- Vaulot, D. (1989) CYTOPC : Processing software for flow cytometric data. *Signal and Noise* **2**(8).
- Vaulot, D., Marie, D., Olson, R.J., and Chisholm, S.W. (1995) Growth of *Prochlorococcus*, a photosynthetic prokaryote, in the equatorial Pacific Ocean. *Science* **268**(5216): 1480-1482.
- Vaulot, D., Eikrem, W., Viprey, M., and Moreau, H. (2008) The diversity of small eukaryotic phytoplankton (< or =3 µm) in marine ecosystems. *FEMS Microbiology Reviews* **32**(5): 795-820.
- Vaulot, D., Lepère, C., Toulza, E., De la Iglesia, R., Poulain, J., Gaboyer, F., Moreau, H., Vandepoele, K., Ulloa, O., and Gavory, F. (2012) Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. *PLoS One* **7**(6): e39648.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D.Y., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., and Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**(5667): 66-74.
- Vergin, K.L., Tripp, H.J., Wilhelm, L.J., Denver, D.R., Rappé, M.S., and Giovannoni, S.J. (2007) High intraspecific recombination rate in a native population of *Candidatus Pelagibacter ubique* (SAR11). *Environmental Microbiology* **9**(10): 2430-2440.
- Vilibić, I., and Šantić, D. (2008) Deep water ventilation traced by *Synechococcus* cyanobacteria. *Ocean Dynamics* **58**(2): 119-125.
- Vogelmann, T.C., and Scheibe, J. (1978) Action spectra for chromatic adaptation in the blue-green alga *Fremyella diplosiphon*. *Planta* **143**(3): 233-239.
- Von Mering, C., Hugenholtz, P., Raes, J., Tringe, S., Doerks, T., Jensen, L., Ward, N., and Bork, P. (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**(5815): 1126-1130.
- Wagner, J.R., Brunzelle, J.S., Forest, K.T., and Vierstra, R.D. (2005) A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**(7066): 325-331.
- Waleron, M., Waleron, K., Vincent, W.F., and Wilmotte, A. (2007) Allochthonous inputs of riverine picocyanobacteria to coastal waters in the Arctic Ocean. *FEMS Microbiology Ecology* **59**(2): 356-365.
- Wang, K., and Chen, F. (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. *Environmental Microbiology* **10**(2): 300-312.
- Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979) Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277**(5694): 293-294.
- Waterbury, J.B., Willey, J.M., Franks, D.G., Valois, F.W., and Watson, S.W. (1985) A cyanobacterium capable of swimming motility. *Science* **230**(4721): 74-76.
- Waterbury, J.B., Watson, S.W., Valois, F.W., and Franks, D.G. (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Canadian Bulletin of Fisheries and Aquatic Sciences* **214**: 71-120.
- Waterbury, J.B., and Valois, F.W. (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Applied and Environmental Microbiology* **59**(10): 3393-3399.
- Wauthy, B., Desrosières, R., and Le Bourhis, J. (1967) Importance présumée de l'ultraplancton dans les eaux tropicales oligotrophes du Pacifique central sud. *Cahier ORSTOM Série Océanographique* **5**: 109-116.
- Wawrick, B., Paul, J.H., Campbell, L., Griffin, D., Houchin, L., Fuentes-Ortega, A., and Muller-Karger, F. (2003) Vertical structure of the phytoplankton community associated with the coastal plume of the Gulf of Mexico. *Marine Ecology Progress Series* **251**: 87-101.
- Wehrmeyer, W. (1983) *Organization and composition of cyanobacterial and rhodophycean phycobilisomes [Algae, structure and biochemistry]*. New-York.
- Wehrmeyer, W., Zimmermann, C., Ohki, K., and Fujita, Y. (1988) On a new type of cyanobacterial phycobilisome exemplified in *Phormidium persicinum* Gomont. *European Journal of Cell Biology* **46**(3): 539-546.
- Weigle, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., King, J., Hatfull, G.F., Lawrence, J.G., and Hendrix, R.W. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting

- marine *Prochlorococcus* and *Synechococcus*. *Environmental Microbiology* **9**(7): 1675-1695.
- Weinbauer, M.G., Bonilla-Findji, O., Chan, A.M., Dolan, J.R., Short, S.M., Šimek, K., Wilhelm, S.W., and Suttle, C.A.** (2011) *Synechococcus* growth in the ocean may depend on the lysis of heterotrophic bacteria. *Journal of Plankton Research* **33**(10): 1465-1476.
- Welsh, E.A., Liberton, M., Stöckel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R.S., Clifton, S.W., and Jacobs, J.M.** (2008) The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proceedings of the National Academy of Sciences of the United States of America* **105**(39): 15094-15099.
- Wendler, J., John, W., Scheer, H., and Hoewartha, A.** (1986) Energy transfer in trimeric C-phycoyanin studied by picosecond fluorescence kinetics. *Photochemistry and Photobiology* **44**(1): 79-85.
- West, N.J., Schönhuber, W.A., Fuller, N.J., Amann, R.L., Rippka, R., Post, A.F., and Scanlan, D.J.** (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147**(7): 1731-1744.
- West, N.J., Lebaron, P., Stratton, P.G., and Suzuki, M.T.** (2010) A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *The ISME Journal* **5**(6): 933-944.
- Westermann, M., Reuter, W., Schimek, C., and Wehrmeyer, W.** (1993) Presence of both hemidiscoidal and hemiellipsoidal physobilisomes in a *Phormidium* species (Cyanobacteria). *Zeitschrift für Naturforschung* **48**: 28-34.
- Westermann, M., and Wehrmeyer, W.** (1995) A new type of complementary chromatic adaptation exemplified by *Phormidium* sp. C86: changes in the number of peripheral rods and in the stoichiometry of core complexes in phycobilisomes. *Archives of Microbiology* **164**(2): 132-141.
- Wetterstrand, K.A.** (2013). DNA sequencing costs : Data from the NHGRI Genome Sequencing Program (GSP). Accessed February 2013. URL [www.genome.gov/sequencingcosts](http://www.genome.gov/sequencingcosts)
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J.** (1998) Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* **95**(12): 6578-6583.
- Wickstrum, J.R., Skredenske, J.M., Kolin, A., Jin, D.J., Fang, J., and Egan, S.M.** (2007) Transcription activation by the DNA-binding domain of the AraC family protein RhaS in the absence of its effector-binding domain. *Journal of Bacteriology* **189**(14): 4984-4993.
- Wickstrum, J.R., Skredenske, J.M., Balasubramaniam, V., Jones, K., and Egan, S.M.** (2010) The AraC/XylS family activator RhaS negatively autoregulates *rhaSR* expression by preventing cyclic AMP receptor protein activation. *Journal of Bacteriology* **192**(1): 225-232.
- Wiethaus, J., Busch, A.W.U., Kock, K., Leichert, L.I., Herrmann, C., and Frankenberg-Dinkel, N.** (2010) CpeS is a lyase specific for attachment of 3Z-PEB to Cys82 of  $\beta$ -phycoerythrin from *Prochlorococcus marinus* MED4. *Journal of Biological Chemistry* **285**(48): 37561-37569.
- Wilbanks, S.M., and Glazer, A.N.** (1993a) Rod structure of a phycoerythrin-II containing phycobilisome. 1. Organization and sequence of the gene cluster encoding the major phycobliprotein rod components in the genome of marine *Synechococcus* sp. WH8020. *Journal of Biological Chemistry* **268**(2): 1226-1235.
- Wilbanks, S.M., and Glazer, A.N.** (1993b) Rod structure of a phycoerythrin II-containing phycobilisome. II. Complete sequence and bilin attachment site of a phycoerythrin gamma subunit. *Journal of Biological Chemistry* **268**(2): 1236-1241.
- Wilde, A., Churin, Y., Schubert, H., and Börner, T.** (1997) Disruption of a *Synechocystis* sp. PCC 6803 gene with partial similarity to phytochrome genes alters growth under changing light qualities. *FEBS Letters* **406**(1-2): 89.
- Wilhelm, L.J., Tripp, H.J., Givan, S.A., Smith, D.P., and Giovannoni, S.J.** (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biology Direct* **2**(27): doi:10.1186/1745-6150-1182-1127.
- Wilk, K.E., Harrop, S.J., Jankova, L., Edler, D., Keenan, G., Sharples, F., Hiller, R.G., and Curmi, P.M.G.** (1999) Evolution of a light-harvesting protein by addition of new subunits and rearrangement of conserved elements: Crystal structure of a cryptophyte phycoerythrin at 1.63-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **96**(16): 8901-8906.
- Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., Andrews-Pfannkoch, C., Fadrosh, D., Miller, C.S., and Sutton, G.** (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* **3**(1): e1456.
- Wilmotte, A., Démonceau, C., Goffart, A., Hecq, J.H., Demoulin, V., and Crossley, A.C.** (2002) Molecular and pigment studies of the picophytoplankton in a region of the Southern Ocean (42-54°S, 141-144 ° E) in March 1998. *Deep-Sea Research Part II-Topical Studies in Oceanography* **49**(16): 3351-3363.
- Wilmotte, A.** (2004) Molecular evolution and Taxonomy of the Cyanobacteria. In *The Molecular Biology of Cyanobacteria (Advances in Photosynthesis and Respiration)*. Bryant, D.A. (ed): Springer Netherlands, pp. 1-25.

- Wintzingerode, F.V., Göbel, U.B., and Stackebrandt, E.** (2006) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21**(3): 213-229.
- Wislough, S.** (1924) Przyczynki do biologji i genezy szlanow leczniczych na Krymie. *Acta Societatis Botanicorum Poloniae* **2**: 99-129.
- Woese, C.R., and Fox, G.E.** (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74**(11): 5088-5090.
- Woese, C.R.** (1987) Bacterial evolution. *Microbiological Reviews* **51**(2): 221-271.
- Woese, C.R., Kandler, O., and Wheelis, M.L.** (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**(12): 4576-4579.
- Wood, A.M., Phinney, D.A., and Yentsch, C.S.** (1998) Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Marine Ecology-Progress Series* **162**: 25-31.
- Woyke, T., Xie, G., Copeland, A., Gonzalez, J.M., Han, C., Kiss, H., Saw, J.H., Senin, P., Yang, C., Chatterji, S., Cheng, J.F., Eisen, J.A., Sieracki, M.E., and Stepanauskas, R.** (2009) Assembling the marine metagenome, one cell at a time. *PLoS One* **4**(4): 10.
- Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D.Y., McCutcheon, J.P., McDonald, B.R., Moran, N.A., Bristow, J., and Cheng, J.F.** (2010) One bacterial cell, one complete genome. *PLoS One* **5**(4): e10314.
- Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., Malmstrom, R., Stepanauskas, R., and Cheng, J.-F.** (2011) Decontamination of MDA reagents for single cell Whole Genome Amplification. *PLoS One* **6**(10).
- Wright, T.D., Vergin, K.L., Boyd, P.W., and Giovannoni, S.J.** (1997) A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Applied and Environmental Microbiology* **63**(4): 1441-1448.
- Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., and Tindall, B.J.** (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**(7276): 1056-1060.
- Wu, X.-J., Cang, K., Luo, J., Zhou, M., Scheer, H., and Zhao, K.-H.** (2013) Modular generation of fluorescent phycobiliproteins. *Photochemical & Photobiological Sciences* DOI: 10.1039/C3PP25383J.
- Wyman, M., Gregory, R., and Carr, N.** (1985) Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**(4727): 818.
- Wyman, M.** (1999) Diel rhythms in ribulose-1, 5-bisphosphate carboxylase/oxygenase and glutamine synthetase gene expression in a natural population of marine picoplanktonic cyanobacteria (*Synechococcus* spp.). *Applied and Environmental Microbiology* **65**(8): 3651-3659.
- Wyman, M., and Thom, C.** (2012) Temporal orchestration of *glgA* (encoding glycogen synthase) expression and glycogen accumulation in the oceanic picoplanktonic cyanobacterium *Synechococcus* sp. strain WH8103. *Applied and Environmental Microbiology* **78**(13): 4744-4747.

Y

**Yamanaka, G., Glazer, A., and Williams, R.** (1980) Molecular architecture of a light-harvesting antenna. Comparison of wild type and mutant *Synechococcus* 6301 phycobilisomes. *Journal of Biological Chemistry* **255**(22): 11004-11010.

**Yamazaki, T., Nishimura, Y., Yamazaki, I., Hirano, M., Matsuura, K., Shimada, K., and Mimuro, M.** (1994) Energy migration in allophycocyanin-B trimer with a linker polypeptide: analysis by the principal multi-component spectral estimation (PMSE) method. *FEBS Letters* **353**(1): 43-47.

**Yokono, M., Tomo, T., Nagao, R., Ito, H., Tanaka, A., and Akimoto, S.** (2012) Alterations in photosynthetic pigments and amino acid composition of D1 protein change energy distribution in photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**(5): 754-759.

**Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., Yang, E.C., Duffy, S., and Bhattacharya, D.** (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**(6030): 714-717.

**Yooseph, S., Neelson, K.H., Rusch, D.B., McCrow, J.P., Dupont, C.L., Kim, M., Johnson, J., Montgomery, R., Ferreira, S., and Beeson, K.** (2010) Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**(7320): 60-66.

**Yoshihara, S., and Ikeuchi, M.** (2004) Phototactic motility in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Photochemical and Photobiological Sciences* **3**(6): 512-518.

**Youssef, N.H., Blainey, P.C., Quake, S.R., and Elshahed, M.S.** (2011) Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma). *Applied and Environmental Microbiology* **77**(21): 7804-7814.

**Yu, M., and Glazer, A.N.** (1982) Cyanobacterial phycobilisomes. Role of the linker polypeptides in the assembly of phycocyanin. *Journal of Biological Chemistry* **257**(7): 3429-3433.

**Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., and Béjà, O.** (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environmental Microbiology* **9**(6): 1464-1475.

## Z

**Zeng, Y., Novak, R., Shuga, J., Smith, M.T., and Mathies, R.A.** (2010) High-performance single cell genetic analysis using microfluidic emulsion generator arrays. *Analytical Chemistry* **82**(8): 3183-3190.

**Zhang, D.Y., Brandwein, M., HSUIH, T., and LI, H.B.** (2001) Ramification amplification: a novel isothermal DNA amplification method. *Molecular Diagnosis* **6**(2): 141-150.

**Zhang, J., Chiodini, R., Badr, A., and Zhang, G.** (2011) The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics* **38**(3): 95-109.

**Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M.** (2006) Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology* **24**(6): 680-686.

**Zhao, F., and Qin, S.** (2006) Evolutionary analysis of phycobiliproteins: Implications for their structural and functional relationships. *Journal of Molecular Evolution* **63**(3): 330-340.

**Zhao, K.-H., Wu, D., Zhou, M., Zhang, L., Böhm, S., Bubbenzer, C., and Scheer, H.** (2005a) Amino acid residues associated with enzymatic activities of the isomerizing phycoviolobin-lyase PecE/F. *Biochemistry* **44**(22): 8126-8137.

**Zhao, K.H., Deng, M.G., Zheng, M., Zhou, M., Parbel, A., Storf, M., Meyer, M., Strohmman, B., and Scheer, H.** (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 Letters* **469**(1): 9-13.

**Zhao, K.H., Wu, D., Wang, L., Zhou, M., Storf, M., Bubbenzer, C., Strohmman, B., and Scheer, H.** (2002) Characterization of phycoviolobin phycoerythrocyanin- $\alpha$ 84-cystein-lyase-(isomerizing) from *Mastigocladus laminosus*. *European Journal of Biochemistry* **269**(18): 4542-4550.

**Zhao, K.H., Su, P., Böhm, S., Song, B., Zhou, M., Bubbenzer, C., and Scheer, H.** (2005b) Reconstitution of phycobilisome core-membrane linker,  $L_{CM}$ , by autocatalytic chromophore binding to ApcE. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1706**(1): 81-87.

**Zhao, K.H., Su, P., Li, J., Tu, J.M., Zhou, M., Bubbenzer, C., and Scheer, H.** (2006) Chromophore attachment to phycobiliprotein  $\beta$ -subunits. Phycocyanobilin:cystein-  $\beta$ -

84 phycobiliprotein lyase activity of CpeS-like protein from *Anabaena* sp. PCC7120. *Journal of Biological Chemistry* **281**(13): 8573-8581.

**Zhao, K.H., Su, P., Tu, J.M., Wang, X., Liu, H., Plöschner, M., Eichacker, L., Yang, B., Zhou, M., and Scheer, H.** (2007a) Phycobilin: cystein-84 biliprotein lyase, a near-universal lyase for cystein-84-binding sites in cyanobacterial phycobiliproteins. *Proceedings of the National Academy of Sciences of the United States of America* **104**(36): 14300-14305.

**Zhao, K.H., Zhang, J., Tu, J.M., Böhm, S., Plöschner, M., Eichacker, L., Bubbenzer, C., Scheer, H., Wang, X., and Zhou, M.** (2007b) Lyase activities of CpcS-and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin  $\beta$ -subunits. *Journal of Biological Chemistry* **282**(47): 34093-34103.

**Zhao, Y., Temperton, B., Thrash, J.C., Schwalbach, M.S., Vergin, K.L., Landry, Z.C., Ellisman, M., Deerinck, T., Sullivan, M.B., and Giovannoni, S.J.** (2013) Abundant SAR11 viruses in the ocean. *Nature* **494**(7437): 357-360.

**Zhou, J., Gasparich, G.E., Stirewalt, V.L., De Lorimier, R., and Bryant, D.** (1992) The *cpcE* and *cpcF* genes of *Synechococcus* sp. PCC 7002. Construction and phenotypic characterization of interposon mutants. *Journal of Biological Chemistry* **267**(23): 16138-16145.

**Zhu, Y., Graham, J.E., Ludwig, M., Xiong, W., Alvey, R.M., Shen, G., and Bryant, D.A.** (2010) Roles of xanthophyll carotenoids in protection against photoinhibition and oxidative stress in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Archives of Biochemistry and Biophysics* **504**(1): 86-99.

**Zilinskas, B.A., Greenwald, L.S., Bailey, C.L., and Kahn, P.C.** (1980) Spectral analysis of allophycocyanin I, II, III and B from *Nostoc* sp. phycobilisomes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **592**(2): 267-276.

**Zilinskas, B.A., and Greenwald, L.S.** (1986) Phycobilisome structure and function. *Photosynthesis Research* **10**(1): 7-35.

**Zinser, E.R., Coe, A., Johnson, Z.I., Martiny, A.C., Fuller, N.J., Scanlan, D.J., and Chisholm, S.W.** (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology* **72**(1): 723-732.

**Zinser, E.R., Johnson, Z.I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S.W.** (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography* **52**(5): 2205-2220.

**Zinser, E.R., Lindell, D., Johnson, Z.I., Futschik, M.E., Steglich, C., Coleman, M.L., Wright, M.A., Rector, T., Steen, R., and McNulty, N.** (2009) Choreography of the transcriptome, photophysiology, and cell cycle of a

- minimal photoautotroph, *Prochlorococcus*. *PLoS One* **4**(4): e5135.
- Zolla, L., and Rinalducci, S.** (2002) Involvement of active oxygen species in degradation of light-harvesting proteins under light stresses. *Biochemistry* **41**(48): 14391-14402.
- Zubkov, M.V., Sleigh, M.A., Burkill, P.H., and Leakey, R.J.G.** (2000) Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. *Progress in Oceanography* **45**(3): 369-386.
- Zuckerandl, E., and Pauling, L.** (1965) Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8**(2): 357-366.
- Zwirgmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M.S., Zubkov, M.V., and Scanlan, D.J.** (2007) Basin-scale distribution patterns lineages in the Atlantic Ocean. *Environmental Microbiology* **9**(5): 1278-1290.
- Zwirgmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaultot, D., Not, F., Massana, R., Ulloa, O., and Scanlan, D.J.** (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology* **10**(1): 147-161.
- Zwirgmaier, K., Spence, E., Zubkov, M.V., Scanlan, D.J., and Mann, N.H.** (2009) Differential grazing of two heterotrophic nanoflagellates on marine *Synechococcus* strains. *Environmental Microbiology* **11**(7): 1767-1776.

---

# ANNEXES

---



---

---

## ANNEXE A : ARTICLE III

### IMPACT DU RÉCHAUFFEMENT CLIMATIQUE SUR LA DISTRIBUTION DES ÉCOTYPES DE *PROCHLOROCOCCUS* ET *SYNECHOCOCCUS* EN MER MEDITERRANÉE

---

**Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes  
in the Mediterranean Sea affected by global warming? (2011)**

*Biogeosciences* **8** (9): 2785-2804.

Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahe, F., Bariat, L., Courties, C., Marie, D., Ras,  
J. Mauriac, R., Jeanthon, C., Bendif, E.M., Ostrowski, M, Scanlan, D.J. and Garczarek, L.

## I. RÉSUMÉ

La mer Méditerranée est située à la limite Nord du Tropique du Cancer, et il a souvent été suggéré que cette région pourrait être particulièrement affectée par le phénomène de réchauffement global. De nombreuses études récentes ont ainsi reporté des changements significatifs dans la distribution et/ou la diversité de certains organismes méditerranéens, comme les poissons et les invertébrés marins (Coll et al., 2010; Lejeune et al., 2010), et il a été démontré que de plus en plus d'espèces tropicales et subtropicales colonisaient cette région océanique, en remplaçant souvent les espèces locales (Bianchi, 2007). Ces observations soulèvent la question de savoir si les communautés microbiennes sont également affectées par ce phénomène, en particulier dans le bassin du Levantin, situé à l'est de la Mer Méditerranée où les températures de surface de la mer ont considérablement augmenté au cours des 25 dernières années ( $0.50 \pm 0.11^\circ\text{C}$  en moyenne par décennie,  $p < 0.01$ ).

Dans cet article de Mella et collaborateurs publié en septembre 2011, la diversité de *Prochlorococcus* et *Synechococcus* en Mer Méditerranée a été étudiée en utilisant deux approches moléculaires : hybridations par la technique des dot-blots sur le gène codant pour l'ARNr 16S et des bibliothèques de clones ciblant la région intergénique située entre l'ARNr 16S et l'ARNr 23S ou ITS 16S-23S (Internal Transcribed Spacer). Cette diversité a été examinée durant deux campagnes océanographiques couvrant les bassins Ouest et Est : la campagne PROSOPE (septembre 1999) et la campagne BOUM (juin-juillet 2008). Ma participation à ce travail, que je cosigne en 3<sup>ème</sup> auteur, a consisté à préparer des banques de clones d'ITS ("16S-23S ribosomal DNA Internal Transcribed Spacer") à partir des échantillons de 5 stations à deux profondeurs et à faire les analyses phylogénétiques des séquences obtenues, qui sont venues confirmer et compléter les résultats acquis par l'approche Dot-Blot (Mella-Flores et al., 2011). Pour *Prochlorococcus*, l'écotype HLI, typique des eaux tempérées, domine la couche supérieure de la zone euphotique, alors que l'écotype HLII, majoritaire dans les eaux chaudes tropicales et subtropicales et notamment en Mer Rouge, a seulement été trouvé à très faible concentration (une seule séquence ITS d'HLII a été obtenue sur 400 séquences de *Prochlorococcus* au total). La profondeur du maximum profond de Chl était quant à elle dominée par l'écotype LLI, typiquement situé à une position intermédiaire entre les écotypes de surface (HL) et les écotypes de profondeur (LLIII et LLIV), dont la distribution est restreinte au bas de la couche éclairée (Johnson et al., 2006). Les populations de *Synechococcus* sont dominées par les clades I, II et IV dans les eaux mésotrophes du Nord-Ouest du Golfe du Lion. Le reste du transect est dominé par le clade III et des groupes génétiquement proches des clades WPC1 et VI. Une seule séquence du clade II, le clade typique des eaux chaudes (Bouman et al., 2006; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008), a été obtenue.

