

# Acclimatation du picoeucaryote photosynthétique *Pelagomonas calceolata* aux changements environnementaux

*Acclimation of photosynthetic picoeukaryote Pelagomonas calceolata to  
environmental changes*

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577, Structure et dynamique des systèmes vivants (SDSV)  
Spécialité de doctorat : Biologie Cellulaire  
Graduate School : Life Sciences and Health. Référent : Université d'Évry Val d'Essonne

Thèse préparée dans l'unité de recherche **Génomique métabolique** (Université Paris-Saclay, Univ Evry, CNRS, CEA), sous la direction de **Patrick WINCKER**, Directeur de Recherche et la co-direction de **Quentin CARRADEC**, Chercheur.

Thèse soutenue à Paris-Saclay, le 15 décembre 2023, par

**Nina GUÉRIN**

## Composition du Jury

Membres du jury avec voix délibératives

<b>Gwenaël PIGANEAU</b> Directrice de recherche CNRS Sorbonne Université	Présidente du jury & Rapporteur
<b>Fabrice NOT</b> Directeur de recherche CNRS Sorbonne Université	Rapporteur & Examineur
<b>Helena CRUZ DE CARVALHO</b> Professeure assistante École Normale Supérieure de Paris	Examinatrice
<b>Justine MARCHAND</b> Maîtresse de conférences Le Mans Université	Examinatrice



**Titre :** Acclimatation du picoeucaryote photosynthétique *Pelagomonas calceolata* aux changements environnementaux

**Mots clés :** *Pelagomonas calceolata* ; Tara Oceans ; acclimatation ; changements environnementaux ; métatranscriptomique ; génomique

**Résumé :** Les picoeucaryotes photosynthétiques (PPE) sont abondants dans tous les océans et constituent une part importante de la biomasse et de la production primaire. Les modèles climatiques prédisent une extension des zones oligotrophes dans les prochaines décennies ce qui pourrait fortement augmenter l'abondance et l'impact écologique des PPE. Parmi eux, la microalgue *Pelagomonas calceolata* (Stramenopiles/Pelagophyceae) est très largement répandue dans les océans, mais son rôle dans le cycle du carbone et son impact sur la chaîne trophique restent méconnus. Des analyses *in situ* et *in vitro* suggèrent que *P. calceolata* peut s'adapter aux variations environnementales grâce à une importante capacité de modulation de l'expression des gènes. L'objectif de cette thèse est de comprendre comment *P. calceolata* s'adapte aux variations environnementales des nombreux milieux qu'elle occupe. Dans le premier chapitre, le génome de *P. calceolata* est assemblé, annoté puis comparé à ceux d'autres PPE. Grâce aux données de métagénomiques et métatranscriptomiques issues de l'expédition Tara Oceans, la biogéographie et l'activité transcriptomique de *P. calceolata* dans les différentes conditions environnementales a permis de mieux comprendre la distribution présente et future de cette algue, ainsi que les gènes impliqués dans son succès écologique. Dans le deuxième chapitre, je me suis particulièrement intéressés aux capacités d'acclimatation de *P. calceolata* face aux changements de quantité et de source d'azote. Les gènes différenciellement exprimés (DEG) chez *P. calceolata* en fonction de la concentration en nitrate dans les échantillons Tara Oceans ont été comparés avec ceux identifiés lors d'expériences de croissance en conditions contrôlées. *P. calceolata* a été cultivé dans des milieux déplétés en nitrate ou dans lesquelles le nitrate a été remplacé par une autre source d'azote : de l'ammonium, de l'urée ou du cyanate.

La comparaison des DEG issus du laboratoire avec ceux issus des données environnementales permet de mieux comprendre le métabolisme de cette microalgue face au manque de nitrate et quels mécanismes sont mis en place dans l'environnement pour faire face à la variabilité de la disponibilité du nitrate, notamment une capacité à utiliser des sources d'azote organique. Dans le troisième chapitre, nous avons voulu mieux comprendre comment la profondeur impacte la physiologie de *P. calceolata*. En effet, on retrouve *P. calceolata* dans des échantillons d'eaux de surface et jusqu'à 200m de profondeur. Nous avons constaté que la profondeur d'échantillonnage impactait fortement l'expression des gènes de *P. calceolata* impliqués dans la fixation du carbone et dans les mécanismes de concentration du carbone. Lors de cette thèse de doctorat, la caractérisation des capacités d'adaptation de *P. calceolata* a permis de mieux comprendre comment la régulation transcriptomique lui permet d'être cosmopolite, et montre que cette microalgue peut servir d'organisme modèle grâce à la possibilité de l'étudier simultanément en laboratoire et dans des données multi-omiques environnementales.

**Title:** Acclimation of photosynthetic picoeukaryote *Pelagomonas calceolata* to environmental changes

**Keywords:** *Pelagomonas calceolata* ; Tara Oceans ; acclimation ; environmental changes ; metatranscriptomic ; genomic

**Abstract:** Photosynthetic picoeukaryotes (PPE) are abundant in all oceans and represent a significant proportion of biomass and primary production. Climate models predict an extension of oligotrophic areas in the following decades, which could greatly increase the abundance and ecological impact of PPEs. Among them, the microalga *Pelagomonas calceolata* (Stramenopiles/Pelagophyceae) is widely distributed in the oceans. However, its role in the carbon cycle and impact on the trophic chain remain poorly characterised. *In situ* and *in vitro* analyses suggest that *P. calceolata* can adapt to environmental variations thanks to a significant capacity to modulate gene expression. This thesis aims to understand how *P. calceolata* adapts to environmental variations in the many environments it lives in. In the first chapter, the *P. calceolata* genome is assembled, annotated and compared with other PPEs. Thanks to metagenomic and metatranscriptomic data from the Tara Oceans expedition, the biogeography and transcriptomic activity of *P. calceolata* under different environmental conditions has provided a better understanding of the present and future distribution of this alga, and the genes involved in its ecological success. In the second chapter, we focused on the acclimatisation abilities of *P. calceolata* to changing nitrogen quantities and sources. Differentially expressed genes (DEGs) in *P. calceolata* as a function of nitrate concentration in Tara Oceans samples were compared with those identified during growth experiments under controlled conditions. *P. calceolata* was grown in media depleted in nitrate or in which nitrate was replaced by ammonium, urea or cyanate. The comparison of DEGs obtained in the laboratory with those obtained from environmental data provides a better understanding of the metabolism of this microalga in the face of nitrate shortage, and of the mechanisms put in place in the environment to cope with variability in nitrate

availability, in particular through its ability to use organic nitrogen sources. In the third chapter, we aimed to understand better how depth affects the physiology of *P. calceolata*. This organism is found in water samples from the surface down to a depth of up to 200m. We found that sampling depth strongly impacted the expression of *P. calceolata* genes involved in carbon fixation and carbon concentration mechanisms. During this Ph.D. thesis, the characterisation of the adaptive capacities of *P. calceolata* led to a better understanding of how transcriptomic regulation enables it to be cosmopolitan and shows that this microalga can be used as a model organism thanks to the possibility of studying it simultaneously in the laboratory and environmental multi-omics data.

Seuls les dieux disposent d'assez de temps pour donner un nom à chaque galet d'une plage, mais c'est la patience qui leur manque.

Terry Pratchett, *Procrastination* (2005)



## Remerciements

J'aimerais remercier toutes les personnes qui ont fait de cette thèse une expérience formidable.

Je remercie Patrick Wincker, mon directeur de thèse, pour m'avoir fait confiance et m'avoir permis de travailler dans de très bonnes conditions.

Je remercie mon encadrant, Quentin Carradec, grâce à qui cette thèse a été une expérience aussi enrichissante et positive. Merci d'avoir écouté mes explications parfois confuses et répondu à toutes mes questions. Ton investissement, ta patience, ta gentillesse, ton soutien et ta confiance m'ont accompagné pendant ces 3 dernières années et m'ont permis de mûrir, humainement et scientifiquement. Merci d'être un passionné, qui trouve l'émerveillement dans chaque petite découverte.