Ainsi, même si la Mer Méditerranée Est est située au Nord de la Mer Rouge, dominée par les *Prochlorococcus* HLII et les *Synechococcus* du clade II (Fuller et al., 2003; Zwirgmaier et al., 2008), et relié à elle par le Canal de Suez, les changements qui ont eu lieu dans l'hydrologie de cette région, notamment en raison du réchauffement global, n'ont pas été suffisants à ce jour pour provoquer un remplacement des communautés cyanobactériennes locales par des communautés subtropicales. Les températures observées dans le bassin du Levantin sont pourtant tout à fait compatibles avec la croissance des écotypes HLII de *Prochlorococcus* (Johnson et al., 2006). Le bassin levantin présente des conditions hydrologiques particulières (carence en azote et en phosphore, fortes concentrations en cuivre, etc.) par rapport aux autres environnements marins oligotrophes, ce qui suggère qu'il faut prendre en compte d'autres variables physico-chimiques que la température pour prédire les éventuels changements d'abondance des écotypes de *Prochlorococcus*.

Pour *Synechococcus*, cette étude a également permis de mettre en évidence de nouveaux clades phylogénétiques (MS1 à MS4). Les clades MS1 et MS2 semblent être minoritaires en Mer Méditerranée et la phylogéographie de ces clades méritera d'être précisée. Les clades MS3 et MS4 représentent quant à eux plus de 30% des séquences obtenues au niveau de la Station 5 à 75m lors de la campagne océanographique BOUM, et leur répartition semble être restreinte à des conditions hydrologiques particulières.

## **II. ARTICLE**

Biogeosciences, 8, 2785–2804, 2011  
 www.biogeosciences.net/8/2785/2011/  
 doi:10.5194/bg-8-2785-2011

© Author(s) 2011. CC Attribution 3.0 License.



Biogeosciences

## Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming?

D. Mella-Flores<sup>1,2,\*</sup>, S. Mazard<sup>3,4,\*</sup>, F. Humily<sup>1,2</sup>, F. Partensky<sup>1,2</sup>, F. Mahé<sup>1,2</sup>, L. Bariat<sup>5</sup>, C. Courties<sup>5</sup>, D. Marie<sup>1,2</sup>, J. Ras<sup>6</sup>, R. Mauriac<sup>7</sup>, C. Jeanthon<sup>1,2</sup>, E. Mahdi Bendif<sup>1,2</sup>, M. Ostrowski<sup>3,4</sup>, D. J. Scanlan<sup>3</sup>, and L. Garczarek<sup>1,2</sup>

<sup>1</sup>CNRS, Observatoire Océanologique, UMR7144, Groupe Plancton Océanique, 29680 Roscoff, France

<sup>2</sup>UPMC-Université Paris 06, Station Biologique, Place Georges Teissier, 29680 Roscoff, France

<sup>3</sup>School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

<sup>4</sup>Dept. Chemistry and Biomolecular Science, Macquarie University, North Ryde, NSW 2109, Australia

<sup>5</sup>CNRS/INSU and UPMC-Université Paris 06, Laboratoire Arago, UMS2348, Observatoire Océanologique, 66651 Banyuls-sur-mer, France

<sup>6</sup>CNRS and UPMC-Université Paris 06, UMR7093, Laboratoire d'Océanographie de Villefranche, 06234 Villefranche-sur-mer, France

<sup>7</sup>Laboratoire d'océanographie physique et biogéochimique, Centre d'océanologie de Marseille, case 901, campus de Luminy, 13288 Marseille cedex 09, France

\*These two authors contributed equally to this work

Received: 8 April 2011 – Published in Biogeosciences Discuss.: 3 May 2011

Revised: 8 September 2011 – Accepted: 12 September 2011 – Published: 29 September 2011

**Abstract.** Biological communities populating the Mediterranean Sea, which is situated at the northern boundary of the subtropics, are often claimed to be particularly affected by global warming. This is indicated, for instance, by the introduction of (sub)tropical species of fish or invertebrates that can displace local species. This raises the question of whether microbial communities are similarly affected, especially in the Levantine basin where sea surface temperatures have significantly risen over the last 25 years ( $0.50 \pm 0.11$  °C in average per decade,  $P < 0.01$ ). In this paper, the genetic diversity of the two most abundant members of the phytoplankton community, the picocyanobacteria *Prochlorococcus* and *Synechococcus*, was examined during two cruises through both eastern and western Mediterranean Sea basins held in September 1999 (PROSOPE cruise) and in June–July 2008 (BOUM cruise). Diversity was studied using dot blot hybridization with clade-specific 16S rRNA oligonucleotide probes and/or clone libraries of the 16S-23S ribosomal DNA Internal Transcribed Spacer (ITS) region, with a focus on the abundance of clades that may constitute bioindicators of warm waters. During both cruises, the dominant *Prochlorococcus* clade in the upper mixed layer at all stations was HLI, a clade typical of temperate waters, whereas the HLII clade, the dominant group in (sub)tropical waters, was only present at very low concentrations. The *Synechococcus* com-

munity was dominated by clades I, III and IV in the north-western waters of the Gulf of Lions and by clade III and groups genetically related to clades WPC1 and VI in the rest of the Mediterranean Sea. In contrast, only a few sequences of clade II, a group typical of warm waters, were observed. These data indicate that local cyanobacterial populations have not yet been displaced by their (sub)tropical counterparts.

### 1 Introduction

Due to their ubiquity and abundance, marine cyanobacteria are widely recognized as being major contributors to global photosynthetic biomass and primary production (Campbell et al., 1994; Goericke and Welschmeyer, 1993; Li, 1994; Partensky et al., 1999). Two genera dominate the picophytoplankton size fraction (0.2–2 µm), *Prochlorococcus* and *Synechococcus*. Although these genera co-occur in tropical and temperate areas, their abundance patterns differ not only spatially but also seasonally (Johnson et al., 2006; Tai and Palenik, 2009; Zwirgmaier et al., 2008). Indeed, *Synechococcus* prefers well-lit, nutrient-rich waters and usually blooms in spring when the water column is mixed, whereas *Prochlorococcus* is most abundant in oligotrophic waters and reaches maximal concentrations in summer and fall when waters are stratified, with a deeper sub-surface maximum than *Synechococcus* (Campbell et al., 1994; DuRand et al.,



Correspondence to: L. Garczarek  
 (laurence.garczarek@sb-roscoff.fr)

2001; Lindell and Post, 1995). Both genera are culturable and this has considerably facilitated studies on their physiological and genetic diversity. Each comprises a number of clades or “ecotypes” (sensu Coleman and Chisholm, 2007) exhibiting distinct ecophysiologicals and spatial distributions.

For *Prochlorococcus*, several different ecotypes coexist in the water column. The upper part of the euphotic zone is occupied by high light-adapted (HL) ecotypes. HLII (also called eMIT9312) dominates in permanently stratified waters from the equator to subtropics, and is progressively replaced at high latitude by HLI (also called eMED4; Johnson et al., 2006; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008). The distinct latitudinal distribution pattern of *Prochlorococcus* HLI and HLII ecotypes is well supported by the different temperature ranges for growth of representative isolates (Johnson et al., 2006; Zinser et al., 2007). These complementary distributions make variations of the relative concentrations of these two HL ecotypes a good potential bioindicator of community shifts as a result of temperature changes of oceanic waters. Two further HL clades (HNLC1 and HNLC2) have recently been shown to occur specifically in warm, iron-depleted areas of the ocean, such as the eastern Equatorial Pacific upwelling and the tropical Indian Ocean (Rusch et al., 2010; West et al., 2010).

Until recently, it was thought that the bottom of the euphotic layer was dominated by the low light-adapted LLIV ecotype (also called eMIT9313), but recent evidence using qPCR analyses with refined ecotype-specific primers suggests that it co-occurs with the LLII ecotype (also called eSS120), and the latter one could even outnumber LLIV (Malmstrom et al., 2010). A fifth ecotype (LLI, also called eNATL) was found to occupy an intermediate niche in the water column, in the vicinity of the thermocline (Johnson et al., 2006; Zinser et al., 2006, 2007). Clone library sequencing has also revealed the presence of several additional but still uncultured LL clades, including NC1, a potentially abundant group found at depth at several stations of the Atlantic and Pacific oceans (Martiny et al., 2009), and two LL clades (LLV and LLVI) seemingly specific for low oxygen zones (Lavin et al., 2010). However, at high latitudes, the lower stability of the water column progressively simplifies these depth distribution patterns, since only HLI and LLI ecotypes appear able to survive in mixed, cool water conditions (Johnson et al., 2006).

For *Synechococcus*, some general distribution patterns can also be defined for several of the ten clades defined by Fuller et al. (2003), although the genetic and/or physiological bases behind these patterns are much less well understood than for *Prochlorococcus*. Clades I and IV generally co-occur at latitudes above 30° N/S and seem to be restricted to near coastal waters in the lowest part of their latitudinal distributions (Tai and Palenik, 2009; Zwirgmaier et al., 2007, 2008; Mazard et al., 2011). In contrast, clade II seems to be abundant in warm, coastal or shelf areas and is thought to be the (sub)tropical counterpart of clades I/IV, although some overlap may oc-

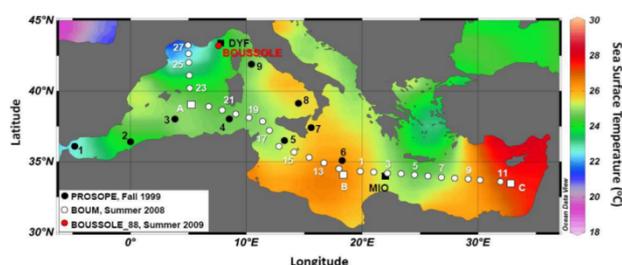
cur in the boundary zone, i.e. between 30° and 35°, in both the southern and northern hemispheres (Zwirgmaier et al., 2008). The latitudinal distribution of clade III appears to be broader, but cells belonging to this ecotype seemingly prefer oligotrophic, offshore waters. All other clades are generally found at lower concentrations than clades I to IV and their distribution patterns are therefore less clearly defined (Zwirgmaier et al., 2008).

A number of recent studies have reported striking changes in the distribution and/or diversity of a variety of Mediterranean organisms, such as fish and marine invertebrates (for recent reviews, see Coll et al., 2010; Lejeune et al., 2009). Indeed, it seems that more and more tropical or subtropical species are colonizing this marine area, often replacing local species (Bianchi, 2007). Whether the community structure of marine prokaryotes is also affected by environmental change, and in particular picocyanobacteria that are key players in the productivity of the Mediterranean Sea, has been little studied so far. In the present paper, we use dot blot hybridization analysis and/or clone library sequencing to analyze the vertical and horizontal distributions of *Prochlorococcus* and *Synechococcus* ecotypes/clades during two cruises in the Mediterranean Sea, PROSOPE and BOUM, held some nine years apart in the Mediterranean Sea and which followed comparable transects through both western and eastern basins.

## 2 Methods

### 2.1 Sampling

Water samples were collected at several stations during two independent cruises in the Mediterranean Sea, the PROSOPE cruise (Productivity of Oceanic Pelagic Systems) that was held from 9 September to 1 October, 1999 aboard the French research vessel R/V *La Thalassa* (for details on hydrology; see Garczarek et al., 2007; Marie et al., 2006; Crombet et al., 2011) and the BOUM cruise (Biogeochemistry from the Oligotrophic to the Ultra-oligotrophic Mediterranean) held from 22 June to 18 July, 2008 aboard the R/V *L'Atalante* (Fig. 1). Additional samples were collected at the BOUSSOLE site (located near the PROSOPE cruise DYF site; see Fig. 1) during the BOUSSOLE 88 cruise held on 15 June 2009 aboard the R/V *Tethys II*. At each station, 4 to 81 water samples were retrieved from 4 to 6 discrete depths using 12 l Niskin bottles, fitted on a Seabird CTDO rosette sampler (model SBE 911) equipped with in situ conductivity, temperature, depth, oxygen as well as chlorophyll fluorescence sensors. Seawater was subsequently filtered through a 20 µm meshed silk (Whatman Co.) and 47 mm diameter, 3 µm pore size polycarbonate Nuclepore filters (Whatman, Maidstone, UK). Picoplanktonic cells were then collected by filtration onto 47 mm diameter, 0.2 µm pore size Pall Supor-200 filters (Gelman Inc., Ann Arbor, MI). The



**Fig. 1.** False-color composite satellite image of sea surface temperature of the Mediterranean Sea in July 2008 (data: NOAA). The sampling stations of the PROSOPE, BOUM and BOUSSOLE 88 cruises, numbered in order of occupation, are indicated in black, white and red, respectively. Short- and long-term stations are indicated by dots and squares, respectively.

filter was transferred into a cryovial containing 1.8 ml of DNA lysis buffer (20 mM EDTA, 400 mM NaCl, 0.75 mM sucrose, 50 mM Tris pH 9), quickly frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until extraction. At all stations and depths, samples were taken to determine nitrate and phosphate concentrations (Raimbault et al., 1990) and perform HPLC pigment analyses (Ras et al., 2008). *Prochlorococcus* and *Synechococcus* cell concentrations were measured by flow cytometry, as detailed previously (Marie et al., 1999; Obernosterer et al., 2008). Due to the dim fluorescence of *Prochlorococcus* cells in the upper mixed layer, populations were not completely resolved from the background noise at most stations, so that for integrating cell numbers up to the surface, we assumed that the concentration measured at the bottom of the mixed layer was the same throughout the top layer. All these data are available on-line on the BOUM cruise database (<http://www.com.univ-mrs.fr/BOUM/>).

## 2.2 DNA extraction

DNA was extracted from the filters largely following the method described by Massana et al. (2004) but with slight modifications. Briefly, filters from selected stations and depths were first cut into small pieces using clean scissors and resuspended by strong vortexing in lysis buffer, used as storage buffer. Samples were then incubated at  $37^{\circ}\text{C}$  for 45 min in the presence of lysozyme at a final concentration of  $1\text{ mg ml}^{-1}$ , then 1 h at  $50^{\circ}\text{C}$  with proteinase K and SDS (final concentrations of  $0.2\text{ mg ml}^{-1}$  and 1% w/v, respectively). Nucleic acids were extracted consecutively with phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1) before being precipitated with 2.5 vol of 100% (v/v) cold ethanol and 0.1 vol of 3 M sodium acetate (pH 5.2) at  $-20^{\circ}\text{C}$ . Nucleic acids were recovered by centrifugation, washed twice with 70% ethanol (v/v), resuspended in sterile water and stored at  $-80^{\circ}\text{C}$ .

Strains representative of the different *Prochlorococcus* and *Synechococcus* clades, obtained from the Roscoff Culture

Collection (RCC, <http://www.sb-roscoff.fr/Phyto/RCC/>) or held at Warwick University, were used as controls in dot blot hybridization experiments (Table 1). Their DNA was extracted as described by Eguchi et al. (2001). Briefly, 50 ml of late logarithmic phase cells were pelleted and then resuspended in 567  $\mu\text{l}$  of TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). Cells were lysed at  $60^{\circ}\text{C}$  for 4 h after addition of SDS (0.5% w/v final) and proteinase K ( $250\text{ mg ml}^{-1}$  final). NaCl (100  $\mu\text{l}$  of 5 M) and CTAB (80  $\mu\text{l}$  of 10% w/v hexadecyltrimethyl ammonium bromide in 0.7 M NaCl; Sigma-Aldrich, St Louis, MO, USA) were added and mixed thoroughly and samples incubated at  $65^{\circ}\text{C}$  for 10 min. Nucleic acids were purified from CTAB complexes by successive chloroform:isoamyl alcohol (24:1; 1 vol) and phenol:chloroform:isoamyl alcohol (25:24:1; 1 vol) extractions before being precipitated by addition of 0.6 vol isopropanol. Nucleic acids were then purified successively in 50%, 70% and 100% (v/v) ethanol, resuspended in sterile water and kept at  $-20^{\circ}\text{C}$  for further use. The quality and quantity of environmental or strain DNA was monitored using a Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA).

## 2.3 PCR amplification for dot blot hybridization analysis

16S rDNA sequences were amplified from control strains and environmental DNA using the primers OXY107F (GGACGGGTGAGTAACGCGTG) and OXY1313R (CTTCAYGYAGGCGAGTTGCAGC), which specifically target sequences from oxygenic phototrophs (Fuller et al., 2003).

PCR reactions were carried out in a total reaction volume of 100  $\mu\text{l}$  with 1.0–10 ng of template DNA per reaction. The reaction mix consisted of 200 nM forward and reverse primers, 200  $\mu\text{M}$  dNTPs, 2.0 mM  $\text{MgCl}_2$  and 2.5 U of Taq DNA polymerase (Qiagen, Valencia, CA). For amplification of environmental DNAs, BSA (bovine serum albumin, Sigma-Aldrich, St Louis, MO, USA) was added at  $1.0\text{ mg ml}^{-1}$  and hot start PCR was routinely performed. The reaction cycle consisted of an initial denaturation step of 4 min at  $96^{\circ}\text{C}$  and 1 min at  $80^{\circ}\text{C}$ , then primers, dNTPs and Taq were added to the reaction mix, followed by 25–30 cycles of  $94^{\circ}\text{C}$  for 45 s,  $55^{\circ}\text{C}$  for 30 s,  $72^{\circ}\text{C}$  for 1 min, with a final extension at  $72^{\circ}\text{C}$  for 6 min. PCR products were purified using the Qiaquick PCR purification kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The amplification reaction was repeated until sufficient DNA amount was available for each sample. PCR products were then pooled for dot blot hybridization analysis.

## 2.4 Dot blot hybridization

Hybridization conditions were as previously described (Fuller et al., 2003; West and Scanlan, 1999). Briefly,

**Table 1.** Characteristics of 16S rDNA oligonucleotide probes used for dot blot hybridization. Abbreviations: *Syn.*, *Synechococcus*; *Proc.*, *Prochlorococcus*; n.a., not applicable.

Probe names	Sequence (5'-3')	Wash Temp. (°C)	Target group	Reference genome <sup>1</sup>	Control sample <sup>2</sup>	References
EUB338	GCTGCCTCCCGTAGGAGT	46	Eubacteria	n.a.	all	(Amann et al., 1990)
SYN1006	CTCTCAAGTTTCCAAGAG	44	<i>Syn.</i> clade I	CC9311	CC9311	(Fuller et al., 2003)
SYN1006RS	CTCTCCTGTTTCCAAGAG	42	<i>Syn.</i> clade II	CC9605	RS9903	Idem
SYN262	GATGCCTTGGTAGGCCTT	53	<i>Syn.</i> clade III	WH8102	WH8103	Idem
SYN635	AAGCCCTCAGTTTCCAC	50	<i>Syn.</i> clade IV	BL107, CC9902	5x15	Idem
SYN1280	GAGCCACGGTTTATGAGA	51	<i>Syn.</i> clade V/VI/VII	WH7803/WH7805/-	WH7803/WH8018/ MITS9220	Idem
SYN620	CACTGCCACGATGGAGTT	51	<i>Syn.</i> clade VIII	RS9917	RS9913	Idem
SYN1000	GGTTTCCAGAAATTCGC	45	<i>Syn.</i> clade IX	RS9916	RS9916	Idem
SYN1007	ACCCTCCGGTTTCCAGA	48	<i>Syn.</i> clade X <sup>3</sup>	RCC307	RCC307	idem
S2PRO640R	ATACTCAAGCCTTTCAGTTT	41	<i>Proc.</i> HLII	MIT9312	TAK9803-2	(Zwirgmaier et al., 2007)
S1PRO634R	GCCGATCAGTTTCCACTG	50	<i>Proc.</i> HLI	MED4	PCC9511	(West and Scanlan, 1999)
DPRO634R	GCCAATCAGTTTCCACTG	46	<i>Proc.</i> LLI/LLIV	NATL1/MIT9313	NATL1/MIT9303	Idem
SARG634R	GCCCTCAGTTTCCACTG	44	<i>Proc.</i> LLII	SS120	SS120	idem

<sup>1</sup> Reference strain (with sequenced genome) representative of the target group, when available (Dufresne et al., 2008).

<sup>2</sup> Strains (or environmental sample in the case of clade IV) representative of the target group and used here as controls for dot blot hybridizations. DNA from the same samples was also used to draw hybridization curves for each probe (cf. materials and methods).

<sup>3</sup> Now called sub-cluster 5.3 (Dufresne et al., 2008).

oligonucleotide probes (30 ng, dual HPLC grade purified) targeting each *Prochlorococcus* and *Synechococcus* clade (Table 1) were end-labeled with  $\gamma$ -<sup>32</sup>P-ATP using T4 polynucleotide kinase in 1X forward kinase buffer (Invitrogen, Carlsbad, CA) at 37 °C for 1 h. Unincorporated nucleotides were then removed by purification through a mini Quick Spin DNA Column (Roche, Basel, Switzerland). Purified 16S rDNA amplicons from BOUM environmental DNA and control strains were blotted onto Zeta-Probe nylon membranes using a Bio-Dot microfiltration apparatus (both from BioRad, Laboratories Ltd., Hemel Hempstead, Herts, UK) and hybridized with labeled clade-specific oligonucleotide probes as described previously (Fuller et al., 2003; see Table 1). Hybridization signals were quantified using a Phosphorimager (Typhoon Trio) and ImageQuant TL software (both from GE Healthcare, Buckinghamshire, UK).

Relative hybridization of clade-specific probes to total oxygenic phototroph 16S rDNA sequences were calculated as described previously (Fuller et al., 2003) by normalizing the individual *Prochlorococcus* and *Synechococcus* probe signals to the signal for total amplified DNA, as determined with the EUB338 probe. The dot blot hybridization data as well as other environmental parameters were visualized using the Ocean Data View software, designed by R. Schlitzer (<http://odv.awi.de>).

## 2.5 Clone libraries and phylogenetic analyses

16S-23S ribosomal DNA Internal Transcribed Spacer (hereafter ITS) regions were amplified using the 16S-1247f (CG-TACTACAATGCTACGG) and 23S-241r (TTCGCTCGC-CRCTACT) primer set (Rocap et al., 2002). PCR reactions were performed in 25  $\mu$ l volumes containing 1.25 U Taq polymerase (Promega, Madison, WI), 1X buffer provided with

the polymerase, 1.5 mM MgCl<sub>2</sub>, 1.25 mM of each dNTP, 0.5–1  $\mu$ m of each primer and 0.5–1  $\mu$ l of environmental DNA template. The reaction cycle consisted of an initial denaturation step of 5 min at 95 °C, followed by 25–30 cycles of 1 min at 94 °C, 1 min at 52 °C and 2 min at 72 °C with a final extension at 72 °C for 10 min. To minimize PCR biases, the number of PCR cycles was adjusted between 25 to 30 cycles in order to obtain a very faint band on agarose gels. Ten parallel reactions were performed for each sample followed by a three-cycle reconditioning step (Thompson et al., 2002) by transferring 2.5  $\mu$ l each of the initial reaction mixture into 22.5  $\mu$ l of fresh PCR mixture. PCR products were then pooled, purified using the Qiaquick kit PCR purification kit (Qiagen, Valencia, CA) and cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA). The presence of an insert was screened by PCR amplification of colonies using the vector specific M13 primers. The size of inserts, as measured by agarose gel electrophoresis, was also used to specifically enrich *Synechococcus* sequences since the ITS sequence is generally longer in *Synechococcus* than in *Prochlorococcus* (except LLIV strains; Rocap et al., 2002). Between 69–80 clones from each sample were sequenced bidirectionally on an ABI3130 automated sequencer using Big Dye terminator Cycle sequencing kits version 3.0 (both from Applied Biosystems, Foster City, CA), according to the manufacturer's instructions.

A total of 746 environmental ITS sequences were obtained. In further analysis of these sequences primer regions were excluded. To avoid redundancy, only 149 sequences, including only one sequence per cd-hit cluster, as defined with a percent identity cutoff of 99% (Li and Godzik, 2006), and all BOUM sequences from the four new clades described in this study, were deposited in the Genbank nucleotide sequence database under the following accession numbers:

JN006522-JN006718. Their phylogenetic affiliation was determined using Pplacer v1.0, which places sequences into a reference tree (Matsen et al., 2010). First, a reference alignment of full-length ITS sequences was made with 97 selected sequences of isolates and environmental clones representative of the whole diversity of *Synechococcus* and *Prochlorococcus* clades/ecotypes known so far, including four novel groups that were revealed by preliminary phylogenetic analyses of our own sequences (see results). The alignment was done using MAFFT v6.847 taking into account the RNA secondary structure (Q-INS-i option; Katoh and Toh, 2008), then manually refined using BioEdit v7.0.5.3 (Hall, 1999). The general time-reversible (GTR+I+G) model of nucleotide substitution was selected according to JModeltest v0.1.1 (Posada, 2008) and the corrected Akaike information criterion (Posada, 2009) for the phylogenetic inference of each sequence by Maximum Likelihood (ML) using PhyML v3.0 (Guindon and Gascuel, 2003) with gamma and invariant sites distributions.

Environmental sequences were then aligned with a HMM built from the reference alignment using tools from the HMMER v3.0 suite (<http://hmmer.janelia.org/>), with default parameters. Using this alignment and the reference tree, the most probable location for each environmental sequence was determined using Pplacer and this was visualized by additional branches in the tree (see Fig. S2 in the Supplement). The same affiliation of environmental sequences was obtained with two independent methods (not shown): (i) a clustering strategy based on pairwise-distances between reference and environmental sequences (Needleman-Wunsch algorithm as implemented in Emboss; Needleman and Wunsch, 1970) and (ii) a strategy based on maximum likelihood analyses of a large multiple sequence alignment including both reference and environmental sequences using RaxML (Stamatakis, 2006).

While a full-length ITS sequence reference alignment was necessary for Pplacer, phylogenetic analyses were performed after removing a number of hypervariable regions and gaps, leaving a total of 1135 positions, including tRNAs. Bayesian inference was then conducted using Mr Bayes v3.1.2 (Huelsenbeck and Ronquist, 2001), starting with a random tree, run two million generations in four chains and with a burn-in of 2000 generations, using parameters similar to the ML method. Neighbor-Joining (NJ) based on the F84 distance model and Maximum Parsimony (MP) trees were inferred using Phylip v3.69 (Felsenstein, 1989). Robustness of inferred topology was supported by bootstrap resampling (1000) with ML, NJ and MP. Very similar tree topologies were obtained for the full-length and the cut versions of the reference alignment except for slight modifications of the relative position of a few groups, essentially within sub-cluster 5.1B (see results).

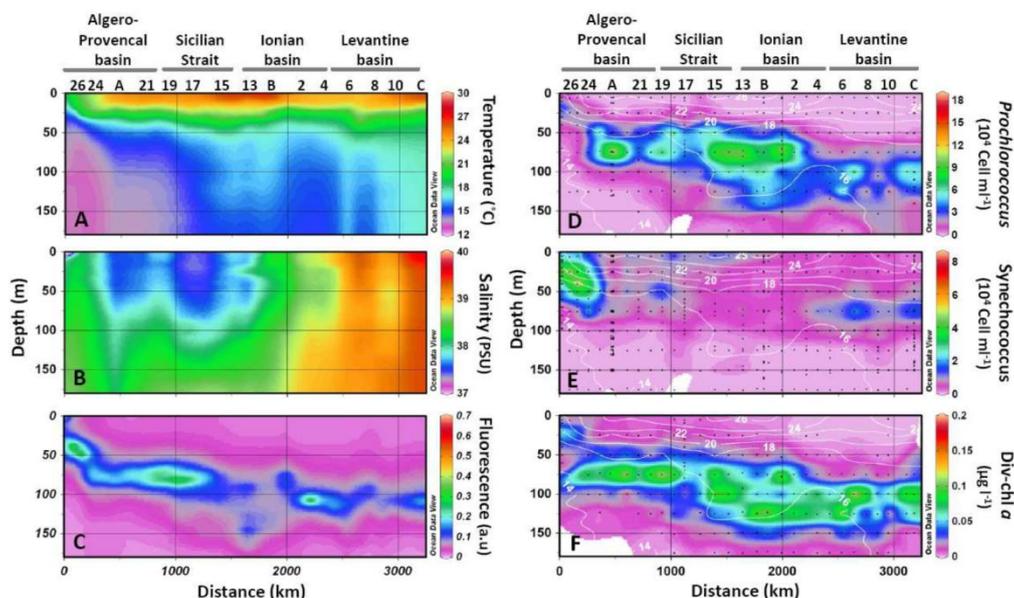
### 3 Results

#### 3.1 Hydrology

The Mediterranean Sea is a semi-enclosed sea undergoing extensive water mixing during winter, a seasonal phenomenon that injects nutrients from deep waters into surface waters (Marty et al., 2002). These nutrients are rapidly consumed by phytoplankton during late winter/early spring blooms, which occur in surface waters in the northern Mediterranean Sea (as seen by satellite imagery; D'Ortenzio and Ribera d'Alcala, 2009) and at depth further south (Crombet et al., 2011). By early summer, the upper mixed layer becomes oligotrophic and the chlorophyll biomass reaches annual minima over the whole Mediterranean Sea (D'Ortenzio and Ribera d'Alcala, 2009).

Figure 2a–c shows high resolution plots of the vertical distribution of temperature, salinity and fluorescence along the BOUM cruise transect held in late June-early July 2008. These graphs show the occurrence of a strong vertical thermal stratification along the whole transect (Fig. 2a) and a strong horizontal salinity gradient increasing from W to E (Fig. 2b), with a marked halocline located in the Ionian Sea. This gradient separates western Atlantic waters entering the Mediterranean Sea in the Straits of Gibraltar, from the high salinity waters of the east Mediterranean Sea. Also notable is the progressive eastwards deepening of the fluorescence maximum (Fig. 2c), a proxy for the deep chlorophyll maximum (DCM), associated with a comparable deepening of nutrient lines (not shown, but see Crombet et al., 2011; Pujo-Pay et al., 2011). These summertime hydrological characteristics were globally similar to those previously reported for the PROSOPE cruise in late summer-early fall (Garczarek et al., 2007; Marie et al., 2006), including sea surface temperatures (SST) above 24 °C (except at the three northernmost stations of BOUM and the westernmost station of PROSOPE), and comparable depths of the upper mixed layer (25–30 m) in the common parts of both cruises. The most notable hydrological difference between the two cruises was the position of the halocline, located more eastward during BOUM (in the Ionian Sea) than during PROSOPE (off the Algerian coasts; Garczarek et al., 2007).

SST data for the 1985–2010 period were retrieved from the National Oceanic and Atmospheric Administration (<http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/>; Reynolds et al., 2002) in order to calculate the summer SST anomaly for the same period, i.e. the difference between the average SST for the three warmest months (July, August and September) of any given year and the mean summer SST averaged over the last 25 years. Figure 3 shows both the monthly SST variations and the evolution of summer SST anomalies in the 1° latitude-longitude grid areas surrounding the three long-term stations (hereafter Sta.) of the BOUM cruise. The latter stations were in the middle of three anticyclonic eddies, located in the Algero-Provencal basin (Sta. A),



**Fig. 2.** Depth profiles of temperature (A), salinity (B), fluorescence (C), *Prochlorococcus* abundance (D), *Synechococcus* abundance (E) and concentration of the *Prochlorococcus*-specific pigment divinyl-Chl *a* (F) along the BOUM cruise track. For each panel, the *x* axis represents the cumulative distance in kilometers from a site near the French coast (43.21° N, 4.93° E) and the *y* axis represents depth in meters. The positions of the main marine regions are indicated at the top of the graph as well as the number (or letter) of every two stations along the transect. For figures (D) to (F), white contour lines represent temperature in °C and black dots correspond to sampled depths. Note that to better emphasize the differential distributions of *Prochlorococcus* and *Synechococcus* along the BOUM transect, different scales have been used in Fig. 2d and e.

the Ionian basin (Sta. B) and the Levantine basin (Sta. C). Over this 25-year period, the SST measured during the coldest (March) and warmest (August) months showed no significant trend at Sta. A, whereas at Sta. B and C, SST increased for both March and August (Fig. 3a–c). Similarly, the summer SST anomaly showed a significant trend only in the eastern Mediterranean Sea (Fig. 3d–f), with the strongest rate of increase ( $0.050 \pm 0.011$  °C per year,  $P < 0.01$ ) observed at the easternmost Sta. C (Fig. 3f). It must be noted though that when considering only the period between the two cruises (1999–2008), there were no significant SST increase at the latter station.

### 3.2 Vertical distributions of marine picocyanobacteria

Figure 2d–e shows the vertical distributions of *Prochlorococcus* and *Synechococcus* during the BOUM cruise, as determined by flow cytometric cell counting. The highest *Synechococcus* abundances (range:  $3.5$ – $8.3 \times 10^4$  cells  $\text{ml}^{-1}$ ; Table 2) were observed in subsurface, mesotrophic waters of the Gulf of Lions, i.e. at the northernmost stations of the Algero-Provençal basin (Sta. 25–27), and the depth of maximum abundance progressively increased southwards to reach 75 m at Sta. A. From Sta. 22 to 18, fairly high *Synechococcus* abundances (range:  $0.63$ – $1.93 \times 10^4$  cells  $\text{ml}^{-1}$ ) were found down to 50 m, whereas from the Sicily Strait (Sta. 17) to

Sta. B cell densities dropped very rapidly with depth, with typically a 2- to 3-fold decrease between surface (ca. 5 m) and 25 m waters. Comparatively, in the same area during the PROSOPE cruise (Sta. 5 and 6), surface abundances of *Synechococcus* were about two-fold higher than on the BOUM cruise and fairly high cell densities ( $>10^4$  cells  $\text{ml}^{-1}$ ) were observed down to 70 m (Garczarek et al., 2007). East of BOUM Sta. 3, the *Synechococcus* maximum systematically occurred around 75 m, i.e. about 25 m above the DCM (Table 2). Since the PROSOPE transect did not extend as far east as the BOUM cruise, it is not possible to say whether this marked deep *Synechococcus* maximum, possibly related to the extreme oligotrophy of the upper layer in the eastern Mediterranean Sea waters, already existed in 1999 (but see Fig. 5d in Tanaka et al., 2007, for a profile made in May 2002). In terms of integrated *Synechococcus* concentrations, the highest value of the BOUM cruise occurred at Sta. 25 ( $2.50 \times 10^8$  cells  $\text{cm}^{-2}$ ) and the lowest at Sta. B ( $2.81 \times 10^7$  cells  $\text{cm}^{-2}$ ; Table 2), consistent with the very different nutrient regimes of these two stations (Crombet et al., 2011).

For *Prochlorococcus*, the highest abundance during the whole BOUM cruise was observed at Sta. 1 at a depth of 71 m ( $1.84 \times 10^5$  cells  $\text{ml}^{-1}$ ). This mid-depth abundance maximum extended over most of the transect from Sta. A

**Table 2.** Maximal and integrated concentrations of *Prochlorococcus* and *Synechococcus* at each station of BOUM and PROSOPE cruises.

Cruise	Station	DCMD (m)	MAC <i>Proc.</i> (cells ml <sup>-1</sup> )	MAD <i>Proc.</i> (m)	IC <i>Proc.</i> (cells cm <sup>-2</sup> )	MAC <i>Syn.</i> (cells l <sup>-1</sup> )	MAD <i>Syn.</i> (m)	IC <i>Syn.</i> (cells cm <sup>-2</sup> )
BOUM	27	30	5.19E+04	19	1.60E+08	3.54E+04	10	1.23E+08
	26	37	1.05E+04	25	n.a.	8.35E+04	25	2.38E+08
	25	51	3.40E+04	25	6.23E+07	7.73E+04	41	2.50E+08
	24	71	3.85E+04	51	2.07E+08	3.23E+04	51	2.07E+08
	23	75	1.01E+05	50	4.30E+08	3.05E+04	50	1.74E+08
	A	88	1.38E+05	75	5.60E+08 ± 1.13E+08 (n = 15)	1.01E+04	75	4.0E+07 ± 5.93E+06 (n = 21)
	22	74	6.69E+04	75	5.90E+08	1.72E+04	25	1.23E+08
	21	87	5.64E+04	51	2.79E+08	1.18E+04	5.9	5.80E+07
	20	78	7.56E+04	76	3.66E+08	1.28E+04	51	8.93E+07
	19	67	7.43E+04	76	2.99E+08	1.93E+04	51	8.80E+07
	18	81	8.97E+04	51	5.67E+08	1.73E+04	5.2	1.15E+08
	17	77	4.74E+04	65	3.61E+08	8.82E+03	6.0	6.71E+07
	16	73	5.58E+04	51	2.96E+08	8.42E+03	5.3	4.30E+07
	15	98	1.06E+05	76	8.08E+08	1.35E+04	5.1	6.47E+07
	14	118	1.00E+05	75	6.35E+08	1.16E+04	4.9	3.78E+07
	13	140	1.17E+05	81	6.73E+08	1.23E+04	5.1	5.90E+07
	12	120	8.15E+04	75	5.65E+08	1.22E+04	4.9	4.81E+07
	B	141	8.42E+04	75	6.55E+08 ± 1.19E+08 (n = 23)	8.09E+03	5.0	2.81E+07 ± 4.85E+06 (n = 22)
	1	78	1.84E+05	71	7.27E+08	5.05E+03	101	4.10E+07
	2	n.a.	5.67E+04	100	5.66E+08	6.47E+03	75	4.70E+07
	3	104	4.02E+04	125	2.38E+08	7.38E+03	50	5.90E+07
	4	98	2.13E+04	101	1.93E+08	1.43E+04	75	7.14E+07
	5	114	6.94E+04	100	2.57E+08	1.55E+04	75	5.43E+07
6	117	8.75E+04	125	4.07E+08	9.28E+03	75	6.17E+07	
7	104	9.38E+04	101	2.59E+08	2.42E+04	76	1.41E+08	
8	98	3.68E+04	100	3.30E+08	1.60E+04	76	7.91E+07	
9	128	4.83E+04	121	2.73E+08	9.55E+03	76	6.47E+07	
10	108	4.34E+04	100	3.65E+08	1.02E+04	75	5.56E+07	
11	109	6.35E+04	100	3.96E+08	1.31E+04	75	6.34E+07	
C	108	5.86E+04	100	2.73E+08 ± 4.076E+07 (n = 23)	1.81E+04	75	6.61E+07 ± 1.11E+07 (n = 23)	
PROSOPE	1	49	4.54E+04	29	2.54E+08	9.89E+04	49	5.64E+08
	2	40	3.80E+04	29	1.67E+08	5.61E+04	15	2.49E+08
	3	5	2.23E+05	36	3.27E+08	2.44E+04	5	9.07E+07
	4	78	9.57E+04	60	4.52E+08	1.87E+04	12	7.51E+07
	5	55	6.95E+04	55	4.06E+08	1.93E+04	6	6.77E+07
	6	85	1.25E+05	70	6.03E+08	1.37E+04	29	6.28E+07
	MI0	90	1.84E+05	70	6.93E+08 ± 2.11E+08 (n = 28)	2.15E+04	70	1.04E+08 ± 5.89E+07 (n = 28)
	DYF	47	1.15E+05	26	3.30E+08 ± 8.03E+07 (n = 23)	6.26E+04	47	2.24E+08 ± 1.12E+08 (n = 23)

Abbreviations: DCMD, Deep Chlorophyll Maximum Depth; MAC, Maximum Abundance Concentration; MAD, Maximum Abundance Depth; IC, Integrated Concentration over the top 200 m; *Syn.*, *Synechococcus*; *Proc.*, *Prochlorococcus*; n.a., not available; n, number of samples used to calculate the mean and standard deviation.

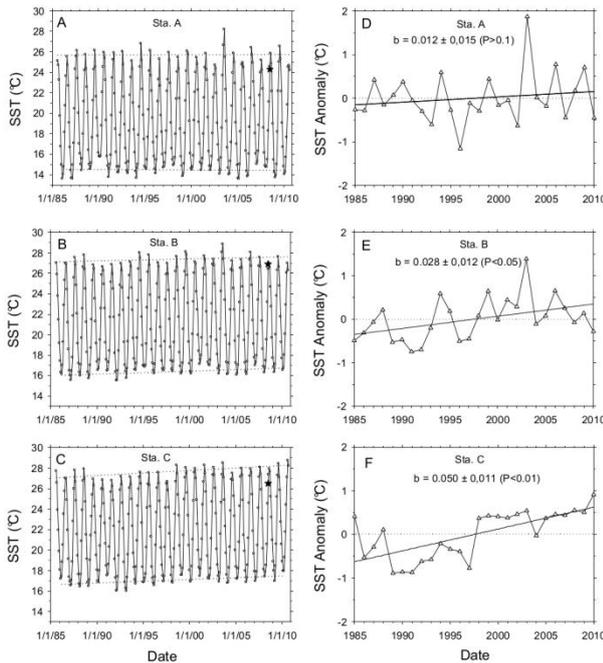
(Algero-Provencal basin) to Sta. 1 (Ionian Sea). At Sta. B, the most oligotrophic station of the transect (i.e., exhibiting the lowest integrated nitrate and phosphate concentrations over the top 150 m layer; see Moutin et al., 2002), the layer of high *Prochlorococcus* cell abundance extended very deeply, with average values as high as  $4.7 \times 10^4$  cells ml<sup>-1</sup> at 140 m. East of Sta. 1, the *Prochlorococcus* abundance maximum progressively deepened down to 100 m with a concomitant decrease of cell densities (Table 2 and Fig. 2d). The maximum concentration of *Prochlorococcus* specific pigment divinyl-Chl *a* followed a similar trend, though it was sometimes deeper in the Ionian basin, likely due to strong photoacclimation in these very transparent, ultra-oligotrophic waters (Fig. 2f). In terms of integrated *Prochlorococcus* concentrations, the highest value of the BOUM cruise occurred

at Sta. 15 ( $8.08 \times 10^8$  cells cm<sup>-2</sup>) and the lowest at Sta. 25 ( $6.23 \times 10^7$  cells cm<sup>-2</sup>; Table 2).

### 3.3 Distribution of picocyanobacterial lineages in the Mediterranean Sea

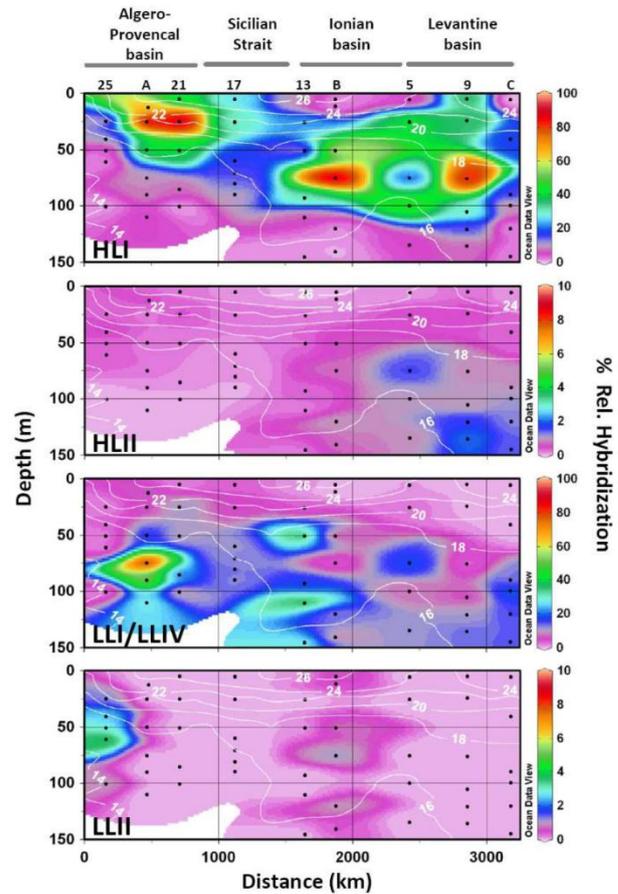
#### 3.3.1 *Prochlorococcus*

The distribution of the main *Prochlorococcus* ecotypes at selected stations of the BOUM cruise (Fig. 4) and at Station BOUSSOLE (sampled in June 2009; see Fig. S1 in the Supplement), was analyzed by dot blot hybridization using 16S rRNA-targeted oligonucleotides. Results showed that over the whole transect, the dominant HL ecotype was HLI, whereas HLII was present at very low concentrations. Analysis of the depth distribution of HLI showed that its peak



**Fig. 3.** Time plots over the 1985–2010 period of sea surface temperature (SST) as measured by satellite (A, B, C) and average SST anomaly for the three summer months (D, E, F) in the  $1^{\circ}$  latitude-longitude grid around the long term stations of the BOUM cruise (Sta. A, B and C; data: NOAA). The 25-year trend for the coldest (March) and warmest (August) months (dotted lines) and the actual SST measured on board during the BOUM cruise (stars) are indicated on plots A–C. The trend curve (plain line) and corresponding slope (b) and P-values are shown on plots E–F.