Je remercie Gwenaël Piganeau, Fabrice Not, Helena Cruz de Carvalho, Justine Marchand et Sophie Filleur, d'avoir accepté d'être membre de mon jury de thèse. Je suis très honorée et reconnaissante d'avoir pu échanger avec vous.

Je remercie Lucie Bittner et Ian Probert d'avoir accepté d'être membre de mon comité de suivi individuel. Nos échanges étaient très enrichissants, et je savais que je pouvais compter sur votre soutien.

Je remercie le Laboratoire d'Analyse Génomique des Eucaryotes, pour son accueil pendant ces trois ans. Je remercie tous ses membres, ceux qui sont là depuis le début, qui viennent d'arriver, ou qui n'étaient que de passage.

Un grand merci à tous les doctorants, doctorantes, post-doctorants et post-doctorantes du LAGE : Julie Lê-Hoang, Paul Frémont, Marie Burel, Lucas Pavlovic, Margaux Crédeville, Lucie Cartairade, Clément Leboine, Thibault Antoine, Chloé Seyman, Janaina Rigonato et Téo Lemane. Merci d'avoir illuminé cette thèse et les pause-café avec votre bonne humeur, votre bon sens, votre intelligence, votre écoute, vos bons conseils, votre humour, et beaucoup de sucre. Je suis heureuse de vous avoir eu à mes côtés dans cette aventure.

Un grand merci à Laurie Bertrand, Céline Orvain et Adrien Thurotte pour votre aide précieuse à la paillasse. Si ces heures sous la hotte ou au microscope ont été aussi riches en apprentissages et en découvertes, c'est grâce à vous.

Merci aux ingénieures et chercheurs du LAGE : Sophie Mangenot, Julie Poulain, Tom Delmont, Morgan Gaïa, Olivier Jaillon, Eric Pelletier et Betina Porcel. Merci pour vos conseils, votre bonne humeur et votre enthousiasme scientifique.

Merci aux stagiaires et alternants qui vont et viennent avec les saisons : Mathieu, Achraf, Louis, Victor, Théo, Julien, Kayna, Nayla, Rhita et les autres. Merci particulièrement à Chloé, pour avoir été ma première stagiaire, et à Élisabeth et Marta, qui

ont réalisé pendant leurs stages le travail préliminaire qui a permis le bon déroulement de cette thèse.

Je remercie le Genoscope, tous ses membres et toutes ces équipes, pour avoir offert un cadre idéal et m'avoir permis de travailler avec plein de scientifiques passionnants.

Un merci particulier au Laboratoire de Bioinformatique pour la Génomique et la Biodiversité et au Laboratoire de Séquençage, qui ont permis une grande partie de cette thèse, de l'extraction du matériel génétique jusqu'à l'assemblage génome complet de *P. calceolata*. Merci notamment à Jean-Marc Aury, Caroline Besler, Benjamin Istace, Benjamin Noel, ainsi qu'à Pedro Oliveira, Karine Labadie, Odette Beluche, Elodie Brun, Corinne Cruaud, Ghislaine Magdelenat, Alice Moussy et Benoit Vacherie.

Merci à Madeleine Bouzon, Volker Döring, Ivan Dubois, Alain Perret et Nadia Perchat, pour votre enthousiasme et votre gentillesse, intacte et communicative malgré les défis techniques et les ratés. Gardons en tête ce proverbe Shadok : « Plus ça rate, plus il y a de chances que ça marche. »

Je remercie également à l'ensemble des doctorants du Genoscope, et particulièrement ceux que j'ai pu côtoyer au cours de ma thèse : Jérôme Arnoux, Angelina Beavogui, Briec de Bonneville, Clément Chalet, Laurine Ducrot, Deborah Martin, Humbeline Paupelin, Florent Poubanne, Océane Touré, Yasmeen Yousfi et Tom Zaplana, pour ne citer qu'eux.

Je remercie aussi le Laboratoire d'Analyse Bioinformatique pour la Génomique et le Métabolisme, particulièrement David Vallenet, Violette Da Cunha, Eddy Elisée, Jean Mainguy et David Roche, pour votre gentillesse et votre positivité.

Merci aux équipes informatiques et au secrétariat du Genoscope, pour fournir tant d'énergie et d'efforts pour que l'on puisse travailler dans les meilleures conditions.

Merci aux organisateurs du GenoPub, monument de la vie nocturne Evryenne : Mathieu Dubois, Florent Poubanne et Alexandre Protat.

Merci à la Station Biologique de Roscoff et à l'EMBRC pour son accueil chaleureux. Un merci tout particulier à Priscillia Gourvil de m'avoir formé pendant un mois à la culture d'algue et à l'expérimentation. Un grand merci pour ton énergie, ta patience, ta sympathie et pour tout ce que j'ai appris à tes côtés. Merci également à Ian Probert, Charles Bachy, Martin Gachenot, Sarah Romac, Christophe Six et tous les Roscovites qui m'ont accueilli avec beaucoup de générosité

Merci à Gilles Peltier pour nos échanges sur *P. calceolata* et les mécanismes de concentration du carbone. Notre travail ensemble a entrouvert une porte, j'ai hâte de savoir ce qu'il se cache derrière.

Un grand merci à tous les membres du Consortium *Tara Oceans*, pour nos discussions souvent scientifiques, parfois personnels, mais toujours passionnantes. Merci particulièrement à Lucia Campese, Damien Eveillard, Marinna Gaudin, Daniele Iudicone, Charlotte Nef, Lucas Paoli, Antoine Régimbeau et Flora Vincent pour nos échanges, à Nice et ailleurs.

Merci à Valérie Chaudru et Élisabeth Petit-Teixeira pour m'avoir donné l'opportunité d'enseigner à l'université.

Merci à Emmanuelle Petit, Mélanie Béraud et Aude Perdereau pour avoir fait de la fête de la Science un moment de partage scientifique formidable.

Merci à Marie Meffre, Katie Roche, Fiddy Semba Prasetya et à Victor Murisson que j'ai eu le plaisir de rencontrer en conférence. J'espère vous recroiser bientôt, en congrès ou ailleurs.

Je remercie mes relectrices, Aurore Fierobe et Marianne Voltaire. Merci pour votre lutte acharnée contre mes fautes de grammaire, et vos commentaires cocasses. Merci d'être mes meilleures amies. C'est grâce à vous que j'ai pu mener à bien cette bataille, qui n'est pas que physique, mais psychologique.

Je remercie tous mes amis, pour votre présence, votre écoute, votre soutien et votre amour. Merci à Juliette, Amandine, Anastasia, Cécile, Emma, Laurie, Mandy, Vicky, Lauriane, Charly, François, Corentin, Julien, Léonie, Lina, tous mes amis de FDV, et tous mes camarades de MVE et MES, d'être à mes côtés et d'être des personnes formidables.

Un merci immense à ma famille, Maman, Papa, Raph, Poussin, Papy et Mamie, Tata, Koukou et Cathy. Je vous embrasse fort. Merci à Marinette d'être la plus adorable des patounettes.

Un grand merci à *Pelagomonas calceolata*, pour s'être plié à mes petites expériences avec plus ou moins de bonne volonté.

Merci à vous de lire cette thèse, fruit de 3 ans et 3 mois de travail.