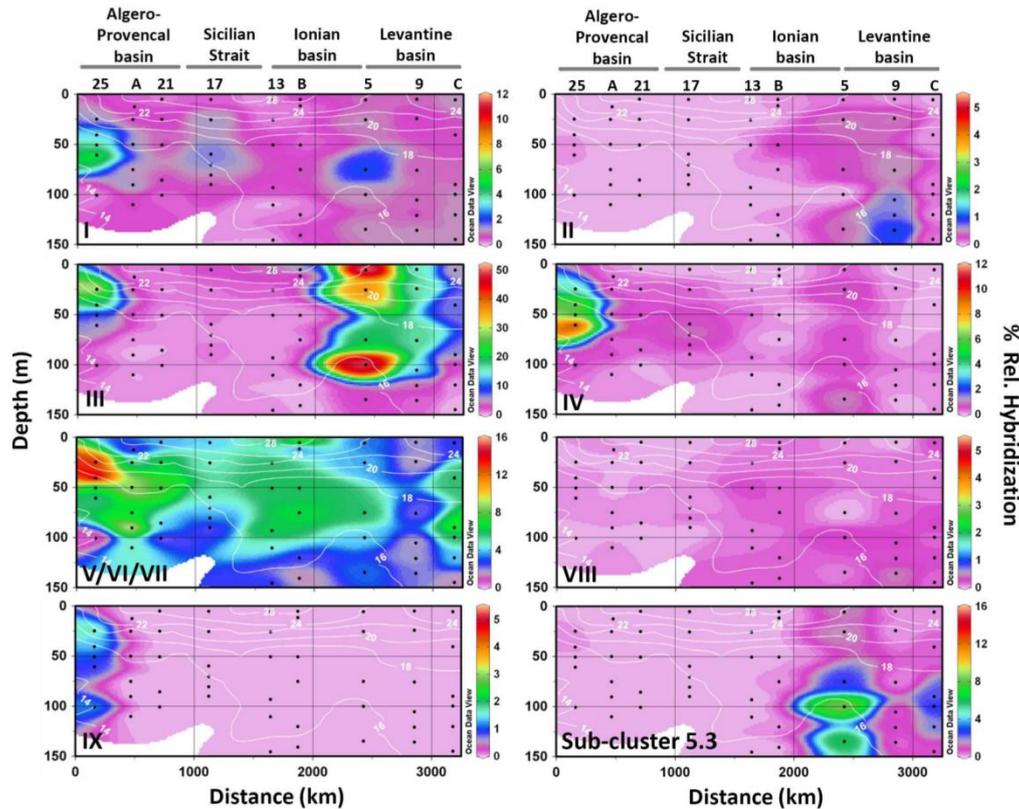
abundance generally occurred around 25 m in the western basin and 75 m in the eastern basin (except at Sta. 5) then rapidly decreased below these depths. Deep *Prochlorococcus* populations were dominated by LL ecotypes belonging to the LLI and/or LLIV clades; the oligonucleotide probe used here did not allow us to discriminate between these two populations, but previous studies have shown that LLI (eNATL) occurs at mid-depth (generally at the base of the thermocline) and LLIV (eMIT9313) at the bottom of the euphotic zone (Johnson et al., 2006; Zinser et al., 2006). At Sta. 13, the vertical distribution of LL ecotypes exhibited two clearly separated abundance maxima (with the upper one likely dominated by LLI and the bottom one by LLIV), a feature apparently related to an anomaly in the hydrological structure. Indeed, a low-salinity/high-nutrient layer was observed at 50 m, probably due to a recent event of downwelling of modified Atlantic Waters below Mediterranean waters (see Fig. 2b for salinity and Crombet et al., 2011, for nutrients). The *Prochlorococcus* LLII (eSS120) ecotype was only abundant in the northwestern part of the Mediterranean Sea, BOUM Sta. 25 and Sta. BOUSSOLE, where notable hy-



**Fig. 4.** Vertical distribution of *Prochlorococcus* ecotypes along the BOUM cruise track analyzed by dot blot hybridization using 16S rRNA gene specific oligonucleotide probes targeting HLI (high light ecotype I or eMED4), HLII (high light ecotype II or eMIT9312), LLI/LLIV (low light ecotypes I or eNATL and IV or eMIT9313) and LLII (low light ecotype II or eSS120). Contour plots indicate the % relative hybridization. The positions of the main marine regions are indicated at the top of the graph as well as the number (or letter) of the stations that have been sampled for dot blot hybridization. Details of panels are as described in Fig. 2. Note the different ranges of % relative hybridization values between panels.

bridization signals could be detected from the DCM down to the bottom of the photic layer (see Fig. 4 and Fig. S2 in the Supplement).

The distribution of the main *Prochlorococcus* clades was previously analyzed in September 1999 during the PROSOPE cruise using the same molecular approach (Garczarek et al., 2007) and gave comparable results. Indeed, the HLI population dominated in the upper layer over the whole transect and HLII was negligible, except at Sta. 5 (located near BOUM Sta. 16; Fig. 1), where the HLII hybridization signal reached 8% at 25 m, one fifth of the HLI signal.



**Fig. 5.** Vertical distribution of different phylogenetic groups of *Synechococcus* along the BOUM cruise track, as analyzed by dot blot hybridization using 16S rRNA gene specific oligonucleotide probes targeting sub-cluster 5.1 clades I to IX and sub-cluster 5.3. See Table 1 for correspondence to sequenced strains. Details of panels are as described in Fig. 2. Note the different ranges of % relative hybridization values between panels.

### 3.3.2 *Synechococcus*

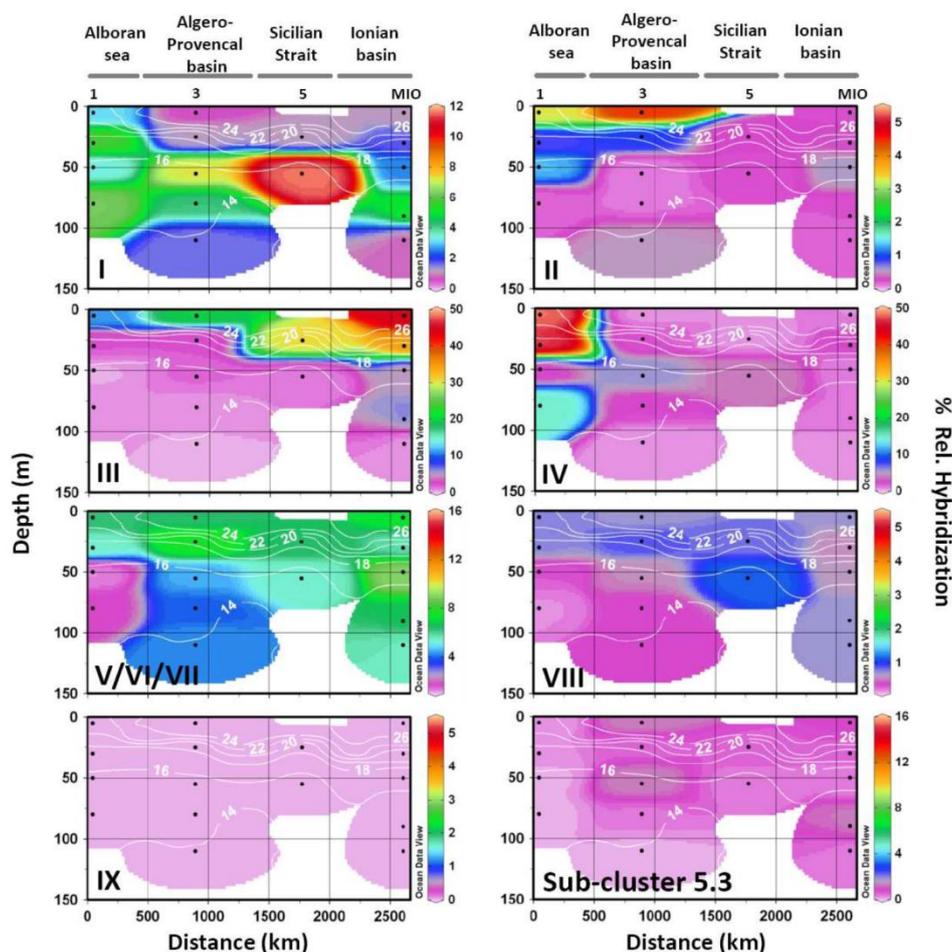
The set of clade-specific oligonucleotide probes designed by Fuller et al. (2003), which targets the nine major marine *Synechococcus* clades of sub-cluster 5.1 (i.e. individual probes against clades I, II, III, IV, VIII and IX, and a single probe against clades V, VI and VII) as well as sub-cluster 5.3 (formerly sub-cluster 5.1, clade X; see Dufresne et al., 2008), was used to study the distribution of these clades in the Mediterranean Sea at selected stations of the BOUM (Fig. 5) and PROSOPE cruises (Fig. 6), as well as at Sta. BOUSSOLE (See Fig. S1 in the Supplement).

Clade I was most abundant in the vicinity of the DCM at mesotrophic stations of the Gulf of Lions (i.e. Sta. 25 during BOUM, Sta. DYF during PROSOPE and the nearby Sta. BOUSSOLE during BOUSSOLE 88), and was therefore found in fairly cold waters (<17 °C). In the rest of the BOUM transect, clade I was sometimes found at low abundance at depth, generally within the deep *Synechococcus* maximum. Clade IV generally exhibited a similar distribution pattern to clade I. However, its relative abundance was

significantly higher than clade I in the northern part of the Algero-Provençal basin (Sta. 25 and BOUSSOLE) and lower at all other stations of the BOUM cruise. Similarly during the PROSOPE cruise, clade IV was dominant over clade I at Sta. 1 (Straits of Gibraltar) and the reverse at Sta. 5, where a significant abundance maximum of clade I cells occurred at 60 m (Fig. 6).

Clade III was present in the upper mixed layer at all stations of the BOUM transect with its maximum signal observed in surface waters. It was the predominant *Synechococcus* clade at Sta. 25, 5 and 9 (Fig. 5) and the hybridization signal extended throughout the top 100 m of most of the Levantine basin. Similarly, during the PROSOPE cruise, clade III was also the dominant clade in surface waters (except in the Alboran Sea) and its relative abundance as well as vertical extension increased eastwards (Fig. 6).

The probe targeting clades V, VI and VII gave strong relative hybridization signals (up to 15.5 %) at all stations and all depths examined during both the PROSOPE and BOUM cruises. During the latter cruise, it was the dominant *Synechococcus* group at most depths from Sta. A to B, with a



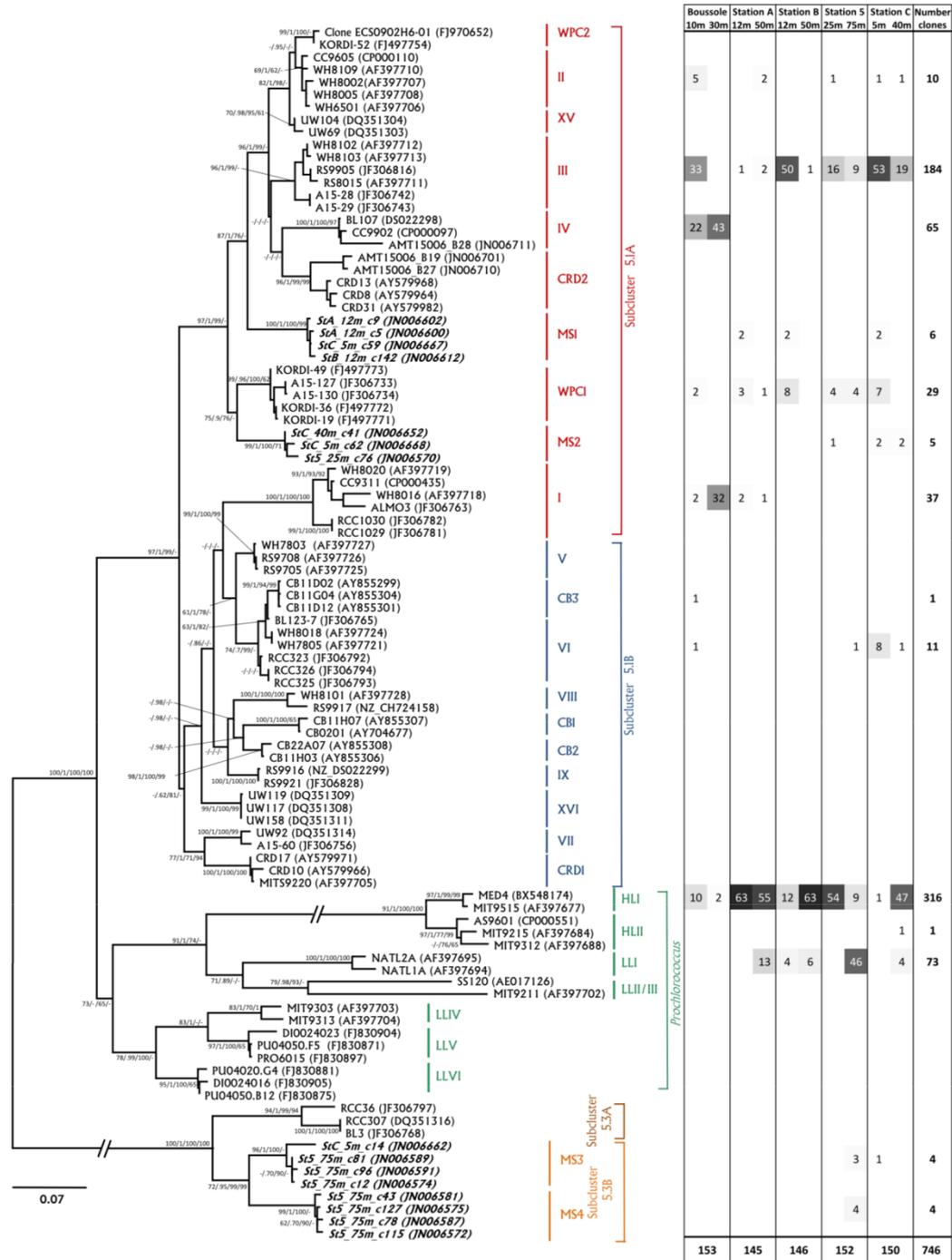
**Fig. 6.** Vertical distribution of different phylogenetic groups of *Synechococcus* along the W-E part of PROSOPE cruise track, as analyzed by dot blot hybridization using 16S rRNA gene specific oligonucleotide probes targeting sub-cluster 5.1 clades I to IX and sub-cluster 5.3. See Table 1 for correspondence to sequenced strains. For each panel, the x axis represents the cumulative distance in kilometers from a site W of Sta. 1 (36.2° N–5.8° W) and the y axis represents depth in meters. Details of panels are as described in Fig. 2. Note the different ranges of % relative hybridization values between panels.

maximum relative abundance found at the level of the *Synechococcus* density maximum.

The hybridization signals for clade II were never significant (<1.0%; Fig. 5) during the BOUM cruise. In contrast, during the PROSOPE cruise, low but significant clade II hybridization signals (up to 4.4%) were observed in surface waters of the southwestern Mediterranean basin (Fig. 6), corresponding to a population that likely originated from the Atlantic Ocean. Clade VIII was virtually undetectable at all examined stations of both BOUM and PROSOPE cruises. Clade IX was detectable only at the mesotrophic Sta. 25 during BOUM at all sampled depths (Fig. 5, Table 2) and at Sta. BOUSSOLE (not shown) but not during PROSOPE. Finally, cells belonging to sub-cluster 5.3 were only detected in the Levantine basin during BOUM, with highest abundance in the DCM, i.e. below the *Synechococcus* density maximum.

### 3.4 ITS clone libraries

In order to explore in greater detail the genetic diversity of the Mediterranean Sea populations of picocyanobacteria, we sequenced clone libraries generated from the ITS region using general primers targeting both *Prochlorococcus* and *Synechococcus* (Rocap et al., 2002). Clone libraries were obtained from two selected depths at four stations of the BOUM cruise (A, B, 5 and C), as well as two depths of the BOUSSOLE site, sampled in mid-June 2009. In order to maximize the number of *Synechococcus* sequences in the data set, the two depths were chosen in the upper mixed layer and above the *Prochlorococcus*-dominated DCM (hereafter called “surface” and “mid-depth” samples, respectively). We sequenced 69–80 clones from each library and performed phylogenetic analyses using four methods (ML, Bayesian,



**Fig. 7.** Reference Maximum Likelihood (ML) tree of ITS sequences (based on a sequence alignment of 1135 positions, including tRNA) and table summarizing the number of clones of the different *Synechococcus* and *Prochlorococcus* clades/ecotypes for each clone library obtained from samples collected during the BOUM cruise. Novel environmental groups discovered during BOUM (MS1–4) were included in the reference tree and are shown in bold italic letters. The nomenclature used to name the BOUM samples is as follows: stX\_Ym\_cZ where X indicates the station number, Y, the sampling depth in meters and Z, the clone number. For sequence submission, the prefix “BOUM.” has been added in front of the sequence name. The clade/ecotype number or name as reported in the literature (Ahlgren and Rocap, 2006; Chen et al., 2006; Choi and Noh, 2009; Fuller et al., 2003; Dufresne et al., 2008) are indicated in colored letters for each group. Sequences belonging to the *Synechococcus* marine sub-cluster 5.3 were used as an outgroup. Numbers at nodes of the phylogenetic tree correspond to bootstrap values for ML analyses, posterior probability of Bayesian method (ranging between 0 and 1), and bootstrap values for Neighbor-Joining and Parsimony methods, respectively. Bootstraps, represented as a percentage, were obtained through 1,000 repetitions and only values higher than 60% are shown. The scale represents 0.07 substitutions per nucleotide. Branches corresponding to the *Prochlorococcus* HL ecotypes and *Synechococcus* marine sub-cluster 5.3 have been cut for readability (their full length is 0.36 and 0.5 substitutions per nucleotide, respectively). The table on the right indicates the number of sequences in clone libraries constructed for each station and depths; their vertical position in the table corresponds to their clade affiliation in the tree, as determined by Pplacer (see text). Numbers in the table are highlighted with increasing degree of shading.

NJ and MP; Fig. 7). In only 2 out of 8 BOUM clone libraries (Sta. B at 12 m and Sta. C at 5 m) as well as at both depths at Sta. BOUSSOLE, the proportion of sequences attributable to *Synechococcus* was higher than *Prochlorococcus*, and the large variability of this ratio (from 74:1 to 1:69) likely reflected the variable relative abundances of these two genera in the starting samples. Environmental sequences from each individual library were positioned using Pplacer v1.0 (see methods) in a reference tree containing 97 ITS sequences from strains or environmental samples representative of the whole ITS diversity reported so far in the *Prochlorococcus* and *Synechococcus* groups, including four new groups described below (see Fig. 7 and Fig. S2 in the Supplement, generated using Pplacer for each library).

Most sequences related to *Prochlorococcus* were assignable to the HLI clade, with the majority of sequences originating from “surface” samples at Sta. BOUSSOLE, A and 5 and from “mid-depth” samples at Sta. B and C. *Prochlorococcus* LLI were also present in abundance, but only in “mid-depth” samples, except at Sta. B where 4 sequences were obtained from the ‘surface’ sample (Fig. 7). There were no sequence attributable to either LLII, LLIII or LLIV in the whole data set but, as mentioned earlier, this result was expected since these ecotypes are abundant only near the bottom of the euphotic zone.

For *Synechococcus*, clade III was the dominant group at most stations except in the “mid-depth” sample from BOUSSOLE and in surface waters of Sta. A (Fig. 7 and Fig. S2 in the Supplement). In contrast, many clade I and IV sequences were retrieved from Sta. BOUSSOLE and a few clade I sequences from Sta. A, but no sequence from either clade from the three stations of the eastern Mediterranean basin. The next most abundant group of sequences belonged to the WPC1 clade, a lineage recently described based on culture isolates from the east China Sea and East Sea (Choi and Noh, 2009). This group was systematically present in surface waters at all sites and sometimes extended to mid-depth. All other *Synechococcus* clades contributed for less than 12 sequences and can therefore be considered as minor (at least for the dataset considered here). These included 10 sequences attributable to clade II, originating from various stations and depths, with half coming from surface waters of Sta. BOUSSOLE. Eight clade VI sequences were retrieved from surface waters at Sta. C, while single sequences of this clade and the closely related clade CB3 (previously only observed in Chesapeake Bay; Chen et al., 2006) were retrieved from 10 m at BOUSSOLE.

Interestingly, sequences from four new, uncultured *Synechococcus* clades, so-called MS1 through MS4, were observed in clone libraries (Fig. 7). MS1 was inferred to be phylogenetically located at the base of the group formed by clades II (which includes groups XV and WPC2), III, IV and CRD2 (Ahlgren and Rocap, 2006). It is noteworthy, however, that in trees made with full-length ITS sequences, MS1 in fact grouped with clade III, though with low bootstrap sup-

port (see Fig. S2 in the Supplement). MS2 clustered with the above mentioned WPC1 clade using all phylogenetic methods except MP. The final two clades, MS3 and MS4, were related to one another and together formed a new group within sub-cluster 5.3. Given their phylogenetic position, supported by strong bootstrap values, and the relatively large distance between these two new clades and members of the initial sub-cluster 5.3, we propose to split the latter sub-cluster into two groups, so-called 5.3A and 5.3B (see Fig. 7).

MS1 was found both in the western and eastern basins (Sta. A, B and C) and comprised only “surface” sequences, whereas the three other uncultured clades were only retrieved in the eastern part of the transect, with the latter two mainly found in the *Synechococcus* abundance maximum.

## 4 Discussion

### 4.1 Hydrology-driven variations in picocyanobacterial population structure

Despite its relatively modest size, representing about 0.7% of the cumulative surface of marine areas, the Mediterranean Sea displays a large variety of hydrological conditions in summer, with a strong N–S gradient of SST (17–24 °C; Fig. 2a) and W–E gradient of salinity (37.4–39.6 PSU; Fig. 2b). Nutrient availability also varied tremendously along the BOUM transect, with a progressive N–S and W–E deepening of nutriclines, that was much steeper for the phosphocline than for the nitracline (Moutin et al., 2011; Pujol-Pay et al., 2011).

The spatial heterogeneity of hydrological conditions observed along the BOUM transect induced large variations in the vertical structure of both *Prochlorococcus* and *Synechococcus* populations, which seemed to respond differently to these physico-chemical changes. In the northern part of the Algero-Provencal basin, both genera formed subsurface abundance maxima and they either exhibited comparable cell densities (e.g. Sta. 27) or the latter genus outnumbered the former one (e.g. at Sta. 25 at all depths). This type of distribution is typical of mesotrophic areas (Partensky et al., 1996; Partensky et al., 1999). In the southern part of the Algero-Provencal basin, the Sicily Strait and the eastern Ionian Sea, *Prochlorococcus* was the dominant group at depth, while *Synechococcus* was most abundant in the upper mixed layer, with abundances usually above  $10^4$  cells ml<sup>-1</sup>. However, because *Prochlorococcus* cells could not be precisely enumerated in the top 50 m or so (see methods), we cannot ascertain which of the two picocyanobacterial genera numerically dominated in this layer. This near surface *Synechococcus* abundance maximum is somewhat unusual compared to other oligotrophic areas with surface chlorophyll concentrations  $<0.1$  µg Chl1<sup>-1</sup>, such as the Atlantic (Li et al., 1992; Olson et al., 1988; Olson et al., 1990; Zubkov et al., 2000) or Pacific gyres (Campbell et al., 1997; Grob

et al., 2007), where *Synechococcus* are generally more homogeneously distributed in the upper 100–150 m layer and maximal cell densities are typically 2–5 times lower. This raises the question of why do such high *Synechococcus* cell densities occur in surface waters of this nutrient-poor region of the Mediterranean Sea? Given the relatively close proximity of aerosol source emissions, which at the time of the BOUM cruise predominantly originated from the European continent (Ternon et al., 2011), it is possible that nutrient enrichment of surface waters by atmospheric dust deposition may have played a significant role in this phenomenon (Bonnet et al., 2005). Indeed, aerosols are enriched in inorganic nutrients (Fe, N and P, but with a N:P ratio usually much higher than the Redfield ratio of 16:1; Paytan et al., 2009) and natural, nutrient-starved *Synechococcus* populations can respond quickly to such enrichment by rapid increases in cell concentrations (Hutchins et al., 2003; Mackey et al., 2009; Tanaka et al., 2011, Ternon et al., 2011).