# Table des matières

<b>INTRODUCTION</b> .....	<b>1</b>
<b>1 - LE PHYTOPLANCTON</b> .....	<b>3</b>
LE PHYTOPLANCTON, PRODUCTEUR PRIMAIRE DES OCÉANS .....	3
LA GÉNOMIQUE, UNE RÉVOLUTION POUR L'ÉTUDE DU PHYTOPLANCTON .....	8
L'EXPÉDITION TARA OCEANS, L'ÉTUDE DU PLANCTON À L'ÉCHELLE GLOBALE .....	11
LES LIMITES DE LA GÉNOMIQUE ENVIRONNEMENTALE DES EUCARYOTES .....	16
<b>2 - LE PHYTOPLANCTON DANS UN OCÉAN OLIGOTROPHE</b> .....	<b>20</b>
L'AZOTE, UNE RESSOURCE INÉGALEMENT DISTRIBUÉE.....	20
<i>Le cycle de l'azote marin et son assimilation par le phytoplancton</i> .....	20
<i>La variabilité spatiale et temporelle de la disponibilité de l'azote marin</i> .....	23
<i>L'acclimatation du phytoplancton au manque d'azote fixé</i> .....	26
LE FER, UN MICRONUTRIMENT RARE .....	28
<i>Le fer dans l'océan</i> .....	28
<i>L'acclimatation du phytoplancton au manque de fer</i> .....	30
LA COLONNE D'EAU, UN ENVIRONNEMENT HÉTÉROGÈNE .....	33
LES IMPACTS DU CHANGEMENT CLIMATIQUE SUR L'OCÉAN ET LE PHYTOPLANCTON.....	35
<b>3 - COMPRENDRE LES PICOEUCARYOTES PHOTOSYNTHÉTIQUES POUR COMPRENDRE LE PHYTOPLANCTON</b>	<b>38</b>
LE PICOPHYTOPLANCTON, UN RÔLE ÉCOLOGIQUE AMENÉ À PRENDRE DE L'IMPORTANCE .....	38
LES PELAGOPHYCEAE, DES PICOEUCARYOTES PHOTOSYNTHÉTIQUES PEU ÉTUDIÉS.....	41
<i>Caractéristique, abondance et distribution des Pelagophyceae</i> .....	41
<i>Diversité des Pelagophyceae</i> .....	42
<i>Aureococcus anophagefferens et ces blooms toxiques en milieu oligotrophe</i> .....	46
PELAGOMONAS CALCEOLATA, UN PÉLAGOPHYTE D'INTÉRÊT .....	48
<b>4 - HYPOTHÈSE GÉNÉRALE ET OBJECTIFS DE LA THÈSE</b> .....	<b>52</b>
<b>CHAPITRE 1 - GENOMIC ADAPTATION OF THE PICOEUKARYOTE PELAGOMONAS CALCEOLATA TO IRON-POOR OCEANS REVEALED BY A CHROMOSOME-SCALE GENOME SEQUENCE</b> .....	<b>55</b>
<b>CHAPITRE 2 - NITROGEN METABOLISM IN THE PICOALGAE PELAGOMONAS CALCEOLATA: DISENTANGLING CYANATE LYASE FUNCTION UNDER NUTRIENT STRESS</b> .....	<b>71</b>
<b>1 - ABSTRACT</b> .....	<b>73</b>
<b>2 - INTRODUCTION</b> .....	<b>73</b>
<b>3 - RESULTS</b> .....	<b>76</b>
<i>IN SITU</i> GENE EXPRESSION LEVELS OF <i>P. CALCEOLATA</i> ACCORDING TO NITRATE CONCENTRATION .....	76
GROWTH AND GENE EXPRESSION VARIATIONS UNDER LOW-N CONDITIONS .....	78
<i>P. CALCEOLATA</i> GENE EXPRESSION LEVELS CULTIVATED WITH DIFFERENT NITROGEN COMPOUNDS. ....	80
NITROGENOUS COMPOUND TRANSPORT .....	82
NITRATE SENSING .....	84
NITRATE REDUCTION AND STORAGE .....	84
CYANATE LYASE, UREASE AND UREA CYCLE.....	86
<b>4 - DISCUSSION</b> .....	<b>86</b>
<i>P. CALCEOLATA</i> GROWTH UNDER VARIOUS NITROGEN SOURCES.....	86
GLOBAL RESPONSE TO LOW NITRATE CONDITIONS .....	86
GLOBAL RESPONSE TO ALTERNATIVE NITROGEN SOURCE .....	87
AMMONIUM USAGE IN <i>P. CALCEOLATA</i> .....	87
UREA CYCLE IN <i>P. CALCEOLATA</i> .....	88
URIC ACID/XANTHINE PERMEASE .....	88
ROLE OF THE CYANATE LYASE IN <i>P. CALCEOLATA</i> .....	89

<b>5 - MATERIALS &amp; METHODS .....</b>	<b>91</b>
DIFFERENTIAL ANALYSIS OF GENE EXPRESSION LEVELS IN ENVIRONMENTAL DATASETS. ....	91
CULTURES OF <i>P. CALCEOLATA</i> RCC100 STRAIN .....	92
RNA EXTRACTION AND SEQUENCING .....	93
ANALYSIS OF <i>P. CALCEOLATA</i> GENE EXPRESSION LEVELS IN DIFFERENT NITROGEN CONDITIONS .....	93
<b>CHAPITRE 3 – L’IMPACT DE LA PROFONDEUR SUR LE MÉTABOLISME DU CARBONE DE <i>PELAGOMONAS CALCEOLATA</i> .....</b>	<b>95</b>
<b>1 - INTRODUCTION.....</b>	<b>97</b>
<b>2 - MÉTHODES.....</b>	<b>98</b>
<b>3 - RÉSULTATS ET DISCUSSION .....</b>	<b>99</b>
L’EXPRESSION DES GÈNES DE <i>P. CALCEOLATA</i> EST FORTEMENT IMPACTÉE PAR LA PROFONDEUR .....	99
L’IMPACT DE LA PROFONDEUR SUR L’ASSIMILATION DE L’AZOTE .....	103
LE CYCLE DE CALVIN ET LA PHOTORESPIRATION : DES MÉTABOLISMES SUREXPRIMÉS EN SURFACE .....	105
<i>Le cycle de Calvin</i> .....	106
<i>La photorespiration</i> .....	108
<i>Les mécanismes de concentration du carbone</i> .....	110
<i>Niveau d’expression des enzymes impliquées dans la photorespiration, le cycle de Calvin et des CCM dans l’environnement.</i> .....	112
<b>4 - CONCLUSION.....</b>	<b>115</b>
<b>DISCUSSION &amp; PERSPECTIVES.....</b>	<b>117</b>
<b>1 - L’IMPACT DU CHANGEMENT CLIMATIQUE SUR LE MÉTABOLISME DE L’AZOTE DES PPE.....</b>	<b>118</b>
<b>2 - DE L’IMPORTANCE DES INTERACTIONS BIOTIQUES .....</b>	<b>120</b>
<b>3 - <i>PELAGOMONAS CALCEOLATA</i> COMME MODÈLE POUR L’ÉTUDE DES PPE.....</b>	<b>122</b>
<b>4 - CONCLUSION.....</b>	<b>125</b>
<b>BIBLIOGRAPHIE .....</b>	<b>127</b>
<b>ANNEXES.....</b>	<b>141</b>
<b>1 - ANNEXES DU CHAPITRE 1.....</b>	<b>141</b>
<b>2 - ANNEXES DU CHAPITRE 2.....</b>	<b>161</b>
<b>3 - ANNEXES DU CHAPITRE 3.....</b>	<b>163</b>

## Tables des abréviations

### Abréviations générales

ASL	Argininosuccinate lyase
ASS	Arginosuccinate synthase
CAH	Anhydrase carbonique ( <i>Carbonic Anhydrase</i> )
CCM	Mécanisme de concentrations du carbone ( <i>Carbon Concentration Mechanism</i> )
CPS	Carbamyl-phosphate synthétase
DCM	Profondeur de concentration de chlorophylle maximale ( <i>Deep Chlorophyll Maximum</i> )
DEG	Gènes différentiellement exprimés ( <i>Differentially Expressed Genes</i> )
DIN	Azote inorganique dissous ( <i>Dissolved Inorganic Nitrogen</i> )
DOM	Matière organique dissoute ( <i>Dissolved Organic Matter</i> )
DON	Azote organique dissous ( <i>Dissolved Organic Nitrogen</i> )
FBA	Fructose-biphosphate aldolase
GOGAT	Glutamate synthase
GS	Glutamine synthétase
ISIP	Protéine induite par la privation en fer ( <i>Iron starvation-induced protein</i> )
MAG	Génome assemblé à partir de métagénome ( <i>Metagenome Assembled Genome</i> )
OCT	Ornithine carbamoyltransférase
OTU	Unité taxonomique opérationnelle ( <i>Operational Taxonomic Unit</i> )
PPE	Picoeucaryote photosynthétique ( <i>Photosynthetic Picoeukaryote</i> )
SUR	Surface
TPM	Transcrit par millions