Most stations between sites A and B exhibited a N:P ratio between 20 and 40 (integrated over the top 150 m; see Crombet et al., 2011), indicating mild P-depletion in these waters. Oceanic strains of *Synechococcus* (e.g. WH8103 and WH8012) display a cellular N:P ratio in the same range (24–43), and therefore have a low P requirement, which can even be further reduced by 3- or 4-fold when grown under P depletion (Bertilsson et al., 2003). Additionally, it was previously shown that *Synechococcus* populations from the Mediterranean Sea have a high affinity for orthophosphate and an elevated P uptake rate (Moutin et al., 2002). Thus, *Synechococcus* seems to be particularly well adapted to withstand the moderate P depletion occurring between Sta. A and B. Nevertheless, east of Sta. B, the intensity of P depletion seemingly increased since integrated N:P ratios were between 50 to more than 100 at most stations (Crombet et al., 2011). This was confirmed by nutrient enrichment experiments using surface water, which showed that *Synechococcus* cells were clearly co-limited by both N and P at Sta. C, whilst they were only N-starved at Sta. A and B (even though, at the latter station, N+P addition induced higher cell concentrations than N only; Tanaka et al., 2011).

This increased P depletion in eastern waters might explain in part the dramatic change in picocyanobacterial population structure observed between the Ionian sea and the Levantine basin. Indeed, in the easternmost region, we observed very low concentrations of both genera in the upper mixed layer (as indirectly indicated for *Prochlorococcus* by undetectable levels of its specific pigment divinyl-Chl *a*; Fig. 2f), and deep abundance maxima of both groups, located around 75 m for *Synechococcus* and 100 m for *Prochlorococcus*. These atypical vertical profiles might also be partially explained by factors other than macronutrient depletion, and in particular by copper toxicity, since the annual Cu deposition rate due to atmospheric aerosols is quite high in the eastern Mediterranean Sea ( $1\text{--}2\text{ mg m}^{-2}\text{ yr}^{-1}$ ; Paytan et al., 2009). Indeed, *Prochlorococcus* growth was shown to be strongly

affected by fairly low concentrations of this element, while *Synechococcus* cells can usually withstand higher concentrations (Mann et al., 2002). A more recent report claimed that some Cu-rich Saharan dust had a more severe toxic effect on the growth of *Synechococcus* in comparison to *Prochlorococcus* growing in the Red Sea (Paytan et al., 2009). It is therefore possible that, after events of dust deposition, abundances of both picocyanobacteria in the upper mixed layer of the eastern Mediterranean Sea may strongly vary over short time scales, in a direction depending on the nature and origin of aerosols.

The population structure of marine picocyanobacteria was globally comparable along common parts of the BOUM and PROSOPE cruises, despite the fact that sampling was held on different months and about a decade apart. It is worth noting though that in surface waters south of the Algero-Provençal basin and the Sicily Strait, *Synechococcus* abundances were two-fold lower during BOUM than during PROSOPE and the corresponding integrated cell concentrations were also 20–40% lower (Table 2 and data not shown). Whether this results merely from seasonal differences or translates some more profound change in the degree of oligotrophy of the area still needs to be determined.

## 4.2 Factors determining the distribution of picocyanobacterial ecotypes

### 4.2.1 *Prochlorococcus*

A number of recent studies have shown that the phylogenetic composition of *Prochlorococcus* populations is tightly driven by environmental factors, primarily by light and secondarily by temperature. When the water column is stratified, there is a clear vertical partitioning of ecotypes, with HL (i.e. either HLI or HLII) occupying the upper layer, LLI an intermediate layer (generally the thermocline depth) and LLII-IV the base of the lit layer (Johnson et al., 2006; Zinser et al., 2006). Interestingly, this vertical partitioning has a genomic basis, since HL and LLI strains possess a number of adaptations to the deleterious effects of high irradiance and UV stress (including DNA repair genes, such as photolyases) that are absent in “strictly LL” ecotypes, i.e., LLII-IV (Kettler et al., 2007; Partensky and Garczarek, 2010; Rocap et al., 2003). In the present study, the LL-specific oligonucleotide probe used for dot blot hybridization analysis did not allow us to distinguish LLI (or eNATL) from LLIV (or eMIT9313) ecotypes, but the fact that ITS clone libraries from “mid-depth” (30 to 75 m; see Fig. 7) contained many LLI and no “strictly LL” sequences clearly indicated that a vertical partitioning of ecotypes comparable to that previously observed in the Atlantic and Pacific Ocean (Johnson et al., 2006; Malmstrom et al., 2010) is also occurring in the Mediterranean Sea. Surprisingly, the LLII- (or eSS120-) specific probe detected significant levels of this ecotype only at the northwestern Mediterranean Sea stations (Sta. DYF during PROSOPE,

Sta. 25 during BOUM and Sta. BOUSSOLE during BOUSSOLE 88). This suggests that the distribution of LLII in the Mediterranean Sea is restricted to temperate, mesotrophic areas. However, this clade was found to be abundant at depth at both long-term stations BATS (Bermuda Atlantic Time Series) and HOT (Hawaii Ocean Time series; Malmstrom et al., 2010) and its occurrence in such tropical and subtropical, oligotrophic waters somewhat contradicts this assumption. So it is probable that our LLII-specific probe in fact did not target exactly the same population as did the qPCR primers used by Malmstrom and coauthors.

The second most important factor determining *Prochlorococcus* ecotype distribution is temperature. Indeed, a clear latitudinal partitioning of HLI and HLII ecotypes was observed during two consecutive AMT transects through the Atlantic ocean in late summer/early fall (Johnson et al., 2006; Zwirgmaier et al., 2007). Although both HL clades co-occurred at low latitudes, HLII was always the dominant clade from 28° S to 33° N; then, the ratio was progressively reversed at higher latitude, so that HLI became dominant above 32° S or 38° N. If one extrapolates the HLI and HLII concentrations reported in the N Atlantic by Johnson et al. (2006) to the latitudinal range covered by the BOUM and PROSOPE transects in the Mediterranean Sea (33.6–43.2° N), one should have expected to find the HLI:HLII abundance ratio ranging roughly from 1:1 in the Ionian and Levantine basins to 2000:1 at the northernmost Sta. 27. However, the HLII hybridization signal was most often lower than 1% of the EUB signal, except in subsurface waters at PROSOPE Sta. 5 (see Fig. 3 in Garczarek et al., 2007). This low abundance of the HLII ecotype in the Mediterranean Sea was independently confirmed by the occurrence of only one HLII sequence (at Sta. C) in our ITS clone libraries compared to 316 HLI sequences.

The scarcity of the HLII ecotype in the southern part of the Mediterranean Sea, in particular in the Levantine basin, is somewhat puzzling. Indeed, the growth temperature range of strains representative of this ecotype (MIT9312 and MIT9215) is 16–30 °C, with growth optima between 23 and 28 °C. In contrast, HLI strains (MED4 and MIT9515) can grow in the 11–28 °C range, with growth optima between 22 and 26 °C (Johnson et al., 2006; Zinser et al., 2007). In field populations, temperature vs. abundance curves differ even more dramatically between the two HL ecotypes: HLII abundance increases steadily with temperature from less than 10<sup>2</sup> cells ml<sup>-1</sup> at 14 °C up to a plateau of 1–4 × 10<sup>5</sup> cells ml<sup>-1</sup> at 25–28 °C, whereas HLI abundance is already high (ca. 10<sup>3</sup> cells ml<sup>-1</sup>) at 14 °C, peaks around 19 °C then declines drastically above 25 °C (Zinser et al., 2007).

The average SST over the 1999–2008 period for the coldest and warmest months (i.e. March and August, respectively) were 14.5–25.8 °C, 16.6–27.6 °C and 17.3–28.0 °C for Sta. A, B and C, respectively (Fig. 2a–c). Although the wintertime SST at Sta. A was close to the lower limit of the HLII growth temperature range, suggesting that this ecotype

might not be able to survive year-round in these waters, the SST range recorded at Sta. B and C were completely compatible with its growth temperature range. Furthermore, the Mediterranean Sea is connected to the Red Sea via the Suez Canal and it was shown that the dominant *Prochlorococcus* ecotype year-round in the Gulf of Aqaba is HLII (Fuller et al., 2005), constituting a possible reservoir of HLII cells for the southeastern Mediterranean Sea. It is even possible that coastal waters of the Levantine basin might display a higher HLII to HLI ecotype ratio than offshore waters, as suggested by a recent analysis of a microbial community metagenome collected in surface waters near Israeli coasts (Feingersch et al., 2010).

Thus, to explain the scarcity of HLII cells in open Levantine surface waters in summer, despite seemingly optimal temperatures for growth, we hypothesize that, in this area, the upper mixed layer exhibits a combination of features that make these waters unfavorable to growth of *Prochlorococcus* HL populations in general, as attested by undetectable divinyl-Chl *a* concentrations (see Fig. 2). These features likely include extremely low macronutrient concentrations, triggering N and P co-limitation of growth, and/or high and potentially toxic levels of copper (Mann et al., 2002; Paytan et al., 2009), though some other yet-to-be identified factors may also be involved. We further assume that local HLI populations are able to survive at low cell concentrations under these stress conditions, whereas alien HLII populations entering the Mediterranean Sea via the Suez Canal cannot, and are rapidly outcompeted and eliminated. It would be interesting to determine whether conditions may become more favorable in the Levantine basin at another time of the year by following the abundance of *Prochlorococcus* populations on a monthly scale, since such time series performed at the long-term Sta. HOT (off Hawaii) and BATS (off Bermuda) revealed strong seasonal variations of the relative abundance of the different ecotypes, with distinct annual patterns for each ecotype (Malmstrom et al., 2010).

#### 4.2.2 *Synechococcus*

Whilst for *Prochlorococcus* the same phylogenetic groups were found at comparable relative depths from one end of the Mediterranean Sea to the other (Fig. 4; see also Garczarek et al., 2007), we observed large variations in the genetic diversity of *Synechococcus* populations both vertically and horizontally along the BOUM and PROSOPE cruise transects (Figs. 5 and 6). It must be noted that the two molecular approaches used here to assess genetic diversity within this genus provided slightly different pictures of the *Synechococcus* community structure. For instance, dot blot hybridization seemingly indicated that the two most abundant *Synechococcus* groups in the Mediterranean Sea were clade III and the phylogenetic cluster formed by the closely related clades V/VI/VII. It is worth noting though, that the clade V/VI/VII probe also recognizes *Synechococcus*

strain MITS9220, which is phylogenetically affiliated to the CRD1 clade (Fig. 7). While the predominance of clade III was confirmed by clone library sequencing, since the majority of *Synechococcus* ITS sequences belonged to this clade, there were fewer sequences attributable to clades V, VI (and the closely related group CB3), VII and/or CRD1 than would have been expected from hybridization results (Fig. 7). Moreover, there is a single mismatch between this probe and sequences of the recently described clade WPC1 (Choi and Noh, 2009), which appears to be widely distributed in the Mediterranean Sea (Fig. 7). It is therefore possible that the clade V/VI/VII probe can also recognize this particular group. In any case, cells targeted by this probe were found everywhere in the Mediterranean Sea, with highest abundances (up to 15.5 % relative hybridization) in the nutrient-rich waters of the northern Algero-Provençal basin (Figs. 5 and 6). This is consistent with previous literature on other regions of the ocean, since this group was observed throughout the AMT transect in the Atlantic Ocean, with the highest concentrations in temperate waters of the southern and northern hemispheres (e.g., up to 28 % relative hybridization off Cape Town) and in upwelling waters of the northern Arabian Sea (up to 53 % relative hybridization; Zwirgmaier et al., 2007, 2008).

Clade III was most abundant in the surface layer at all stations of the western basin and the Ionian Sea (Fig. 5). This result was confirmed by the much larger number of clade III sequences obtained in clone libraries from “surface” than “mid-depth” (except at Sta. A, where only three sequences were retrieved at this station). This is consistent with earlier studies which showed that clade III was preferentially found in the upper mixed layer (Toledo et al., 1999; Toledo and Palenik, 2003). However, in the Levantine basin, its depth distribution extended to the whole euphotic layer and this seems to be directly related to the deepening of the *Synechococcus* density maximum in this area (Fig. 2e). Zwirgmaier et al. (2008) previously suggested that clade III was an “oligotrophic” clade, since they observed that members of this clade were confined to a fairly narrow window of nitrate and phosphate concentration, even though they displayed no obvious latitudinal preference. Genomic analysis of the clade III isolate *Synechococcus* sp. WH8102, indeed shows that it contains a large number of nutrient transporters but a reduced regulatory machinery (Dufresne et al., 2008; Palenik et al., 2003; Scanlan et al., 2009), strengthening the idea that members of this clade are well adapted to stable, low-nutrient environments. Moreover, this strain exhibits a well-developed cellular response to P-stress that includes the expression of genes encoding high-affinity periplasmic P-binding proteins and at least three organic phosphatases that could provide a competitive advantage for scavenging P from alternative sources when inorganic P is scarce (Ostrowski et al., 2010; Tetu et al., 2009).

Clades I and IV were both most abundant in mesotrophic waters of the Mediterranean Sea (though with higher signals for clade IV at all depths), suggesting that both clades are adapted to high-nutrient/low-temperature waters. Faint hybridization signals were also observed in the southern Mediterranean Sea, with maximal values for both clades often occurring at the deep *Synechococcus* maximum, which was generally found below the 18 °C isoline (Fig. 2), confirming the preference of these clades for cool waters. The co-occurrence of *Synechococcus* clades I and IV has been well documented in the Atlantic and Pacific Oceans, where they constitute the dominant *Synechococcus* groups at latitudes above 30° N/S (Brown and Fuhrman, 2005; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008; Tai et al., 2011; Tai and Palenik, 2009). Nevertheless, the ratio of abundances of the two clades may vary both spatially and temporally. Indeed, it was observed at a coastal site off South California that these clades displayed different seasonal cycles (Tai and Palenik, 2009). Whilst clade IV was the dominant type over most of the year, clade I outnumbered clade IV just before the annual abundance peak of *Synechococcus* cell density. Genome sequences of one clade I strain (CC9311) and two clade IV strains (CC9902 and BL107) have been obtained so far (Dufresne et al., 2008; Palenik et al., 2006; Scanlan et al., 2009). The former has a slightly larger genome than the latter, and encodes a larger set of proteins involved in signal sensing (such as histidine kinases and response regulators) and metal binding (including multiple genome encoded copies of ferritin and Zn-binding metallothioneins). Thus, clade I might have been expected to be better adapted to the variability experienced by coastal/mesotrophic environments than its clade IV counterpart, which somewhat contradicts the higher relative abundance of clade IV over clade I in nutrient-rich areas (Tai and Palenik, 2009; this study). Hence, understanding the factors that allow the co-occurrence of clades I and IV, but also the dominance of one clade over the other still awaits extensive comparisons of genomes from many more strains of both clades and/or refined comparative physiological studies in culture (including co-cultures).

The presence of sub-cluster 5.3 at depth in the Levantine basin during BOUM, with the highest abundance observed at Sta. 5, is consistent with ITS sequencing, since seven sequences of this group were retrieved at this station from the 75 m sample. Previously, this group (ex sub-cluster 5.1, clade X) had only been reported from subsurface samples in warm, coastal waters of the Arabian Sea, i.e., in a fairly different niche (Fuller et al., 2003; Fuller et al., 2006). Yet, we showed here that this sub-cluster is more diverse than previously thought and thus it is not unexpected that, like for sub-cluster 5.1 (Dufresne et al., 2008; Scanlan et al., 2009), different clades within this group might be adapted to different habitats.

Both dot blot hybridization and clone library sequencing suggest that clades II, VIII and IX are only minor components of the *Synechococcus* community in the Mediterranean Sea. So far, all characterized clade VIII strains, including a number isolated from the Gulf of Aqaba (Fuller et al., 2003; Fuller et al., 2006), are phycocyanin-containing halotolerant isolates, and this may explain why members of this clade have never been found in significant numbers in open ocean samples (Zwirgmaier et al., 2008). Clade IX has only been found at low levels in warm and coastal sub-surface waters of the northern Arabian Sea (Fuller et al., 2006) and was undetectable elsewhere, so its preferred niche remains unclear. The low abundance of clade II in the Mediterranean Sea, suggested by dot blot hybridization and confirmed by the occurrence of few ITS sequences in clone libraries, is somewhat more surprising. In the California Current (Toledo and Palenik, 2003), the Red Sea (Fuller et al., 2003), the northern part of the Arabian Sea (Fuller et al., 2006) and the region of the Morocco upwelling (Zwirgmaier et al., 2007), this clade constituted a large proportion of the *Synechococcus* population (up to 75 % relative hybridization) in surface waters in tropical and subtropical areas. Indeed, Zwirgmaier et al. (2008) showed that this clade was particularly well represented in warm waters, ranging from 22 °C to 28 °C, while other clades either possess a larger temperature tolerance (7–28 °C for clades III and V/VI/VII) or are restricted to cooler waters (7–18 °C for clades I and IV). Thus, one might have expected to find notable abundances of clade II in the Mediterranean Sea, at least at the easternmost Sta. C, where SST reaches 27 °C in summer (Fig. 2a). Like for *Prochlorococcus* HLII, some environmental factors other than temperature possibly inhibit the proliferation of this clade, at least during the warmest months of the year.

## 5 Conclusions

A number of recent studies suggest that a global increase in the temperature of the Mediterranean Sea has occurred in the last few decades (Saaroni et al., 2003 and references therein) and that this oceanic region is particularly vulnerable to global change (see e.g., Belkin, 2009). However, the intensity of this rise may vary locally, as shown here by differences in multi-annual trends in summer SST anomalies over the 1985–2010 period, in the vicinity of the three long-term stations of the BOUM cruise, located in the center of distinct anticyclonic eddies. Interestingly, the increasing trend was only significant in the eastern Mediterranean Sea and reached 0.5 °C per decade in the Levantine basin. This is consistent with climate models, which forecast that by the end of the 21st century the Mediterranean Sea SST's will be, on average, 3 to 4 °C warmer than today (Coll et al., 2010; Hertig and Jacobeit, 2008; Somot et al., 2006). Even if these multi-annual SST variations may seem relatively modest compared to seasonal SST variations (Fig. 3), this change

could potentially induce drastic changes in the composition of marine microbial communities (including picocyanobacteria) at mid (decades) or long term (centuries). Indeed, a rise in the annual temperature range by only a few degrees should theoretically favor subtropical groups over local populations, leading to their progressive replacement.

Here, we studied the genetic diversity and distribution of *Prochlorococcus* and *Synechococcus* clades/ecotypes during the PROSOPE and BOUM cruises held almost a decade apart, using two different and complementary molecular approaches. For *Prochlorococcus*, there was a large phylogenetic diversity down the water column at any given site but the same ecotypes were found from one end of the Mediterranean Sea to the other. The dominant clade in the upper mixed layer was HLI during both cruises. Thus, even though the Mediterranean Sea is located in the boundary zone between the distribution areas of HLI and HLII ecotypes and we have evidence that HLII is present in this area, changes that have occurred in the hydrology of the Mediterranean Sea as a result of global warming, have not been sufficient yet to provoke a significant change in the dominant HL population, even though the annual SST range in the Levantine basin is seemingly suitable for HLII growth. The Eastern Mediterranean Sea is known to have a number of unique hydrological features compared to other oligotrophic areas, and our results suggest that environmental factors other than temperature (and light, if one considers the LL ecotypes) must be taken into account to reliably predict any future changes in the relative abundances of *Prochlorococcus* ecotypes.

For *Synechococcus*, analyses of the distribution of the different clades reported here for the Mediterranean Sea strengthened generalizations made by previous studies in other parts of the world ocean, including the fact that clades I, III and IV are locally the most abundant groups, with clades I and IV co-occurring and being restricted to cool waters, while clade III is ubiquitously found in oligotrophic areas (Bouman et al., 2006; Johnson et al., 2006; Zwirgmaier et al., 2007; Zwirgmaier et al., 2008). However, our study also revealed the presence in the Mediterranean Sea of novel groups (MS1, MS2 and sub-cluster 5.3B) and the large relative abundance of cells detected by the probe targeting clades V/VI/VII/CRD1. These data clearly point to a need for the design of further probes, or the use of alternate approaches, such as quantitative PCR (Tai and Palenik, 2009), hybridization to liquid bead-arrays (Tai et al., 2011) or high throughput sequencing using a recently developed functional gene marker (*petB*) with high taxonomic resolution (Mazard et al., 2011), which allow precise mapping of the distribution of individual clades or genotypes in the environment.

Our study advocates the interest of bioindicators (such as *Prochlorococcus* HLII or the *Synechococcus* clade II) to complement the classical physico-chemical parameters usually measured to monitor the impact of global changes on marine ecosystems. It also points out the need for time series analyses of the microbial diversity in the Mediterranean Sea, especially in the Levantine basin, which is seemingly the most affected by global warming and which exhibits a number of atypical hydrological traits for an oligotrophic area. Such studies would not only show whether the different taxonomic groups display different annual patterns of abundance (Malmstrom et al., 2010; Tai and Palenik, 2009), but also help unveil what factors, besides temperature, control the composition of these microbial communities.

**Supplementary material related to this article is available online at:**

<http://www.biogeosciences.net/8/2785/2011/bg-8-2785-2011-supplement.pdf>.

*Acknowledgements.* This work was supported by the French ANR program PELICAN (PCS-09-GENM-200) and the PROSOPE and BOUM projects of the LEFE-CYBER program. DMF was supported by the National Commission of Scientific and Technological Investigation of Chile (CONICYT), SM was in receipt of a Natural Environment Research Council (NERC) PhD studentship and MO was supported by NERC grant NE/F004249/1 awarded to DJS. FM and EMB were respectively supported by the European programmes ERA-net BiodivERSA (BioMarKs project) and EPOCA (grant 211384). We thank Thierry Moutin for providing us access to BOUM cruise data and to several manuscripts prior to publication. The captains and crews are kindly acknowledged for their help during the PROSOPE, BOUM (both on board the R/V *L'Atalante*) and BOUSSOLE 88 cruises (on board the R/V *Tethys II*). Emilie Diamond is warmly thanked for her help during the BOUSSOLE88 cruise as well as Claudie Marec, Louis Prieur and Marc Picheral for CTD operations during BOUM. We are also grateful to Nyree West for generating dot blots from the DYF site of PROSOPE, Morgan Perennou and Delphine Naquin for their help with ITS sequencing and data treatment and Vincent Moron and Dominique Davoult for helpful discussions on temperature data analysis.