### Enzymes et molécules du cycle de Calvin et de la photorespiration

2PG	2-phosphoglycolate
3PG	3-phosphoglycérate
G3P	Glycéraldéhyde-3-phosphate
GAPDH	Glycéraldéhyde 3-phosphate déshydrogénase
GCS	Système de clivage des glycines
GOX	Glycolate oxydase
HPR	Hydroxypyruvate réductase
PGK	Phosphoglycérate kinase
PRK	Phosphoribulokinase
Rubisco	Ribulose 1,5-bisphosphate carboxylase/oxygénase
RuBP	Ribulose-1,5-bisphosphate
SGAT	Sérine-glyoxylate transaminase
SHMT	Sérine hydroxyméthyltransférase



# Introduction

---

## Préambule

L'océan est un écosystème gigantesque, varié et encore mal caractérisé, du fait des contraintes nombreuses qui limitent son exploration. A la base du réseau alimentaire océanique, on trouve le phytoplancton, qui participe à de nombreux cycles biogéochimiques. Producteur primaire, il présente une grande diversité d'espèces ayant des morphologies variées, chacune jouant des rôles différents selon les écosystèmes. Le plancton, procaryote ou eucaryote, intéresse les chercheurs depuis des siècles, mais la génomique moderne permet de répondre à des questions qu'il était impossible de poser il y a 30 ans. L'essor de la génomique environnementale a permis d'étudier le plancton à une échelle globale, ce que les expérimentations en laboratoire ne permettent pas. Cependant, les études de méta-omiques environnementales présentent une série de biais, notamment liés à la quantification. De plus en plus, on incite à coupler les études *in situ* et les méthodes de laboratoire "traditionnelles", afin de compenser les faiblesses méthodologiques de chacune.

Les nutriments comme l'azote et le fer sont essentiels à la croissance du phytoplancton, mais leur disponibilité dépend de phénomènes complexes, physiques, chimiques ou biologiques. Pour faire face aux environnements limités en nutriments, dits oligotrophes, les microalgues eucaryotes présentent des adaptations particulièrement intéressantes. Le manque d'azote peut les amener à optimiser le transport de l'azote inorganique ou à se tourner vers des molécules organiques azotées. Le manque de fer oblige les organismes à développer une série de mécanismes pour le capter plus efficacement, le stocker ou s'en passer. Le changement climatique devrait durablement intensifier la stratification des océans, ce qui va augmenter la surface des zones oligotrophes et potentiellement impacter le phytoplancton et la production primaire. Face à ces défis actuels et imminents, il est impératif d'améliorer nos connaissances sur le phytoplancton afin de mieux comprendre sa réponse aux changements environnementaux.

Parmi les picoeucaryotes photosynthétiques, les Pélagophytes sont une classe peu étudiée, dans laquelle on retrouve une quinzaine de genres qui sont caractérisés par une thèque (enveloppe solide entourant la cellule) unique. Les données génomiques manquent pour décrire l'écologie et la physiologie de cette classe qui pourrait être amenée à profiter des conséquences du changement climatique. Parmi les Pélagophytes, *Pelagomonas calceolata* présente une abondance et une distribution cosmopolite facilitant son étude *in situ* par métagénomique et métatranscriptomique. En se basant sur la bibliographie existante, nous émettons l'hypothèse que l'abondance de *P. calceolata* devrait augmenter malgré l'augmentation de la taille des zones oligotrophes. Au travers des quatre chapitres de cette thèse, nous nous attèlerons à comprendre comment *P. calceolata* s'acclimate aux changements environnementaux.



# 1 - Le phytoplancton

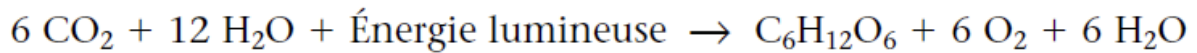
## Le phytoplancton, producteur primaire des océans

L'océan est un écosystème gigantesque qui représente 70 % de la surface terrestre. Son étude reste complexe à cause des contraintes propres à cet environnement unique. Parmi les nombreux organismes qui le peuplent, le plancton, du grec *planktós* signifiant "errant" ou "instable", regroupe l'ensemble des organismes portés par les forces du courant. On y retrouve des méduses, des larves de poissons, des crustacés de petite taille, et l'ensemble des micro-organismes. L'étude de ces derniers ne s'est développée qu'assez récemment. Le premier traité s'intéressant spécifiquement aux bactéries marines, *Die Bakterien des Meeres*, est rédigé par le microbiologiste allemand Bernhard Fischer en 1894. Pourtant, on date généralement la naissance de la microbiologie marine à 1944 avec la publication de *Marine Microbiology: A Monograph on Hydrobacteriology*, par Claude ZoBell. Des années 1950 à nos jours, les chercheurs se sont attelés à décrire la diversité des espèces composant le plancton microscopique, leurs physiologies, leurs rôles écologiques et leurs places dans les grands cycles biogéochimiques. À la base du réseau trophique, les micro-organismes marins permettent la fixation des éléments tels que le carbone, l'azote, le phosphore et bien d'autres, dans la matière organique. Cette fixation permet tout d'abord de rendre les nutriments biodisponibles pour les couches supérieures du réseau trophique, puis de les reminéraliser, ou bien de les exporter et les stocker dans l'océan profond pendant des périodes de l'ordre du cycle géologiques (Guidi et al. 2016).

Traditionnellement et dans une vision simplifiée du plancton microscopique, on différencie d'une part le phytoplancton, d'origine végétale, et d'autre part le zooplancton, d'origine animale. Plutôt qu'une origine évolutive commune, c'est le mode trophique qui différencie ces 2 catégories. Le phytoplancton est autotrophe, c'est-à-dire qu'il utilise l'énergie lumineuse, le dioxyde de carbone et les matières inorganiques de son environnement pour synthétiser sa propre biomasse. Le zooplancton en revanche est hétérotrophe, c'est-à-dire qu'il est incapable de synthétiser sa propre matière organique et utilise donc celle des autres, soit en captant les molécules organiques dissoutes (osmotrophie) soit en consommant d'autres organismes (phagotrophie). Cette version simpliste du plancton omet beaucoup de modes trophiques et d'interactions biotiques. Mais elle permet cependant de bien différencier le compartiment producteur du compartiment consommateur.

Le phytoplancton est une composante essentielle des écosystèmes marins, car il est le producteur primaire dans les océans et forme la base du réseau trophique (Pierella Karlusich, Ibarbalz, and Bowler 2020). Comme chez les plantes terrestres, cette production primaire est possible grâce à la photosynthèse oxygénique. Chez les eucaryotes, elle a lieu dans le chloroplaste, et peut se résumer sous la forme d'une équation décomposable en 2 étapes. Lors de la réaction photochimique, les photons de l'énergie lumineuse permettent d'oxyder l'eau en dioxygène pour en dégager de

l'énergie chimique sous la forme d'électrons. Ces électrons permettent d'activer le cycle de Calvin, qui va fixer le carbone du CO<sub>2</sub> dans du glucose et générer de l'eau en conséquence.