Edited by: T. Moutin

## References

- Ahlgren, N. A. and Rocap, G.: Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies, *Appl. Environ. Microbiol.*, 72, 7193–7204, 2006.
- Amann, R. I., Krumholz, L., and Stahl, D. A.: Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic and environmental studies in microbiology, *J. Bacteriol.*, 172, 762–770, 1990.
- Belkin, I. M.: Rapid warming of large marine ecosystems, *Prog. Oceanogr.*, 81, 207–213, 2009.
- Bertilsson, S., Berglund, O., Karl, D. M., and Chisholm, S. W.: Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea, *Limnol. Oceanogr.*, 48, 1721–1731, 2003.
- Bianchi, C. N.: Biodiversity issues for the forthcoming tropical Mediterranean Sea, *Hydrobiologia*, 580, 7–21, 2007.
- Bonnet, S., Guieu, C., Chiaverini, J., Ras, J., and Stock, A.: Impact of atmospheric inputs on the autotrophic communities in a low nutrient low chlorophyll system, *Limnol. Oceanogr.*, 50, 1810–1819, 2005.
- Bouman, H. A., Ulloa, O., Scanlan, D. J., Zwirgmaier, K., Li, W. K., Platt, T., Stuart, V., Barlow, R., Leth, O., Clementson, L., Lutz, V., Fukasawa, M., Watanabe, S., and Sathyendranath, S.: Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes, *Science*, 312, 918–921, 2006.
- Brown, M. V. and Fuhrman, J. A.: Marine bacterial microdiversity as revealed by internal transcribed spacer analysis, *Aquat. Microb. Ecol.*, 41, 15–23, 2005.
- Campbell, L., Nolla, H. A., and Vaulot, D.: The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean, *Limnol. Oceanogr.*, 39, 954–961, 1994.
- Campbell, L., Liu, H. B., Nolla, H. A., and Vaulot, D.: Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991–1994 ENSO event, *Deep Sea Res. Pt. I*, 44, 167–192, 1997.
- Chen, F., Wang, K., Kan, J., Suzuki, M. T., and Wommack, K. E.: Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S–23S rRNA internal transcribed spacer sequences, *Appl. Environ. Microbiol.*, 72, 2239–2243, 2006.
- Choi, D. H. and Noh, J. H.: Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea, *FEMS Microbiol. Ecol.*, 69, 439–448, 2009.
- Coleman, M. L. and Chisholm, S. W.: Code and context: *Prochlorococcus* as a model for cross-scale biology, *Trends Microbiol.*, 15, 399–407, 2007.
- Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., Ballesteros, E., Bianchi, C. N., Corbera, J., Dailianis, T., Danovaro, R., Estrada, M., Froggia, C., Galil, B. S., Gasol, J. M., Gertwagen, R., Gil, J., Guilhaumon, F., Kesner-Reyes, K., Kitsos, M. S., Koukouras, A., Lampadariou, N., Laxamana, E., Lopez-Fe de la Cuadra, C. M., Lotze, H. K., Martin, D., Mouillot, D., Oro, D., Raicevich, S., Rius-Barile, J., Saiz-Salinas, J. I., San Vicente, C., Somot, S., Templado, J., Turon, X., Vafidis, D., Villanueva, R., and Voultsiadou, E.: The biodiversity of the Mediterranean Sea: estimates, patterns, and threats, *PLoS One*, 5, e11842, 2010.
- Crombet, Y., Leblanc, K., Quéguiner, B., Moutin, T., Rimmelin, P., Ras, J., Claustre, H., Leblond, N., Oriol, L., and Pujo-Pay, M.: Deep silicon maxima in the stratified oligotrophic Mediterranean Sea, *Biogeosciences*, 8, 459–475, doi:10.5194/bg-8-459-2011, 2011.
- D'Ortenzio, F. and Ribera d'Alcalá, M.: On the trophic regimes of the Mediterranean Sea: a satellite analysis, *Biogeosciences*, 6, 139–148, doi:10.5194/bg-6-139-2009, 2009.
- Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., Paulsen, I. T., Tandeau de Marsac, N., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A. F., Hess, W. R.,

- and Partensky, F.: Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria, *Genome Biol.*, 9, R90, 2008.
- DuRand, M. D., Olson, R. J., and Chisholm, S. W.: Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea, *Deep Sea Res. Pt. II*, 48, 1983–2003, 2001.
- Eguchi, M., Ostrowski, M., Fegatella, F., Bowman, J., Nichols, D., Nishino, T., and Cavicchioli, R.: *Sphingomonas alaskensis* strain AFO1, an abundant oligotrophic ultramicrobacterium from the North Pacific, *Appl. Environ. Microbiol.*, 67, 4945–4954, 2001.
- Fingersch, R., Suzuki, M. T., Shmoish, M., Sharon, I., Sabeji, G., Partensky, F., and Béjà, O.: Microbial community genomics in eastern Mediterranean Sea surface waters, *ISME J.*, 4, 78–87, 2010.
- Felsenstein, J.: PHYLIP – Phylogeny inference package (version 3.2), *Cladistics*, 5, 164–166, 1989.
- Fuller, N. J., Marie, D., Partensky, F., Vault, D., Post, A. F., and Scanlan, D. J.: 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, 2003.
- Fuller, N. J., West, N. J., Marie, D., Yallop, M., Rivlin, T., Post, A. F., and Scanlan, D. J.: Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea, *Limnol. Oceanogr.*, 50, 363–375, 2005.
- Fuller, N. J., Tarran, G. A., Yallop, M., Orcutt, K. M., and Scanlan, D. J.: Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages, *Limnol. Oceanogr.*, 51, 2515–2526, 2006.
- Garczarek, L., Dufresne, A., Rousvoal, S., West, N. J., Mazard, S., Marie, D., Claustre, H., Raimbault, P., Post, A. F., Scanlan, D. J., and Partensky, F.: High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer, *FEMS Microbiol. Ecol.*, 60, 189–206, 2007.
- Goericke, R. and Welschmeyer, N. A.: The marine prochlorophyte *Prochlorococcus* contributes significantly to phytoplankton biomass and primary production in the Sargasso Sea, *Deep Sea Res. Pt. I*, 40, 2283–2294, 1993.
- Grob, C., Ulloa, O., Li, W. K. W., Alarcon, G., Fukasawa, M., and Watanabe, S.: Picoplankton abundance and biomass across the eastern South Pacific Ocean along latitude 32.5 degrees S, *Mar. Ecol. Progr. Ser.*, 332, 53–62, 2007.
- Guindon, S. and Gascuel, O.: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *System Biol.*, 52, 696–704, 2003.
- Hall, T. A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucl. Acids Symp. Ser.*, 41, 95–98, 1999.
- Hertig, E. and Jacobeit, J.: Downscaling future climate change: Temperature scenarios for the Mediterranean area, *Global Planet. Change*, 63, 127–131, 2008.
- Huelsenbeck, J. P. and Ronquist, F.: MRBAYES: Bayesian inference of phylogenetic trees, *Bioinformatics*, 17, 754–755, 2001.
- Hutchins, D. A., Pustizzi, F., Hare, C. E., and DiTullio, G. R.: A shipboard natural community continuous culture system for ecologically relevant low-level nutrient enrichment experiments, *Limnol. Oceanogr. Meth.*, 1, 82–91, 2003.
- Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M., and Chisholm, S. W.: Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients, *Science*, 311, 1737–1740, 2006.
- Katoh, K. and Toh, H.: Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework, *BMC Bioinfo.*, 9, 212, 2008.
- Kettler, G., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G., Richardson, P., and Chisholm, S. W.: Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*, *PLoS Genet.*, 3, e231, 2007.
- Lavin, P., González, B., Santibáñez, J., Scanlan, D., and Ulloa, O.: Novel lineages of *Prochlorococcus* dominate the picocyanobacterial community within the oxygen minimum zone of the eastern tropical South Pacific, *Environ. Microbiol. Rep.*, 2, 728–738, 2010.
- Lejeune, C., Chevaldonné, P., Pergent-Martini, C., Boudouresque, C. F., and Pérez, T.: Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea, *Trends Ecol. Evol.*, 25, 250–260, 2009.
- Li, W. K. W., Dickie, P. M., Irwin, B. D., and Wood, A. M.: Biomass of bacteria, cyanobacteria, prochlorophytes and photosynthetic eukaryotes in the Sargasso Sea, *Deep Sea Res. Pt. I*, 39, 501–519, 1992.
- Li, W. K. W.: Primary productivity of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting, *Limnol. Oceanogr.*, 39, 169–175, 1994.
- Li, W. Z. and Godzik, A.: Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics* 22, 1658–1659, 2006.
- Lindell, D. and Post, A. F.: Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat), Red Sea, *Limnol. Oceanogr.*, 40, 1130–1141, 1995.
- Mackey, K. R. M., Rivlin, T., Grossman, A. R., Post, A. F., and Paytan, A.: Picophytoplankton responses to changing nutrient and light regimes during a bloom, *Mar. Biol.*, 156, 1531–1546, 2009.
- Malmstrom, R. R., Coe, A., Kettler, G. C., Martiny, A. C., Frias-Lopez, J., Zinser, E. R., and Chisholm, S. W.: Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans, *ISME J.*, 4, 1252–1264, 2010.
- Mann, E. L., Ahlgren, N., Moffett, J. W., and Chisholm, S. W.: Copper toxicity and cyanobacteria ecology in the Sargasso Sea, *Limnol. Oceanogr.*, 47, 976–988, 2002.
- Marie, D., Brussaard, C., Partensky, F., and Vault, D.: Flow cytometric analysis of phytoplankton, bacteria and viruses, in: *Current Protocols in Cytometry*, edited by: Sons, J. W., International Society for Analytical Cytology, 11.11.11–11.11.15., 1999.
- Marie, D., Zhu, F., Balagué, V., Ras, J., and Vault, D.: Eukaryotic picoplankton communities of the Mediterranean Sea in summer assessed by molecular approaches (DGGE, TTGE, QPCR), *FEMS Microbiol. Ecol.*, 55, 403–415, 2006.
- Martiny, A. C., Tai, A. P., Veneziano, D., Primeau, F., and Chisholm, S. W.: Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*, *Environ. Microbiol.*, 11, 823–832, 2009.
- Marty, J.-C., Chiavérini, J., Pizay, M. D., and Avril, B.: Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999) *Deep Sea Res. Pt. II*, 49, 1965–1985, 2002.

- Massana, R., Balague, V., Guillou, L., and Pedros-Alio, C.: Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches, *FEMS Microbiol. Ecol.*, 50, 231–243, 2004.
- Matsen, F. A., Kodner, R. B., and Armbrust, E. V.: Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree, *BMC Bioinfo.*, 11, 538, 2010.
- Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D. J.: Multiquence analysis, taxonomic resolution and biogeography of marine *Synechococcus*, *Environ. Microbiol.*, <http://onlinelibrary.wiley.com/doi/10.1111/j.1462-2920.2011.02514.x/full>, 2011.
- Moutin, T., Thingstad, T. F., Van Wambeke, F., Marie, D., Slawyk, G., Raimbault, P., and Claustre, H.: Does competition for nanomolar phosphate supply explain the predominance of the cyanobacterium *Synechococcus*?, *Limnol. Oceanogr.*, 47, 1562–1567, 2002.
- Moutin, T., Van Wambeke, F., and Prieur, L.: The Biogeochemistry from the Oligotrophic to the Ultraoligotrophic Mediterranean (BOUM) experiment, *Biogeosciences Discuss.*, 8, 8091–8160, doi:10.5194/bg-8-8091-2011, 2011.
- Needleman, S. B. and Wunsch, C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J. Mol. Biol.*, 48, 443–453, 1970.
- Obernosterer, I., Christaki, U., Lefèvre, D., Catala, P., Van Wambeke, F., and Lebaron, P.: Rapid bacterial mineralization of organic carbon produced during a phytoplankton bloom induced by natural iron fertilization in the Southern Ocean., *Deep Sea Res. Pt. II*, 55, 777–789, 2008.
- Olson, R. J., Chisholm, S. W., Zettler, E. R., and Armbrust, E. V.: Analysis of *Synechococcus* pigment types in the sea using single and dual beam flow cytometry, *Deep Sea Res. Pt. I*, 35, 425–440, 1988.
- Olson, R. J., Zettler, E. R., Armbrust, E. V., and Chisholm, S. W.: Pigment, size and distribution of *Synechococcus* in the North Atlantic and Pacific oceans, *Limnol. Oceanogr.*, 35, 45–58, 1990.
- Ostrowski, M., Mazard, S., Tetu, S. G., Phillippy, K., Johnson, A., Palenik, B., Paulsen, I. T., and Scanlan, D. J.: PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*, *ISME J.*, 4, 908–921, 2010.
- Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E. E., McCarren, J., Paulsen, I., Dufresne, A., Partensky, F., Webb, E. A., and Waterbury, J.: The genome of a motile marine *Synechococcus*, *Nature*, 424, 1037–1042, 2003.
- Palenik, B., Ren, Q. H., Dupont, C. L., Myers, G. S., Heidelberg, J. F., Badger, J. H., Madupu, R., Nelson, W. C., Brinkac, L. M., Dodson, R. J., Durkin, A. S., Daugherty, S. C., Sullivan, S. A., Khouri, H., Mohamoud, Y., Halpin, R., and Paulsen, I. T.: Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment, *Proc. Natl Acad. Sci. USA.*, 103, 13555–13559, 2006.
- Partensky, F., Blanchot, J., Lantoine, F., Neveux, J., and Marie, D.: Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean, *Deep Sea Res. Pt. I*, 43, 1191–1213, 1996.
- Partensky, F., Blanchot, J., and Vaultot, D.: Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review, in: *Marine Cyanobacteria*, edited by: Charpy, L. and Larkum, A., Musée Océanographique, Monaco, 457–475, 1999.
- Partensky, F. and Garczarek, L.: *Prochlorococcus*: Advantages and limits of minimalism, *Ann. Rev. Mar. Sci.*, 2, 305–331, 2010.
- Paytan, A., Mackey, K. R., Chen, Y., Lima, I. D., Doney, S. C., Mahowald, N., Labiosa, R., and Post, A. F.: Toxicity of atmospheric aerosols on marine phytoplankton, *Proc. Natl. Acad. Sci. USA.*, 106, 4601–4605, 2009.
- Posada, D.: jModelTest: phylogenetic model averaging, *Mol. Biol. Evol.*, 25, 1253–1256, 2008.
- Posada, D.: Selection of models of DNA evolution with jModelTest, *Meth. Mol. Biol.*, 537, 93–112, 2009.
- Pujo-Pay, M., Conan, P., Oriol, L., Cornet-Barthaux, V., Falco, C., Ghiglione, J.-F., Goyet, C., Moutin, T., and Prieur, L.: Integrated survey of elemental stoichiometry (C, N, P) from the western to eastern Mediterranean Sea, *Biogeosciences*, 8, 883–899, doi:10.5194/bg-8-883-2011, 2011.
- Raimbault, P., Slawyk, G., Coste, B., and Fry, J.: Feasibility of using an automated colorimetric procedure for the determination of seawater nitrate in the 0 to 100 nM range: Examples from field and culture, *Mar. Biol.*, 104, 347–351, 1990.
- Ras, J., Uitz, J., and Claustre, H.: Spatial variability of phytoplankton pigment distributions in the subtropical South Pacific Ocean: comparison between in situ and modelled data, *Biogeosciences*, 5, 353–369, 2008, <http://www.biogeosciences.net/5/353/2008/>.
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An improved in situ and satellite SST analysis for climate, *J. Clim.*, 15, 1609–1625, 2002.
- Rocap, G., Distel, D. L., Waterbury, J. B., and Chisholm, S. W.: Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences, *Appl. Environ. Microbiol.*, 68, 1180–1191, 2002.
- Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R., and Chisholm, S. W.: Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation, *Nature*, 424, 1042–1047, 2003.
- Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L., and Venter, J. C.: Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions, *Proc. Natl. Acad. Sci. USA*, 107, 16184–16189, 2010.
- Saaroni, H., Ziv, B., Edelson, J., and Alpert, P.: Long-term variations in summer temperatures over the Eastern Mediterranean, *Geophys. Res. Lett.*, 30, doi:10.1029/2003GL017742, 2003.
- Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., Post, A. F., Hagemann, M., Paulsen, I., and Partensky, F.: Ecological genomics of marine picocyanobacteria, *Microbiol. Mol. Biol. Rev.*, 73, 249–299, 2009.
- Somot, S., Sevault, F., and Déqué, M.: Transient climate change scenario simulation of the Mediterranean Sea for the Twenty-first century using a high-resolution ocean circulation model, *Clim. Dynam.*, 27, 851–879, 2006.
- Stamatakis, A.: RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, *Bioinformatics*, 22, 2688–2690, 2006.

- Tai, V. and Palenik, B.: Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site, *ISME J.*, 3, 903–915, 2009.
- Tai, V., Burton, R. S., and Palenik, B.: Temporal and spatial distributions of marine *Synechococcus* in the Southern California Bight assessed by hybridization to bead-arrays *Mar. Ecol. Prog. Ser.*, 426, 133–147, 2011.
- Tanaka, T., Zohary, T., Krom, M. D., Lawe, C. S., Pitta, P., Psarra, S., Rassoulzadegan, F., Thingstad, T. F., Tselepidis, A., Woodward, E. M. S., Flaten, G. A. F., Skjoldal, E. F., and Zodiatis, G.: Microbial community structure and function in the Levantine Basin of the eastern Mediterranean, *Deep Sea Res. Pt. I*, 54, 1721–1743, 2007.
- Tanaka, T., Thingstad, T. F., Christaki, U., Colombet, J., Cornet-Barthaux, V., Courties, C., Grattepanche, J.-D., Lagaria, A., Nedoma, J., Oriol, L., Psarra, S., Pujo-Pay, M., and Van Wambeke, F.: Lack of P-limitation of phytoplankton and heterotrophic prokaryotes in surface waters of three anticyclonic eddies in the stratified Mediterranean Sea, *Biogeosciences*, 8, 525–538, doi:10.5194/bg-8-525-2011, 2011.
- Ternon, E., Guieu, C., Ridame, C., L’Helguen, S., and Catala, P.: Longitudinal variability of the biogeochemical role of Mediterranean aerosols in the Mediterranean Sea, *Biogeosciences*, 8, 1067–1080, doi:10.5194/bg-8-1067-2011, 2011.
- Tetu, S. G., Brahamsha, B., Johnson, D. A., Tai, V., Phillippy, K., Palenik, B., and Paulsen, I. T.: Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102, *ISME J.*, 3, 835–849, 2009.
- Thompson, J. R., Marcelino, L. A., and Polz, M. F.: Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by “reconditioning PCR”, *Nucl. Acids Res.*, 30, 2083–2088, 2002.
- Toledo, G., Palenik, B., and Brahamsha, B.: Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group, *Appl. Environ. Microbiol.*, 65, 5247–5251, 1999.
- Toledo, G. and Palenik, B.: A *Synechococcus* serotype is found preferentially in surface marine waters, *Limnol. Oceanogr.*, 48, 1744–1755, 2003.
- West, N. J. and Scanlan, D. J.: Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern north Atlantic ocean, *Appl. Environ. Microbiol.*, 65, 2585–2591, 1999.
- West, N. J., Lebaron, P., Strutton, P. G., and Suzuki, M. T.: A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean, *ISME J.*, 5, 933–944, 2010.
- Zinser, E. R., Coe, A., Johnson, Z. I., Martiny, A. C., Fuller, N. J., Scanlan, D. J., and Chisholm, S. W.: *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method, *Appl. Environ. Microbiol.*, 72, 723–732, 2006.
- Zinser, E. R., Johnson, Z. I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S. W.: Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean, *Limnol. Oceanogr.*, 52, 2205–2220, 2007.
- Zubkov, M. V., Sleigh, M. A., and Burkill, P. H.: Assaying picoplankton distribution by flow cytometry of underway samples collected along a meridional transect across the Atlantic Ocean, *Aquat. Microb. Ecol.*, 21, 13–20, 2000.
- Zwirgmaier, K., Heywood, J. L., Chamberlain, K., Woodward, E. M. S., Zubkov, M. V., and Scanlan, D. J.: Basin-scale distribution patterns lineages in the Atlantic Ocean, *Environ. Microbiol.*, 9, 1278–1290, 2007.
- Zwirgmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaultot, D., Not, F., Massana, R., Ulloa, O., and Scanlan, D. J.: Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes, *Environ. Microbiol.*, 10, 147–161, 2008.

---

---

## ANNEXE B : ARTICLE IV

### CYANOLYASE : UNE BASE DE DONNÉES DÉDIÉE AUX PHYCOBILINES LYASES

---

**CyanoLyase: a database of phycobilin lyase sequences,  
motifs and functions. (2012)**

*Nucleic Acid Research*. **41** (Database Issue) 396-401

Bretaudeau A., Coste, F., Humily, E., Garczarek, L., Le Corguillé, G., Six, C.,  
Ratin, M., Collin, O., Schluchter, W.M. and Partensky F.2012

## **I. RÉSUMÉ**

Les PBP sont les composants majoritaires des phycobilisomes, et peuvent représenter jusqu'à 40-50% des protéines totales des cellules cyanobactériennes cultivées à faible intensité lumineuse (Glazer, 1989b). Ces structures protéiques sont également retrouvées chez les algues rouges (Rhodophyta), les glaucocystophytes et les cryptophytes. A ce jour, quatre types majoritaires de PBP ont été décrits: l'APC ( $\lambda_{\max} \sim 650\text{nm}$ ), la PC ( $\lambda_{\max} \sim 620\text{nm}$ ), la PE ( $\lambda_{\max} \sim 565\text{nm}$ ) et la PEC, spécifique des cyanobactéries d'eau douce ( $\lambda_{\max} \sim 590\text{nm}$ ). La capture de la lumière par est réalisée par les phycobilines. Ces composés tétrapyrroliques, dérivés de l'hème, sont liés aux sous-unités  $\alpha$  et  $\beta$  des PBP et conditionnent leurs propriétés spectroscopiques. Le processus de chromophorylation requiert dans la majorité des cas l'intervention d'enzymes spécifiques appelées phycobilines lyases, bien que l'attachement spontané de phycobilines soit également possible, notamment au niveau du  $L_{\text{CM}}$ , le polypeptide de liaison assurant la cohésion du cœur de l'APC. Ces enzymes, considérées comme des chaperonnes, assurent l'attachement correct des phycobilines, qui conditionne l'efficacité des complexes collecteurs de lumière. Certaines peuvent être bi-fonctionnelles, assurant à la fois la fixation et le changement de configuration chimique (ou isomérisation) du chromophore.

Dans cet article, dont je suis co-auteur, nous décrivons une base de données spécifiquement dédiées aux phycobilines-lyases: Cyanolyase. Cette base de données rassemble à la fois les informations sur les séquences et les structures, et fournit divers outils bioinformatiques permettant la mise en évidence de motifs ou l'annotation de ces enzymes au sein de nouveaux génomes. En rassemblant les informations issues des caractérisations biochimiques et comparaisons génomiques des phycobilines lyases, CyanoLyase constitue un excellent outil pour la compréhension et la prédiction du rôle et du fonctionnement de cette famille protéique, dont le nombre d'applications biotechnologiques ne cesse de croître (Wu et al., 2013).

## **I. ARTICLE**

# CyanoLyase: a database of phycobilin lyase sequences, motifs and functions

Anthony Bretaudeau<sup>1</sup>, François Coste<sup>2</sup>, Florian Humily<sup>3,4</sup>, Laurence Garczarek<sup>3,4</sup>, Gildas Le Corguillé<sup>3,5</sup>, Christophe Six<sup>3,4</sup>, Morgane Ratin<sup>3,4</sup>, Olivier Collin<sup>1</sup>, Wendy M. Schluchter<sup>6</sup> and Frédéric Partensky<sup>3,4,\*</sup>

<sup>1</sup>GenOuest Platform, <sup>2</sup>Dyliss team, INRIA/Irisa, Campus de Beaulieu, 35042 Rennes Cedex, <sup>3</sup>UPMC-Université Paris VI, Station Biologique, Place Georges Teissier, 29680 Roscoff France, <sup>4</sup>CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment, Oceanic Plankton Group, 29680, <sup>5</sup>CNRS, FR2424 Analysis and Bioinformatics for Marine Science (ABiMS), 29680 Roscoff, France and <sup>6</sup>Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA

Received August 29, 2012; Revised October 17, 2012; Accepted October 18, 2012

## ABSTRACT

**CyanoLyase** (<http://cyanolyase.genouest.org/>) is a manually curated sequence and motif database of phycobilin lyases and related proteins. These enzymes catalyze the covalent ligation of chromophores (phycobilins) to specific binding sites of phycobiliproteins (PBPs). The latter constitute the building bricks of phycobilisomes, the major light-harvesting systems of cyanobacteria and red algae. Phycobilin lyases sequences are poorly annotated in public databases. Sequences included in CyanoLyase were retrieved from all available genomes of these organisms and a few others by similarity searches using biochemically characterized enzyme sequences and then classified into 3 clans and 32 families. Amino acid motifs were computed for each family using Protomata learner. CyanoLyase also includes BLAST and a novel pattern matching tool (Protomatch) that allow users to rapidly retrieve and annotate lyases from any new genome. In addition, it provides phylogenetic analyses of all phycobilin lyases families, describes their function, their presence/absence in all genomes of the database (phyletic profiles) and predicts the chromophorylation of PBPs in each strain. The site also includes a thorough bibliography about phycobilin lyases and genomes included in the database. This resource should be useful to scientists and companies interested in natural or artificial PBPs, which have a number of biotechnological applications, notably as fluorescent markers.

## INTRODUCTION

Oxygenic phototrophic prokaryotes (i.e. cyanobacteria) share with the eukaryotic classes Rhodophyta (i.e. red algae) and Cryptophyta the presence of phycobiliproteins (PBPs), which are water-soluble proteins chromophorylated with brilliantly colored, linear-tetrapyrrolic pigments, called phycobilins (1). In red algae and most cyanobacterial species, different PBP types are assembled to form phycobilisomes (PBS), the major light-harvesting systems of these organisms, which are constituted of a central core surrounded by a number of radiating rods [usually six in cyanobacteria (2,3)]. Although the main antenna system of Cryptophyta is a membrane-intrinsic Lhc-type complex, like in all other photosynthetic eukaryotes (except red algae), cryptophytes possess a secondary one made of tightly packed aggregates of one PBP type, either phycocyanin (PC) or phycoerythrin (PE), located in the thylakoid lumen in the proximity of photosystems (4).

Although the PBP and phycobilin composition of the PBS core varies little, because it is always composed of allophycocyanin (APC) that binds phycocyanobilin (PCB) as its only chromophore, the structure of PBS rods is extremely variable between groups, and even within a given genus (3). In marine *Synechococcus*, for instance, six different pigment types have been described so far (5), based on the various PBP composition of their PBS rods. Indeed, the latter can comprise one to three of four possible PBP types: PC, PE-I or PE-II and phycoerythrocyanin [PEC, so far only found in freshwater species; see (3)]. Furthermore, the phycobilin composition of each individual PBP itself varies, because it may bind one to three types of the four different possible phycobilin types: PCB, phycoerythrobin (PEB), phycourobilin (PUB) and phycoviolobilin (PVB), which are isomers with distinct

\*To whom correspondence should be addressed. Tel: +33 2 9829 2564; Fax: +33 2 9829 2324; Email: partensky@sb-roscoff.fr

© The Author(s) 2012. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

spectral properties. PBPs generally consist of two subunits,  $\alpha$  and  $\beta$ , organized into hexamers, and each subunit has either one ( $\alpha$ - or  $\beta$ -APC,  $\alpha$ -PC and  $\alpha$ -PEC), two ( $\alpha$ -PEI and  $\beta$ -PC) or three chromophore binding cysteinyl sites ( $\beta$ -PEI,  $\alpha$ -PEII and  $\beta$ -PEII). Given this complexity, the phycobilin lyases, that is the enzymes that catalyze the ligation of chromophores to PBPs, constitute a particularly wide and diversified group of proteins (6,7). Among those which have been biochemically characterized, most are highly specific *in vivo* as they can ligate only one phycobilin type at one particular PBP binding site. However, CpcS (a member of the S/U clan) can bind either PCB or PEB to one specific site, Cys-82 (consensus numbering), of a variety of PBPs ( $\alpha$ - and  $\beta$ -APC and PE,  $\beta$ -PC and PEC) and is therefore more universal (8). Furthermore, some enzymes of the E/F clan are bifunctional, because they can both bind a chromophore (either PCB for PecE/F or PEB for RpcG) to  $\alpha$ -PC and change its chemical configuration into another isomer [i.e. PVB or PUB, respectively; see (9–13)].

Here, we describe CyanoLyase, a sequence and motif database dedicated to the annotation of phycobilin lyases and related proteins. Indeed, these enzymes are often poorly annotated in public databases, especially sequences coming from genome projects. Given the fact that PBPs have a growing number of biotechnological and biomedical applications [see, e.g. (14) and references therein], this resource should be very useful to all scientists and companies interested in natural or artificial PBPs. Furthermore, the knowledge of the phycobilin lyase content of any given cyanobacterial strain can be used to predict the pigmentation of its PBS, even if the latter was previously unknown, and therefore, CyanoLyase provides a list of the known and predicted chromophores at all binding positions of PBPs for most strains of the database. CyanoLyase also contains bioinformatic tools, BLAST (15) and a new pattern analysis suite Protomata [see (16) and <http://tools.genouest.org/tools/protomata>], that allow users to rapidly retrieve and annotate all lyases present in any new genome using a whole proteome file in Fasta format. In addition, it provides tables specifying the function of lyases and phyletic profiles [i.e. patterns of presence of orthologs in a set of genomes; see, e.g. (17)] allowing the user to determine the co-occurrence of lyase genes in the different strains of the database.

## DATA COLLECTION AND CURATION

The CyanoLyase database is mainly composed of sequences of characterized or presumed phycobilin lyases retrieved from genomes of cyanobacteria, red algae or cryptophytes. However, in view of forthcoming evolutionary studies of this interesting enzyme group, the database also comprises sequences of a number of phylogenetically related proteins, such as NblB that is involved in PBS degradation during nitrogen starvation (18) or IaiH involved in iron–sulfur cluster biosynthesis (19), as well as other proteins with no characterized function to date. At the time of writing, CyanoLyase accounted 954

sequences of phycobilin lyases and related proteins, coming from 84 genomes (mainly cyanobacteria). These sequences have been classified into three main clans [i.e. proteins sharing a common 3D structure; see, e.g. (20)] and 30 different families [i.e. groups of orthologous sequences; see, e.g. (21)], a modification and extension of the previous classification proposed by Schluchter *et al.* (7). For members of the S/U and E/F clans, the 3D structure was predicted using the Protein Fold Recognition Server Phyre 2 (22), while there is so far no structure that fits members of the T clan in public databases. The E/F clan was further subdivided into two subclans, based on both the phylogeny and the fact that enzymes belonging to E/F subclan 1 form either heteroduplexes or fusion proteins, whereas members of E/F subclan 2 apparently do not. Each family in our classification gathers proteins that we assume to have the same biological function. Some families were further divided into subfamilies, based on phylogenetic analyses, which for instance often split apart marine picocyanobacteria sequences from other cyanobacteria (7).

To build the dataset included in the CyanoLyase database, an initial set of biochemically characterized phycobilin lyases was compiled from an extensive literature survey and similarity searches were conducted using BLAST to retrieve highly homologous sequences in all cyanobacterial genomes available in RefSeq. After attributing these sequences to a given family and/or subfamily, conserved amino acid motifs were computed using Protomata learner. Then, using both BLAST and Protomatch (see Tools section below for details about Protomata learner and Protomatch), more distantly related sequences were identified in public databanks (e.g. RefSeq protein) and added to existing or newly created families.

Sequences stored in the database are tagged as ‘sure’ or ‘unsure’. ‘Unsure’ sequences are sequences that, based on similarity or the presence of a conserved motif, have been affiliated to a given family, but for which matching scores are too low to ascertain that they have the same function as other members of this family (e.g. CMR092C from *Cyanidioschyzon merolae* or the two paralogous sequences AM1\_4215 and AM1\_C0217 from *Acaryochloris marina* have been attributed to the CpcS family, but tagged ‘unsure’ because the identity to other members of this well-conserved family is lower than 53%). ‘Unsure’ sequences are indicated in italics in the family description pages and are not used for motif and phylogeny inferences.

All the sequences stored in CyanoLyase were manually curated, and modifications were made for some open reading frames (ORFs) that were not accurately defined in public databanks (GenBank). For instance, some protein sequences were missing a few residues at the N-terminus and were therefore extended, whereas others seemingly had too long N-termini due to a misplaced start codon and were then shortened. Also, two sequences (CpcSIII in *Cylindrospermopsis raciborskii* CS-505 and in *Raphidiopsis brookii* D9) were found to have a bacterial group II intron insertion, which led to the prediction of two independent ORFs by the annotation algorithms. In

D398 *Nucleic Acids Research*, 2013, Vol. 41, Database issue

this case, the intronic region was suppressed and the two ORFs fused. Whenever such an alteration was made to a sequence, this was reported in the corresponding remark field.

Because most of the sequences of CyanoLyase come from GenBank, NCBI record IDs are included in the database, when available, to keep track of the origin of the data.

## DATA ACCESS

A free, public access to the CyanoLyase database, tools and other features are available at <http://cyanolyase.genouest.org>. A brief description of this group of enzymes, some database statistics and direct access to bioinformatic tools are available from the 'Home Page'. Several browsing methods are available to access the information contained in the database: curated data, applications and additional features are arranged under independent pull-down menus, the main ones called 'Genomes', 'Families', 'Functions', 'Pigmentation', 'Blast', 'Protomatch', 'Phyletic profile' and 'References'.

The 'Genomes' page lists all the genomes where at least one 'true' phycobilin lyase has been identified (i.e. not only a related sequence) and provides some information about strain taxonomy, classification, environment and the sequencing center and status. This list, like all others in CyanoLyase, is sortable and filterable to ease the navigation. It can also be exported in various formats (csv and pdf). Clicking on a genome displays information about this genome, links to its RefSeq record (if available), some bibliographic references and the list of phycobilin lyase or related sequences that were found in this genome.

The 'Families' page displays the classification of sequences included in the database that are divided in clans, subclans, families and subfamilies. For each level, a hyperlink leads to a brief description, some bibliographic references and the list of sequences belonging to this group and, for most of the two lower levels, additional links give access to amino acid motifs. Each sequence stored in CyanoLyase has a dedicated page with details about the genome where the sequence comes from and the family in which it is classified.

Because CyanoLyase keeps track of NCBI record IDs (GenBank, RefSeq) of sequences and genomes when available, it offers users the possibility to display the genomic context of each phycobilin lyase gene (Figure 1) using the NCBI Sequence Viewer (<http://www.ncbi.nlm.nih.gov/projects/sviewer/>). Using this tool, the user gets access to the genomic organization around lyase genes. This option is of particular interest because the latter genes are frequently organized in clusters with other genes involved in PBS biosynthesis and regulation (5).

## TOOLS

Some bioinformatic tools are directly available on the CyanoLyase website to perform analyzes of new sequences to find novel members of the phycobilin lyase family. A BLAST (version 2.2.26+) form allows users to search

for sequence similarity of query sequences in CyanoLyase databanks. These databanks (that can be selected using a scroll-down menu) comprise not only all individual phycobilin lyase and related protein families but also the whole proteomes of all cyanobacteria and red algae included in the database. BLAST results can be downloaded in various formats. Sequences already recorded in CyanoLyase and having an associated GenBank ID are highlighted in the result page by the presence of the abbreviation 'CL' (for CyanoLyase), just before the ORF ID in the BLAST result.

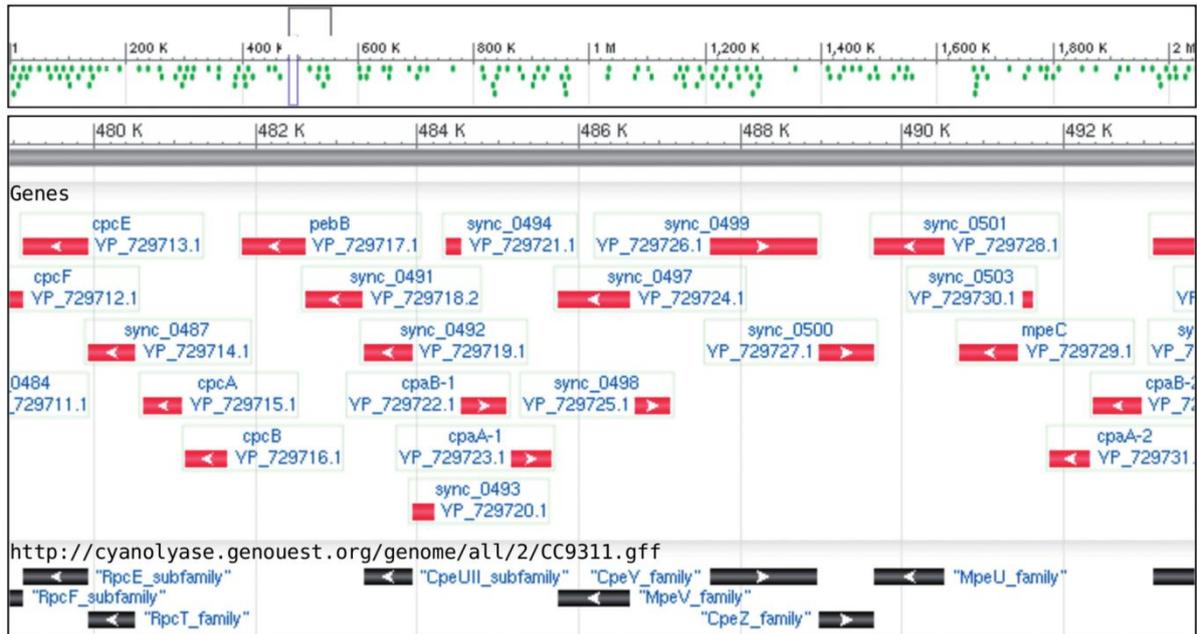
CyanoLyase also gives access to an original motif discovery and matching tool suite: Protomata (version 2.0; <http://tools.genouest.org/tools/protomata>). This software allows the user to discover motifs in sets of related sequences, focusing only on most conserved regions, represented as blocks with a letter size proportional to the amino acid frequency at any position (Figure 2). Most often, multiple blocks are detected for each dataset, and each block can be found in all the sequences or only a subset of them. With this tool, it is possible to detect the regions common to all the sequences of a protein family, but also regions shared only by some sequences that may constitute a subfamily. Using Protomata learner, a motif has been generated for most lyase family and subfamily. These motifs are available online and can be used directly from the web interface to search for motif matching on new protein sequences using Protomatch.

Using both BLAST and Protomatch allows users to rapidly retrieve all the putative phycobilin lyases or related sequences present in any new genome within a few minutes, directly in a web browser. The same tools can also be used to search for such sequences in public databanks such as RefSeq protein or NR. This type of search will be done on a regular basis by CyanoLyase authors to keep the database updated.

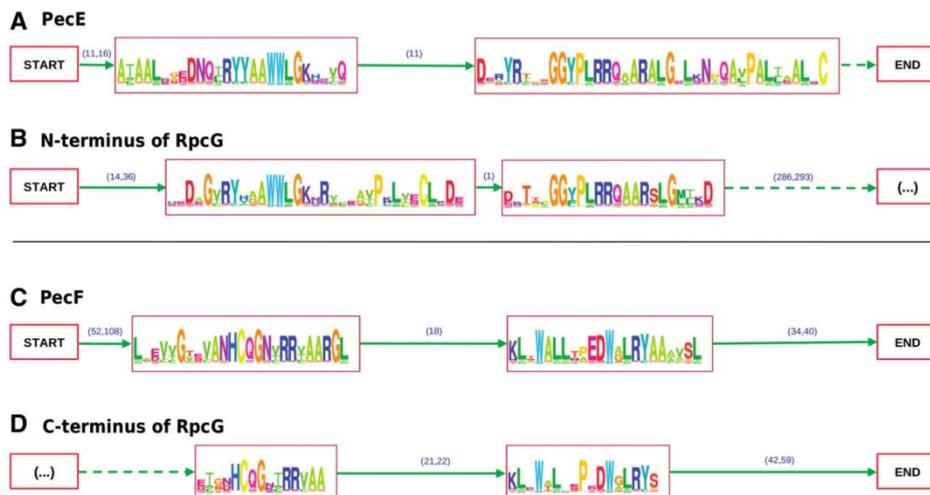
In addition, a phylogeny program has been integrated into CyanoLyase. It performs a succession of three tasks: (i) using Muscle (23), it automatically generates multiple alignments of protein sequences for each level from subfamilies to clans, (ii) using trimAl (24), it eliminates gaps and low-quality regions and (iii) using PhyML (25), it generates Maximum Likelihood trees. The multiple alignments and phylogenetic trees (under both Newick format and radial visualization) are available in the 'Phylogeny' boxes that appear in the description of each group of sequences at all levels of the classification.

## ADDITIONAL FEATURES

The 'Function' menu summarizes in a table the characterized or predicted functions of phycobilin lyases in terms of the bound chromophore (PCB or PEB), enzymatic activity (lyase or lyase-isomerase), substrate (apo-protein targeted by the enzyme) and binding sites (cysteine positions at which phycobilins are bound). These data refer to literature data listed in the 'References' menu (see below). It is important to note that phycobilin lyases often have a lower specificity *in vitro* than *in vivo* (9,26), but we chose to display their *in vivo* properties.



**Figure 1.** Use of the Browser NCBI to visualize the genome context of lyase genes in the genome of *Synechococcus* sp. CC9311. The upper track shows the region with a low level of zoom and the middle track with a higher level of zoom and show (in red) all ORFs with their IDs (or gene names) in the RefSeq record. The bottom track displays (in black) genes present in the CyanoLyase database and their corresponding name.



**Figure 2.** Example of motifs created by Protomata learner for members of the E/F subclass 1. Motifs for the PecE (A) and PecF (C) subunits of the PCB:Cys-84  $\alpha$ -PEC lyase-isomerase PecE/F(12) and for the corresponding N-terminus (B) and C-terminus (D) of the closely related fusion protein RpcG, a PEB:Cys-84  $\alpha$ -PC lyase-isomerase (9). Note the similarity of motifs independently found by Protomata learner for these two sets of protein sequences, and in particular the presence of the well-conserved motifs YyaAWWL, biochemically characterized as essential for the lyase activity of PecE (13) and nHCQGn, conferring its isomerase activity to PecF (13), in the N- and C-terminus of RpcG, respectively.

The 'Pigmentation' page displays the PBP content for each strain listed in the genome page and the chromophores found at the different binding sites of PBP subunits. The latter information was either determined biochemically (bold letters) or predicted by CyanoLyase authors (normal letters), based on the presence of specific phycobilin lyases in the genome.

The 'Phyletic profiles' page shows the distribution of phycobilin lyases or related proteins in the different genomes of the CyanoLyase database. It allows the user to easily spot sequences that are present in a comparable set of genomes, and this is particularly useful to predict potential heterodimers. For instance, this tool allowed us to find that marine *Synechococcus* spp. likely possess a

D400 *Nucleic Acids Research*, 2013, Vol. 41, Database issue

CpcU-like lyase (that we called CpcU-II), which likely makes an heterodimer with CpcS-II, by analogy with CpcU-I, which forms a heterodimer with CpcS-I in a number of freshwater cyanobacteria (7,27). In contrast, CpcS-III does not seem to form heterodimers (8).

The 'References' page comprises an exhaustive bibliography about the function of the different lyases and the description of genomes included in the database. These references are referred to, when appropriate, in the different sections of the CyanoLyase website.

## IMPLEMENTATION AND DATA EXPORT

CyanoLyase was developed using the Symfony 2 PHP framework (<http://symfony.com/>). Some modules that were used to build the application are available under an open source license on github (<https://github.com/genouest/>). The web interface was designed with XHTML, CSS and JQuery, and data were integrated into a MySQL database. These data can be exported in Fasta format from the web interface either as individual sequences or as multiple sequence files, for example if one wishes to retrieve all members of a particular family or of a genome.

## CURRENT SCOPE AND FUTURE PERSPECTIVE

The first release of CyanoLyase provides an extensive collection of phycobilin lyases and related proteins, classified in clans, subclans, families and subfamilies. The website also gives access to bioinformatic tools to ease the annotation of these sequences in forthcoming genomes of PBP-containing organisms. As such, the website will be updated regularly as new data become available and will therefore be a long-term resource. Users can monitor directly from the web interface the latest changes that have occurred in the database using the corresponding scrollbar menu.

CyanoLyase aims to be a reference resource about the classification of phycobilin lyases and their respective predicted or characterized functions. Future versions could include sequences from metagenomic samples or viruses, and the same kind of resources could be built for other complex and poorly annotated groups of protein sequences, such as the polypeptide linkers that maintain the PBS assembly (3).

## ACKNOWLEDGEMENTS

The Authors thank Dr M. Czjzek for her useful hints concerning protein structures that helped them designing the new classification of phycobilin lyases used in the database. Cyanolyase would not be possible without the genome data publicly available at the National Center for Biotechnology Information.