Cette simple équation permet de capter l'importance écologique et climatique du plancton. Le dioxyde de carbone nécessaire à la photosynthèse est capté sous forme dissoute dans l'eau de mer, facilitant ainsi la dissolution du CO<sub>2</sub> atmosphérique dans l'océan. L'incorporation du carbone dans la matière organique par la synthèse du glucose rend ce carbone biodisponible pour l'ensemble du réseau trophique. Il est estimé que bien que ne représentant que 1 % de la biomasse végétale, le phytoplancton est responsable de 45 % de la production primaire nette globale, c'est-à-dire 50 Gt de carbone par an (Field et al. 1998). La productivité du phytoplancton peut être affectée par la disponibilité de la lumière et des éléments tels que l'azote, le phosphore, la silice ou le fer (Martin and Fitzwater 1988; Berges et al. 1996). Si la matière organique générée par le phytoplancton n'est pas consommée par les échelons trophiques supérieurs, elle peut sédimenter et être séquestrée dans l'océan profond pendant des périodes de l'ordre du million ou milliard d'années. Ce processus de stockage du carbone est appelé pompe biologique à carbone (Guidi et al. 2016).

Le phytoplancton impacte toute l'histoire du climat terrestre depuis son apparition (Figure 1). La photosynthèse s'est dans un premier temps développée il y a 2.4 milliards d'années, chez les cyanobactéries, et a participé à la Grande Oxygénation : le dioxygène est devenu un composant essentiel de l'atmosphère (Fischer, Hemp, and Johnson 2016). Ensuite, le phytoplancton eucaryote est apparu il y a 1.5 milliard d'années, lorsque qu'un eucaryote hétérotrophe a absorbé une cyanobactérie photosynthétique, formant une endosymbiose. La cyanobactérie a perdu l'ensemble des fonctions qui pouvaient être prises en charge par la cellule eucaryote, jusqu'à devenir le chloroplaste. Cet évènement, appelé l'endosymbiose primaire, a donné naissance à l'ancêtre commun des Glaucophytes, Rhodophytes (algues rouges), Chlorophytes (algues vertes) et plantes terrestres. La majorité du phytoplancton marin actuel dérive de l'endosymbiose secondaire : il y a 1 milliard d'années, une cellule eucaryote hétérotrophe a absorbé un Rhodophyte issu de l'endosymbiose primaire, formant par un processus similaire l'ancêtre commun des Cryptophytes, Haptophytes, Straménopiles et Alvéolates. La géochimie terrestre a été fortement influencée par la photosynthèse, mais aussi par l'évolution de la morphologie du phytoplancton. Les apparitions des thèques en carbonate de calcium des coccolithophores et des frustules en silice des diatomées ont participé, de par la chimie de leur synthèse, à transformer la composition de l'atmosphère terrestre. Ces effets combinés ont mené au monde tel que nous le connaissons aujourd'hui (Pierella Karlusich, Ibarbalz, and Bowler 2020).

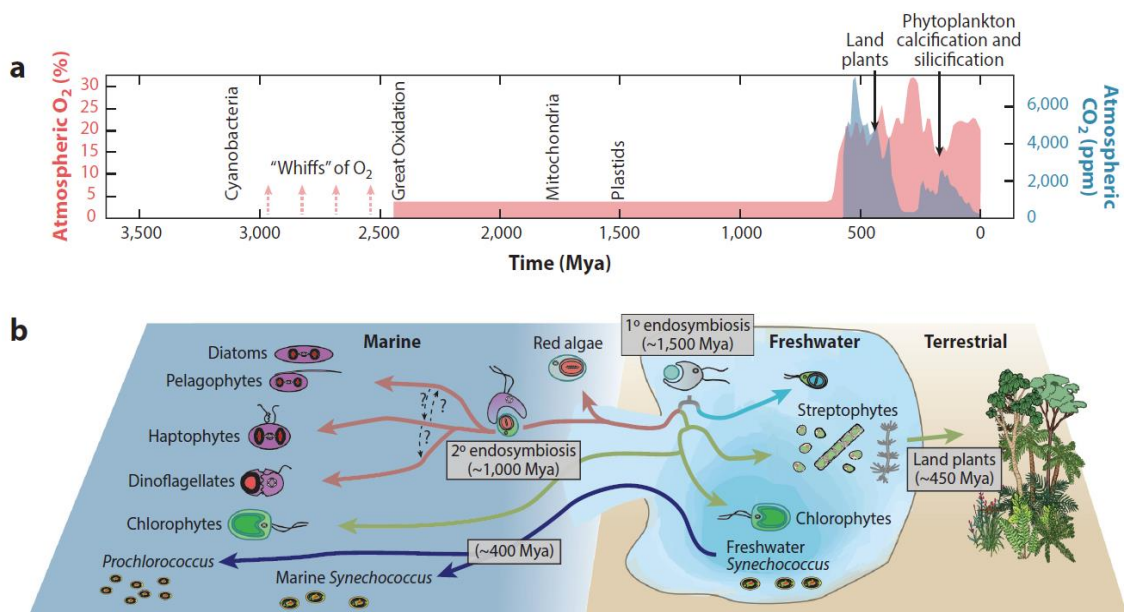


Figure 1. La relation entre l'évolution du climat et l'évolution du phytoplancton. A) Évolution de la composition en O<sub>2</sub> et CO<sub>2</sub> de l'atmosphère terrestre et évènements marquants de l'histoire évolutive des organismes photosynthétiques. B) Évolution et diversification des grands groupes phytoplanctoniques. D'après Pierella Karlusich, Ibarbalz, and Bowler 2020.

Les espèces bactériennes du phytoplancton sont majoritairement réparties dans 4 genres de cyanobactéries : *Prochlorococcus*, *Synechococcus*, *Trichodesmium* et *Crocospaera* (Figure 2A). Faisant partie des bactéries diazotrophes, ces deux dernières sont capables de fixer le diazote atmosphérique en ammonium. Ce processus, qui n'est pas propre aux cyanobactéries, est de première importance car il met l'azote à disposition de tous les espèces non-diazotrophes. Les cyanobactéries peuvent être retrouvées sous forme libre, d'agrégat, en symbiose ou en association avec les représentants eucaryotes du phytoplancton (Zehr 2011).

Le compartiment eucaryote du phytoplancton présente une grande diversité d'espèces et de forme, avec 5 lignées principales réparties sur l'ensemble de l'arbre des eucaryotes (Figure 2B).

Les dinoflagellés (ou Dinophyceae) représentent une grande diversité d'espèces, avec 24 813 OTU différentes identifiées dans la base de données d'ADNr 18S V9 de *Tara Oceans* (Vargas et al. 2015). Sur l'ensemble des espèces de phytoplancton formellement décrites en 2020, 40 % étaient des dinoflagellés (Pierella Karlusich, Ibarbalz, and Bowler 2020). Cette diversité se reflète dans la complexité de leurs modes de vie. Ils peuvent être trouvés sous forme libre, de parasites d'autres protistes ou d'animaux, ou de symbionte d'invertébrés tels que le corail, les anémones, les méduses ou les nudibranches. Ils peuvent être autotrophes, mixotrophes ou hétérotrophes, ou encore faire de la kleptoplastie (phagocyter une proie autotrophe et conserver ses chloroplastes et les utiliser) (Stoecker 1999).

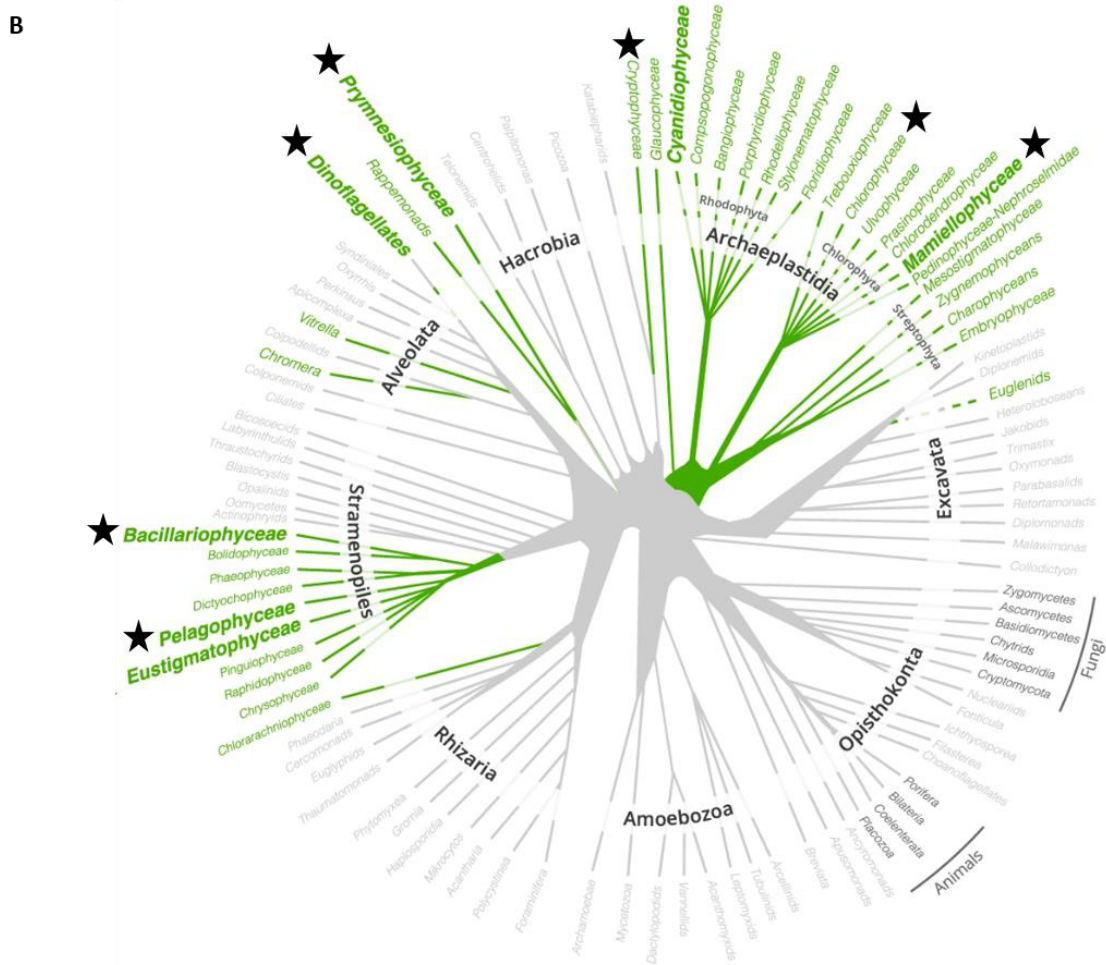
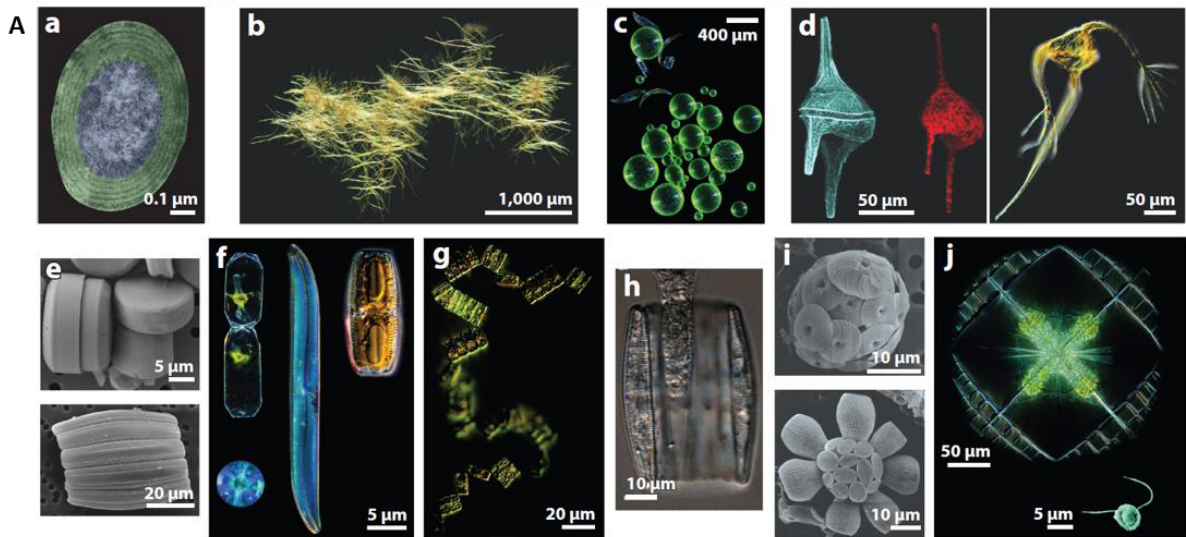


Figure 2. Diversité et taxonomie du phytoplancton marin. A) Diversité structurale et fonctionnelle du phytoplancton marin, en microscopie. a) Cellule de cyanobactérie *Prochlorococcus marinus*, un des organismes photosynthétiques les plus abondants au monde. b) Colonie de cyanobactéries diazotrophes du genre *Trichodesmium*. c) Mélange de cellules de chlorophyte *Halosphaera sp.* (cellules sphériques) et de diatomée *Rhizosolenia sp.* (cellules cylindriques). d) Deux dinoflagellés : *Ceratum candelabrum* (gauche) et *Ceratum ranipes* (droite). La fluorescence bleue permet de visualiser la thèque, et la rouge les chloroplastes. e) Diatomées centrales (haut) et pennales (bas). La microscopie électronique à balayage permet de bien visualiser la frustule. f) Plusieurs espèces de diatomées : *Cerataulina sp.* (en haut à gauche), *Actinopterychus sp.* (en bas à

gauche), *Gyrosigma* sp. (milieu), et *Amphora* sp. (droite). g) Chaîne formée par des cellules de diatomées pennales *Thalassionema nitzschioides*. h) Cellule de diatomée *Fragilariopsis doliolus*. i) Deux Prymnesiophytes : *Dicosphaera tubifera* (haut) et *Scyphosphaera apsteinii* (bas). La microscopie électronique à balayage permet de bien visualiser les coccolithes formant la coccosphère. j) Cellule d'acanthaïre (zooplancton) *Lithoptera fenestrata* en symbiose avec le Prymnesiophyte *Phaeocystis* formant les 4 masses jaunes dans son cytoplasme. En bas, une cellule de *Phaeocystis* sous forme libre. D'après Pierella Karlusich, Ibarbalz, and Bowler 2020. B) Arbre phylogénétique des eucaryotes représentant la diversité taxonomique des lignées photosynthétiques (en vert). Les taxons phytoplanctoniques marins évoqués dans le texte sont marqués d'une étoile noire. Modifié d'après Uwizeye et al. 2021.

Les diatomées (Bacillariophyta) sont des Ochrophytes, un groupe monophylétique rassemblant les Straménopiles photosynthétiques. Les diatomées sont caractérisées par une frustule, un squelette externe en silice. Pélagiques ou benthiques, libres ou fixées, seules ou en colonies, elles font partie avec les dinoflagellés des taxons les plus abondants et diversifiés du plancton marin, représentant 40 % des espèces de phytoplancton décrites, avec 3 276 OTU différents identifiés dans la base de données d'ADNr 18S V9 de *Tara* Oceans (Vargas et al. 2015; Pierella Karlusich, Ibarbalz, and Bowler 2020). Au sein des Ochrophytes, on retrouve également la classe des Pélagophytes (ou Pelagophyceae), une classe de microalgues majoritairement côtières malgré 2 espèces exclusivement pélagique. Encore peu étudiée, nous prendrons le temps de décrire cette classe plus précisément dans la partie 3 de cette introduction.