## FUNDING

Agence Nationale de la Recherche Scientifique (ANR), Microbial Genomics Programme PELICAN,

[ANR-09-GENM-030]; European Union's Seventh Framework Programmes (FP7) MicroB3 and MaCumBa (287589 and 311975, respectively); National Science Foundation grant [MCB-0843664]. Funding for open access charge: ANR contract [ANR-09-GENM-030].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Apt, K.E., Collier, J.L. and Grossman, A.R. (1995) Evolution of the phycobiliproteins. *J. Mol. Biol.*, **248**, 79–96.
2. Glazer, A.N. (1982) Phycobilisomes: structure and dynamics. *Annu. Rev. Microbiol.*, **36**, 173–198.
3. Sidler, W.A. (1994) Phycobilisome and phycobiliprotein structure. In: Bryant, D.A. (ed.), *The Molecular Biology of Cyanobacteria*. Kluwer Academic Publishers, The Netherlands, pp. 139–216.
4. Glazer, A.N. and Wedemayer, G.J. (1995) Cryptomonad biliproteins—an evolutionary perspective. *Photosynth. Res.*, **46**, 93–105.
5. Six, C., Thomas, J.C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., Scanlan, D.J. and Partensky, F. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol.*, **8**, R259.
6. Scheer, H. and Zhao, K.H. (2008) Biliprotein maturation: the chromophore attachment. *Mol. Microbiol.*, **68**, 263–276.
7. Schlachter, W.M., Shen, G., Alvey, R.M., Biswas, A., Saunée, N.A., Williams, S.R., Miller, C.A. and Bryant, D.A. (2010) Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification. *Adv. Exp. Med. Biol.*, **675**, 211–228.
8. Zhao, K., Su, P., Tu, J., Wang, X., Liu, H., Plöschner, M., Eichacker, L., Yang, B., Zhou, M. and Scheer, H. (2007) Phycobilin:cystein-84 biliprotein lyase, a near-universal lyase for cysteine-84-binding sites in cyanobacterial phycobiliproteins. *Proc. Natl Acad. Sci. USA*, **104**, 14300–14305.
9. Blot, N., Wu, X.-J., Thomas, J.-C., Zhang, J., Garczarek, L., Böhm, S., Tu, J.M., Zhou, M., Plöschner, M., Eichacker, L. et al. (2009) Phycourobilin in a unique trichromatic phycobiliprotein is formed post-translationally by a novel phycoerythrobilin lyase-isomerase. *J. Biol. Chem.*, **284**, 9290–9298.
10. Jung, L.J., Chan, C.F. and Glazer, A.N. (1995) Candidate genes for the phycoerythrocyanin alpha subunit lyase. Biochemical analysis of *pecE* and *pecF* interposon mutants. *J. Biol. Chem.*, **270**, 12877–12884.
11. Storf, M., Parbel, A., Meyer, M., Strohmman, B., Scheer, H., Deng, M.G., Zheng, M., Zhou, M. and Zhao, K.H. (2001) Chromophore attachment to biliproteins: specificity of PecE/PecF, a lyase-isomerase for the photoactive 3(1)-cys-alpha 84-phycoviolobilin chromophore of phycoerythrocyanin. *Biochemistry*, **40**, 12444–12456.
12. Zhao, K.H., Deng, M.G., Zheng, M., Zhou, M., Parbel, A., Storf, M., Meyer, M., Strohmman, B. and Scheer, H. (2000) Novel activity of a phycobiliprotein lyase: both the attachment of phycoerythrobilin and the isomerization to phycoviolobilin are catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon. *FEBS Lett.*, **469**, 9–13.
13. Zhao, K.H., Wu, D., Zhou, M., Zhang, L., Böhm, S., Bubenzer, C. and Scheer, H. (2005) Amino acid residues associated with enzymatic activities of the isomerizing phycoviolobilin-lyase PecE/F. *Biochemistry*, **44**, 8126–8137.
14. Sekar, S. and Chandramohan, M. (2008) Phycobiliproteins as a commodity: trends in applied research, patents and commercialization. *J. Appl. Phycol.*, **20**, 113–136.
15. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinfo.*, **10**, 421.
16. Coste, F. and Kerbellec, G. (2005) A similar fragments merging approach to learn Automata on proteins. In: Gama, J., Camacho, R., Brazdil, P., Jorge, A. and Torgo, L. (eds), *Proceeding 16th European Conference on Machine Learning (ECML 2005)*. Springer-Verlag, Berlin, Heidelberg, Porto, Portugal, pp. 522–529.

17. Aravind,L. (2000) Guilt by association: contextual information in genome analysis. *Genome Res.*, **10**, 1074–1077.
18. Dolganov,N. and Grossman,A.R. (1999) A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J. Bacteriol.*, **181**, 610–617.
19. Morimoto,K., Sato,S., Tabata,S. and Nakai,M. (2003) A HEAT-repeats containing protein, IaiH, stabilizes the iron-sulfur cluster bound to the cyanobacterial IscA homologue, IscA2. *J. Biochem.*, **134**, 211–217.
20. Rawlings,N.D., Morton,F.R., Kok,C.Y., Kong,J. and Barrett,A.J. (2008) MEROPS: the peptidase database. *Nucleic Acids Res.*, **36**, D320–D325.
21. Mulikidjanian,A.Y., Koonin,E.V., Makarova,K.S., Mekhedov,S.L., Sorokin,A., Wolf,Y.I., Dufresne,A., Partensky,F., Burd,H., Kaznadzey,D. *et al.* (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl Acad. Sci. USA*, **103**, 13126–13131.
22. Kelley,L.A. and Sternberg,M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
23. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
24. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
25. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *System. Biol.*, **59**, 307–321.
26. Zhao,K.H., Zhang,J., Tu,J.M., Bohm,S., Ploscher,M., Eichacker,L., Bubenzer,C., Scheer,H., Wang,X. and Zhou,M. (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.
27. Saunée,N.A., Williams,S.R., Bryant,D.A. and Schluchter,W.M. (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.

---

---

**ANNEXE C : POSTER I**

**A TARGETED METAGENOMIC APPROACH TO STUDY THE  
PIGMENT DIVERSITY OF MARINE *SYNECHOCOCCUS* AND  
THE DISTRIBUTION OF PIGMENT TYPES *IN SITU***

---

*Conférence Jacques-Monod*

*Génomique écologique intégrative - Integrative ecological  
genomics.*

Roscoff (Bretagne) - 15-19 octobre 2011

Humily F., Marie D., Mazard S., Scanlan D.J., Mary I., Partensky F. and Garczarek L.



---

---

**ANNEXE D : POSTER II**

**DIVERSITY AND FUNCTIONAL CHARACTERIZATION  
OF CHROMATIC ACCLIMATION IN MARINE  
*SYNECHOCOCCUS* CYANOBACTERIA**

---

***14th International Symposium on Phototrophic Prokaryotes***

Porto - Portugal - 5-9 Août 2012

Humily E., Shukla, A., Garczarek L., Six C., Marie D., Schlutchter W.M., Karty J.A., Hammad, L.A. Kehoe  
D.M. and Partensky F

# Diversity and Functional Characterization of Chromatic Acclimation in Marine *Synechococcus* Cyanobacteria

Florian Humily<sup>1</sup>, Animesh Shukla<sup>2</sup>, Laurence Garczarek<sup>1</sup>, Christophe Six<sup>1</sup>, Dominique Marie<sup>1</sup>, Wendy M. Schlachter<sup>3</sup>, Jonathan A. Karty<sup>4</sup>, Loubna A. Hamdad<sup>2</sup>, David M. Kehoe<sup>2</sup> and Frédéric Partensky<sup>1</sup>

<sup>1</sup>UMR7144 Adaptation et Diversité en Milieu Marin, CNRS & UPMC-Université Paris 6, Station Biologique, Place Georges Teissier, 29680 Roscoff, France; <sup>2</sup>Department of Biology, 1001 East Third Street, Indiana University, Bloomington, IN 47405, USA; <sup>3</sup>Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA; <sup>4</sup>METACyt Biochemical Analysis Center, Department of Chemistry Indiana University, Bloomington, IN 47405, USA

## Type IV chromatic acclimation (CA4) in marine *Synechococcus*

Chromatic acclimation (CA) is a process by which some cyanobacterial species can match the properties of their light-harvesting complex (phycobilisome or PBS) to the quality of ambient light [1]. In freshwater cyanobacteria, this process generally consists in changes in the relative composition of the PBS rods in phycocyanin and phycoerythrin (PE), when cells are shifted from red to green light or vice versa. In contrast, marine *Synechococcus* display a unique form of CA, called CA4, that occurs during shifts from blue (BL) to green light (GL) [2] and involves changes in the chromophorylation of both  $\alpha$ -PEI and  $\alpha$ -PEII subunits [3]. Indeed, the ratio of the two chromophores bound to both PEI and PEII, phycourobilin (PUB) and phycocourobilin (PEB), is high under blue light and low under green light [2, 3].

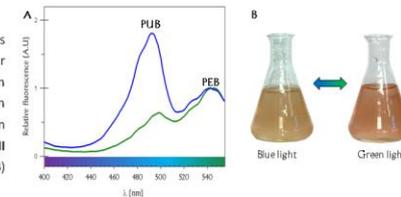


Fig. 1: Fluorescence and phenotypic color changes during CA4 in the marine *Synechococcus* strain RS9916.

A: Fluorescence excitation spectra of RS9916 cells grown in blue light (blue line) or green light (green line) showing changes in the relative fluorescence excitation of the two bilans PUB ( $\lambda_{max}$  ~495 nm) and PEB ( $\lambda_{max}$  ~500 nm), with emission at 580 nm.

B: Different colors of RS9916 cells grown in BL and GL.

## The gene *mpeZ* is involved in the CA4 process

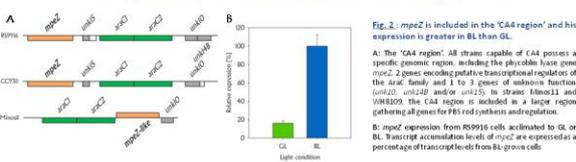


Fig. 2: *mpeZ* is included in the 'CA4 region' and his expression is greater in BL than GL.

A: The 'CA4 region'. All strains capable of CA4 possess a specific genomic region, including the phycobilin lyase gene *mpeZ*, 2 genes encoding putative transcriptional regulators of the *Arac* family and 1 to 3 genes of unknown function (*mpeO*, *mpeN* and/or *mpeL*). In strain H10211 and WH6209 the CA4 region is included in a larger region gathering all genes for PBS rod synthesis and regulation.

B: *mpeZ* expression from RS9916 cells, acclimated to GL or BL. Transcript accumulation levels of *mpeZ* are expressed as a percentage of transcript levels from BL-grown cells.

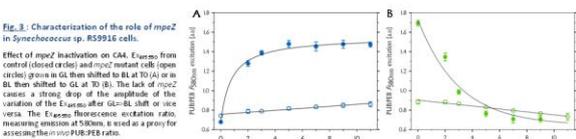


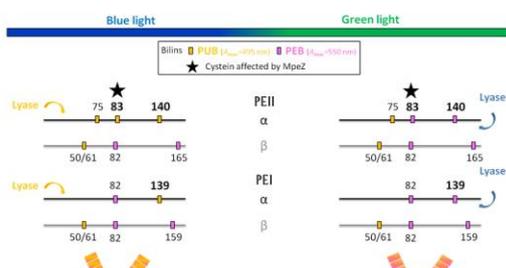
Fig. 3: Characterisation of the role of *mpeZ* in *Synechococcus* sp. RS9916 cells.

Effect of *mpeZ* inactivation on CA4. *E. fusus* from control (closed circles) and *mpeZ* mutant cells (open circles) grown in GL then shifted to BL at TO (A) or in BL then shifted to GL at TO (B). The lack of *mpeZ* causes a strong drop of the amplitude of the variation of the *E. fusus* after GL=BL shift or vice versa. The *E. fusus* fluorescence excitation ratio, measured at 580 nm, is used as a proxy for assessing the in vivo PUB:PEB ratio.

- CA4 requires a specific genomic region, including *mpeZ* or a related *mpeZ*-like gene.
- mpeZ* is highly expressed in BL and is critical for the CA4 process [4].
- The fluorescence emission characteristics are dramatically altered in a *mpeZ*-less mutant

## MpeZ is a phycobilin lyase-isomerase

Fig. 4: Differential chromophorylation of PEI and PEII in BL and GL in RS9916 cells. Chromophores bound at different cysteine binding sites of the two PE types are indicated by rectangles of different colors (PEB=pink, PUB=orange).



- CA4 involves changes at three phycobilin binding sites
- MpeZ* attaches a PEB at Cys-83 of PEII  $\alpha$ -subunit (*MpeA*) and isomerizes it into PUB [4].
- One (or two) lyase-isomerase(s) involved in CA4 still remain(s) to be characterized.

## CA4 requires protein synthesis

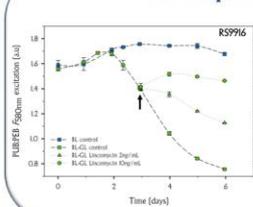


Fig. 5: Addition of lincomycin stops the CA4 process

RS9916 cultures acclimated to BL were shifted to GL. Lincomycin, a protein synthesis inhibitor, was added (arrow) during the course of pigment changes at two final concentrations: 2 ng ml<sup>-1</sup> (triangles) or 10 ng ml<sup>-1</sup> (squares). Cultures maintained in BL (blue squares) and cultures shifted without addition of lincomycin (green squares) serve as experiment control.

- CA4 likely involves *de novo* synthesis of novel PBS constituted of PEII discs matching ambient light quality that progressively replace initial PBS.

## Phenotypic variability among *Synechococcus* strains

Physiological responses of 13 potentially chromatically adapting isolates, acclimated to BL then shifted to BL (or vice versa). This experiment was done at two distinct irradiances: low light (LL) and high light (HL), i.e. 20 and 80  $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ , respectively.

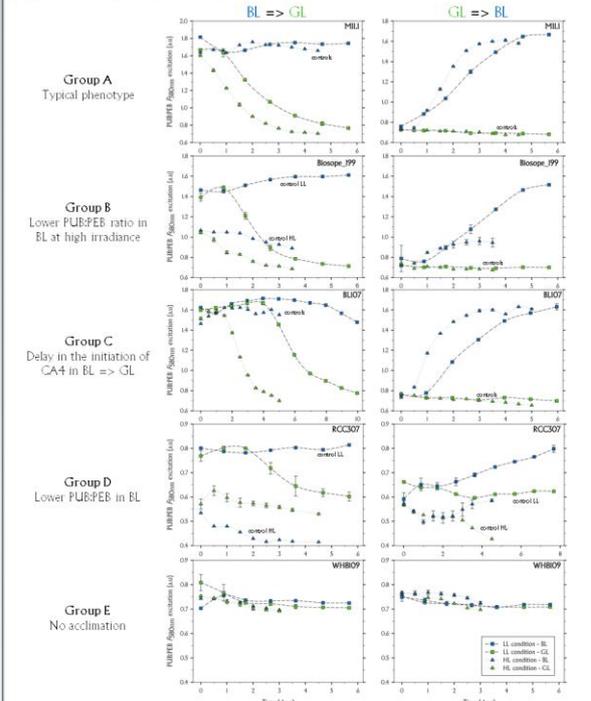


Fig. 6: Behaviors of representative strains, potentially chromatically adapting, of marine *Synechococcus*. Cultures acclimated to BL or GL were shifted at time zero (GL (left panel) and BL (right panel), respectively). The experiment was conducted both in LL (squares) and HL condition (triangles).

- The kinetics and amplitude of variation of the PUB:PEB ratio may vary between different *Synechococcus* strains, independently from their phylogenetic affiliation
- Some phenotypic variability may result from changes in the phycobilin protein composition of PBS between low and high irradiances (Groups B and D).
- Variation of the amplitude of the PUB:PEB ratio in BL (Group B), the delay in the initiation of the CA4 process when cells are shifted to GL (Group C) or even the complete loss of the CA4 process (Group E) likely result from defects in CA4 regulation pathways
- CA4 regulators clearly have a key role, but still need to be characterized (*arc*-like genes are likely candidates)

## CONCLUSIONS

- CA4 is widely distributed within the marine *Synechococcus* radiation
- CA4 involves replacement of 3 PUB chromophores in BL
- Identification of the phycobilin lyase-isomerase *mpeZ*, specific to Cys-83  $\alpha$ -PEII
- High phenotypic variability among strains possessing the CA4 region

## References cited

- Guth, A., and Johnson, D. A. (2011). Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. *Mol. Plant* 5: 1-23.
- Partensky, F. (2003). Chromatic adaptation in marine *Synechococcus* strains. *Appl. Environ. Microbiol.* 67(2): 593-594.
- Everroad, et al. (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp. *Bacteriol.* 188(9): 3345-3356.
- Shukla et al. A phycobilin-specific bilan lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Submitted Proc. Natl. Acad. Sci. U.S.A.*

## ANNEXE E

### NOUVEAUX GÉNOMES DE *SYNECHOCOCCUS* SÉQUENCÉS

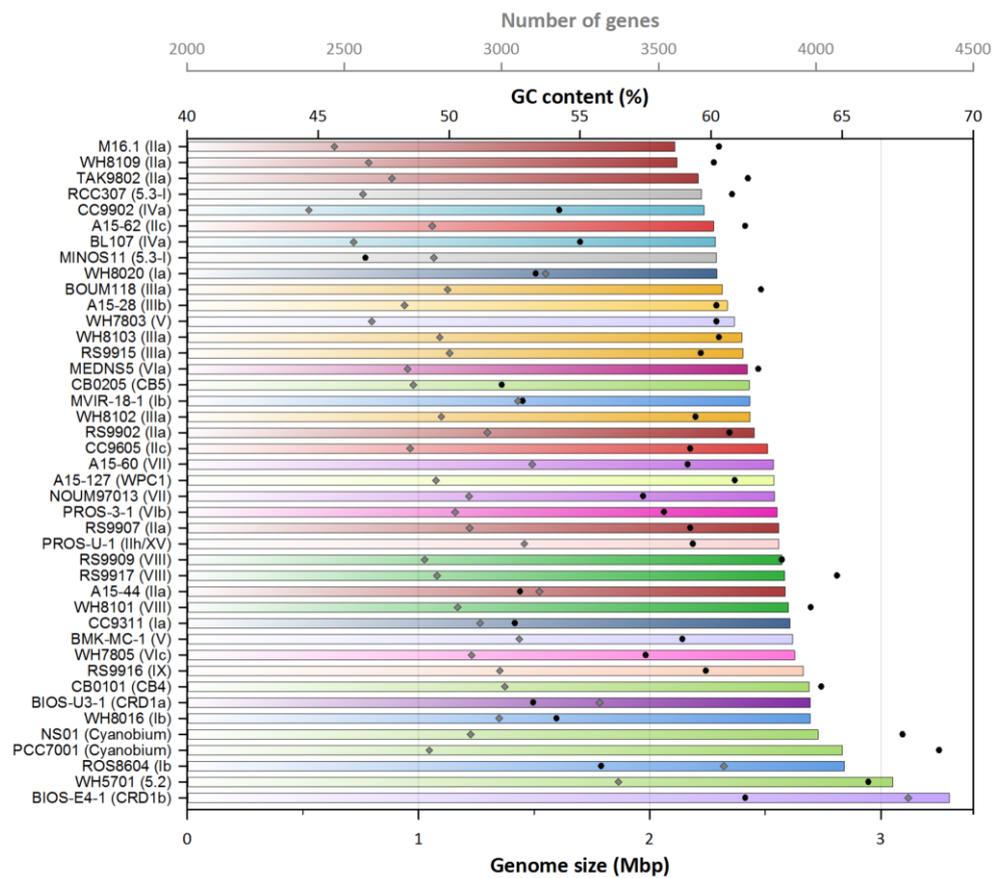


Figure 62 : Taille des génomes des *Synechococcus* et *Cyanobium* marins, contenu en GC et nombre de gènes

## ÉTUDE GÉNOMIQUE, MÉTAGÉNOMIQUE ET PHYSIOLOGIQUE DE LA DIVERSITÉ PIGMENTAIRE CHEZ LES CYANOBACTÉRIES DU GENRE *SYNECHOCOCCUS*

### Résumé

Les picocyanobactéries du genre *Synechococcus* sont présentes dans tous les types d'environnements marins. Cette ubiquité s'explique en partie par la grande diversité pigmentaire de ces cellules, leur permettant de capturer efficacement la lumière sur une large gamme spectrale. La plupart des souches ont une composition pigmentaire fixe mais certaines sont capables d'ajuster leur pigmentation en fonction de la lumière incidente par un processus physiologique appelé acclimatation chromatique de type IV (AC4). Une approche de métagénomique ciblée originale, combinant des techniques sophistiquées, a été développée afin d'étudier la diversité et la distribution des différents types pigmentaires de *Synechococcus in situ*. L'intérêt de cette approche a pu être démontré après des optimisations spécifiques. La disponibilité de 25 nouveaux génomes de *Synechococcus* a permis de faire d'importantes avancées sur la compréhension du mécanisme d'AC4. Un cluster de 4 à 6 gènes, codant pour une phycobiline-lyase et plusieurs régulateurs transcriptionnels, est systématiquement présent chez toutes les souches capables de cette plasticité phénotypique. Deux configurations bien distinctes de ce cluster, nommées AC4-A et AC4-B, ont été découvertes et se retrouvent dans des lignées différentes de *Synechococcus*. Ces deux types de clusters auraient été soumis à des processus évolutifs distincts. Par ailleurs, certaines singularités phénotypiques entre les souches possédant ces deux types de clusters génomiques ont pu être démontrées. Ce travail de thèse soulève de nouvelles hypothèses sur la régulation de cette plasticité phénotypique ainsi que sur les mécanismes biochimiques associés.

Mots-clés : *Synechococcus*, diversité pigmentaire, diversité génétique, métagénomique ciblée, acclimatation chromatique

### Abstract

Picocyanobacteria of the *Synechococcus* genus are present in all marine environments. This ubiquity is partly explained by the large pigment diversity found within this genus, allowing them to efficiently capture photons over a broad spectral range. Most strains have a fixed pigmentation but some of them are able to match their pigmentation with the ambient light quality by a physiological process called type IV chromatic acclimation (CA4). An original targeted metagenomics approach, combining sophisticated techniques, was developed to study the diversity and distribution of pigment types of *Synechococcus in situ*. The interest of our approach has been demonstrated after specific optimizations. Availability of 25 new genomes of *Synechococcus* has allowed us to make significant advances in the understanding of molecular mechanism of CA4. A cluster of 4 to 6 genes, encoding a phycobilin lyase and several transcriptional regulators, is consistently present in all strains capable of this phenotypic plasticity. Two distinct configurations of this cluster, named CA4-A and CA4-B, have been discovered and were found in different *Synechococcus* lineages. These two types of clusters have undergone distinct evolutionary processes. In addition, some phenotypic peculiarities between strains having these two types of genomic clusters have been demonstrated. This thesis raises new hypotheses about the regulation of this phenotypic plasticity as well as the biochemical mechanisms involved.

Keywords : *Synechococcus*, pigment diversity, genetic diversity, targeted metagenomics, chromatic acclimation