Les Haptophytes représentent 10 % des espèces phytoplanctoniques décrites (Pierella Karlusich, Ibarbalz, and Bowler 2020). Ils sont rassemblés dans 3 classes : les Pavlovaceae, les Rappephyceae et les Prymnesiophyceae. Ces derniers, aussi appelés coccolithophores, sont les plus étudiés et représentent 673 des 762 espèces d'Haptophytes décrites. Capables de former d'impressionnantes efflorescences visibles depuis l'espace, ils sont caractérisés par les coccolithes, plaques en carbonate de calcium entourant la cellule, formant une coccosphère. Cette morphologie particulière requiert de la part des cellules un processus de calcification biogénique. En précipitant dans le fond marin avec leurs coccosphères, les coccolithophores participent à l'exportation du carbone et à sa séquestration, et jouent un rôle essentiel dans la pompe à carbone biologique. Ces organismes, comme tous les organismes calcifiants (coraux, coquillages, etc.) devraient durablement souffrir de l'augmentation de l'acidité de l'eau de mer, qui perturbe l'équilibre chimique des ions carbonates et bicarbonates.

Les Chlorophytes ou algues vertes, sont un phylum appartenant à la lignée des Viridiplantae, donc issus de l'endosymbiose primaire et phylogénétiquement plus proches des plantes terrestres. Représentant 6 % des espèces phytoplanctoniques décrites (Pierella Karlusich, Ibarbalz, and Bowler 2020), ce groupe compte un grand nombre de classes : les Chlorophyceae et les Mamiellophyceae sont les plus remarquables. Dans la classe des Chlorophyceae, on retrouve des espèces très étudiées en laboratoire, comme *Volvox carteri*, qui sert de modèle pour comprendre la différenciation cellulaire, ou *Chlamydomonas reinhardtii*, qui permet d'étudier la motilité et la photosynthèse grâce au génie génétique. Les Mamiellophyceae sont également très étudiés en laboratoire, particulièrement 3 genres : *Micromonas*,

*Bathycoccus* et *Ostreococcus*. On considère ce dernier comme le plus petit eucaryote vivant sous forme libre, avec une taille moyenne de 0.8  $\mu\text{m}$ .

Les Cryptophytes sont un petit groupe de phytoplancton (environ 200 espèces) caractérisé par la présence d'éjectosome qui leur sert à se propulser en cas de stress. Certaines espèces seraient mixotrophes, mais ce groupe est encore peu caractérisé et on ignore beaucoup de choses sur les organismes qui le composent.

## **La génomique, une révolution pour l'étude du phytoplancton.**

Depuis les débuts de la microbiologie environnementale, un problème subsiste. Les scientifiques constatent une différence entre le nombre de cellules comptées par microscopie dans un échantillon environnemental, et le nombre de cellules comptées après avoir mis en culture ce même échantillon sur un milieu riche. Ce concept, théorisé en 1985 sous le nom de "Great Plate Anomaly", est observé dès 1959 dans le milieu marin (Jannasch and Jones 1959; Staley and Konopka 1985). L'hypothèse d'alors est que beaucoup d'organismes de l'environnement ne sont pas cultivables en laboratoire. Ainsi, dans le cas particulier du plancton marin, les organismes photosynthétiques et/ou côtiers sont en général plus faciles à cultiver en laboratoire que les espèces pélagiques et/ou hétérotrophe (Keeling et al. 2014). À cause de cette impossibilité à être cultivé, une immense diversité d'espèces, de modes trophiques et de rôles écologiques passent sous les radars des chercheurs. Seul des techniques d'étude des micro-organismes indépendantes de la culture peuvent permettre de contourner ce biais.

Entre 1940 et 1970, les découvertes sur l'ADN, sa composition, sa structure et sa synthèse s'enchaînent et permettent le développement du séquençage. Mises au point simultanément par 2 équipes différentes de part et d'autre de l'Atlantique, les méthodes de séquençage de l'ADN de Maxam et Gilbert et de Sanger, sont toutes deux récompensées par un prix Nobel de chimie en 1980. Le séquençage devient rapidement un outil incontournable pour l'étude des micro-organismes dans l'environnement. En 1986, Olsen et ses collaborateurs mettent au point une méthode pour identifier la composition en espèce d'un échantillon environnementale (Olsen et al. 1986). Le ribotypage consiste à utiliser les séquences d'ADN codantes pour l'ARN ribosomique 16S (ADNr 16S) comme marqueur taxonomique. Ce gène, présent en plusieurs copies dans les cellules, est impliqué dans la synthèse des ribosomes. Au vu du rôle vital de ces structures dans la synthèse des protéines, ces séquences codantes d'environ 1500 pb évoluent très peu. L'alignement de l'ensemble de l'ADNr 16S est donc aisée grâce à la conservation importante de l'ensemble de la séquence au travers du vivant. Toutefois, l'ADNr 16S présente des régions hypervariables (appelée V1 à V9) de 80 à 100 pb, qui permettent de discriminer les espèces les unes par rapport aux autres. Le ribotypage permet de caractériser directement la diversité d'une communauté microbienne environnementale sans passer par la culture. Il se généralise rapidement et d'autres gènes marqueurs taxonomiques sont identifiés, notamment les

gènes ribosomiaux codant pour l'ADNr 18S, les ITS1 et ITS2, et les gènes chloroplastiques *rbcl*, *psaA*, *psaB*, *psbA* et *psbC*. On l'élargit pour étudier la présence de n'importe quel gène dont la séquence est connue, donnant naissance au barcoding moléculaire.

L'idée émerge de s'intéresser à la totalité de l'ADN de l'environnement. Le terme métagénomique est utilisé pour la première fois en 1998, pour qualifier l'analyse d'un mélange d'ADN environnementale d'espèces différentes (Handelsman et al. 1998). Dans cette étude, l'ADN est intégralement extrait de l'échantillon de sol, puis fragmenté grâce à une enzyme de restriction afin d'être cloné dans *Escherichia coli*. Ainsi, la culture de *E. coli* permet d'obtenir suffisamment de quantité d'ADN pour le séquençage. Le génome collectif ainsi obtenu, celui de tous les organismes du sol, est appelé le métagénome. La métagénomique permet d'étudier le potentiel métabolique de l'entière communauté microbienne, sans passer par la culture (Figure 3). Cette pratique lève le voile sur l'immense potentiel des organismes non-cultivables.

À partir de 2006, le séquençage entre dans l'ère du séquençage nouvelle génération hauts-débits (NGS) : les grands représentants de cette période sont le pyroséquençage 454, la technologie SOLiD, et le séquençage Solexa par Illumina (Ansorge 2009). Ces techniques permettent d'accélérer grandement le séquençage et de réduire drastiquement les coûts grâce à l'ajout d'étapes d'amplification de l'ADN. Cette étape permet de contourner la longue et laborieuse étape du clonage. Cependant, ces nouvelles méthodes se limitent aux séquences courtes (lectures de 150 pb en moyenne, alors que celles obtenues avec la méthode Sanger sont de 800 pb en moyenne). Ce problème soulève de nouveaux défis d'ordre bioinformatique, et des besoins de nouveaux algorithmes d'assemblage. Il faut attendre 2016 pour que les séquençages d'Oxford Nanopore Technologies et Pacific Biosciences proposent une nouvelle technique, permettant de séquencer rapidement des lectures de plusieurs dizaines à centaines de kilobases, avec en contrepartie, un taux d'erreur légèrement élevé : 1 à 0.5 % d'erreur contre 0.5 à 0.1 % pour le Solexa (Sahlin and Medvedev 2021; Stoler and Nekrutenko 2021). Ces progrès augmentent drastiquement l'ampleur et la précision des études de génomique et de métagénomique.

Avec le séquençage nouvelle génération et le séquençage hauts-débits, une nouvelle approche émerge en microbiologie environnementale : la métatranscriptomique. De la même façon que la métagénomique étudie l'ensemble des génomes dans un environnement, la métatranscriptomique analyse l'ensemble des molécules d'ARNs transcrites dans un échantillon. Les premières études de métatranscriptomique en microbiologie environnementale débutent en 2005 (Poretzky et al. 2005). Cette approche, rapidement associée au métabarcoding et la métagénomique, permet de mieux comprendre l'activité transcriptomique des communautés.

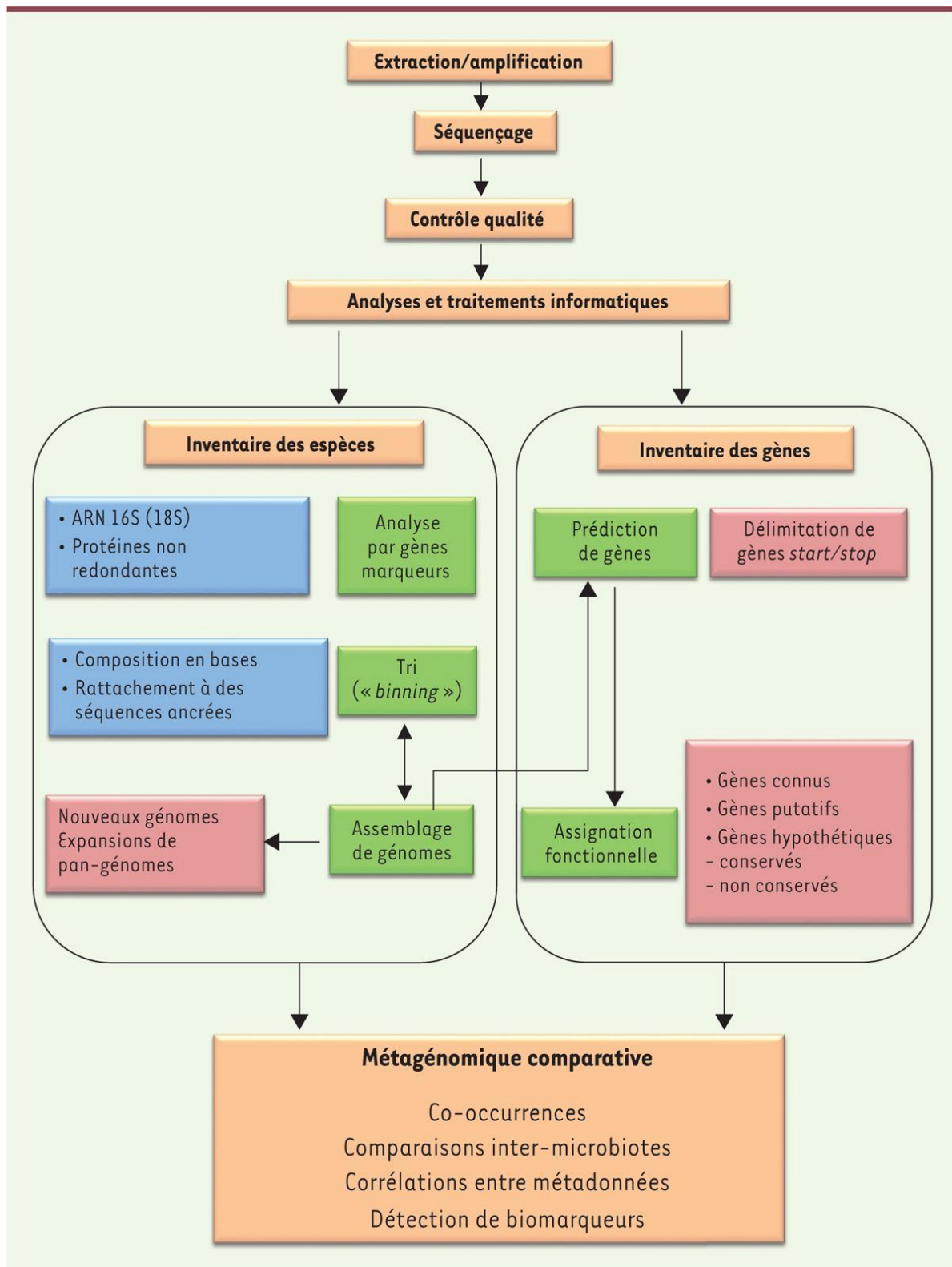


Figure 3. Exemple de pipeline d'analyse métagénomique d'un échantillon de communauté bactérienne. L'ADN environnementale peut être utilisé pour dresser l'inventaire taxonomique ou fonctionnel d'une communauté microbienne. Les catalogues de gènes et/ou d'espèce ainsi obtenus peuvent être comparés, afin d'étudier les variations de microbiotes entre les échantillons. Il faut que les échantillons étudiés aient été traités selon le même protocole d'échantillonnage, d'extraction et de séquençage. D'après Weissenbach and Sghir 2016.

L'étude du plancton bénéficie énormément de ces avancées technologiques. La métagénomique et la métatranscriptomique contribuent à de grands projets d'étude du plancton à l'échelle mondiale. En 2001, les premières études de ribotypage du gène codant pour l'ADNr 18S montrent que le plancton eucaryote a une diversité bien plus importante que celle estimée par l'observation des échantillons au microscope (López-García et al. 2001; Moon-van der Staay, De Wachter, and Vaultot 2001). Très vite, les ambitions d'échantillonnage augmentent. Des campagnes océanographiques visent à donner une vision globale du plancton marin grâce à la génomique. Les expéditions cherchant à échantillonner le plancton ont cours depuis plus de 150 ans. Entre 1872 et 1876, à bord du *HMS Challenger*, Ernst Haeckel décrit, dessine et nomme environ 3 500 nouvelles espèces de radiolaires. En 1884, Adolphe-Adrien Certes, disciple de Pasteur, récolte et cultive des micro-organismes échantillonnés dans les fonds marins sur les navires *le Travailleur* et *le Talisman*. La première expédition appliquant strictement des méthodologies de séquençage à des échantillons de plancton marin est le *Global Ocean Sampling* (GOS). Menée par le Craig Venter Institute, elle se déroule de 2004 à 2006 à bord du *Sorcerer II* et échantillonne sur 41 sites, de la Nouvelle-Écosse à la Polynésie-Française, en passant par la mer des Caraïbes, le canal de Panama et les îles Galápagos. L'assemblage des 7.7 milliards de lectures générées par le séquençage des échantillons permet de générer 44 métagénomiques d'environnements variés, ainsi que 6.12 millions de protéines bactériennes et virales annotées. L'analyse de ces données montre tout d'abord que beaucoup d'espèces procaryotes et virales sont encore inconnues et non étudiées. De plus, des espèces abondantes peuvent présenter de fortes variations génétiques entre les échantillons, ce qui suggère une adaptation aux environnements qu'elles occupent. Finalement, des organismes génétiquement isolés des autres populations sont des marqueurs de préférences environnementales uniques (Rusch et al. 2007). Les études basées sur cette expédition et le pilote mené en 2004 dans la mer des Sargasses façonnent la métagénomique moderne : d'abord, par son application à l'analyse du plancton à grande échelle, ensuite par l'effort de séquençage fourni pour être exhaustif dans l'échantillonnage (Venter et al. 2004; Tyson et al. 2004; Yooseph et al. 2007; Rusch et al. 2007).

### **L'expédition *Tara Oceans*, l'étude du plancton à l'échelle globale**

En 2008, un consortium de recherche mené par Éric Karsenti organise l'expédition *Tara Oceans*. L'objectif est d'explorer la diversité du plancton marin à l'échelle globale, du virus à la larve de poisson. De 2009 à 2013, l'expédition parcourt les océans du globe avec à bord de la goélette *Tara* : au départ de Lorient, elle traverse la Méditerranée, la mer Rouge, l'Océan Indien, l'Atlantique Sud, le Pacifique et l'Atlantique Nord pendant 2 ans et 8 mois. Puis elle repart lors de la mission *Tara Polar Circle*, pour une circumnavigation de l'Arctique pendant 7 mois (Figure 4).

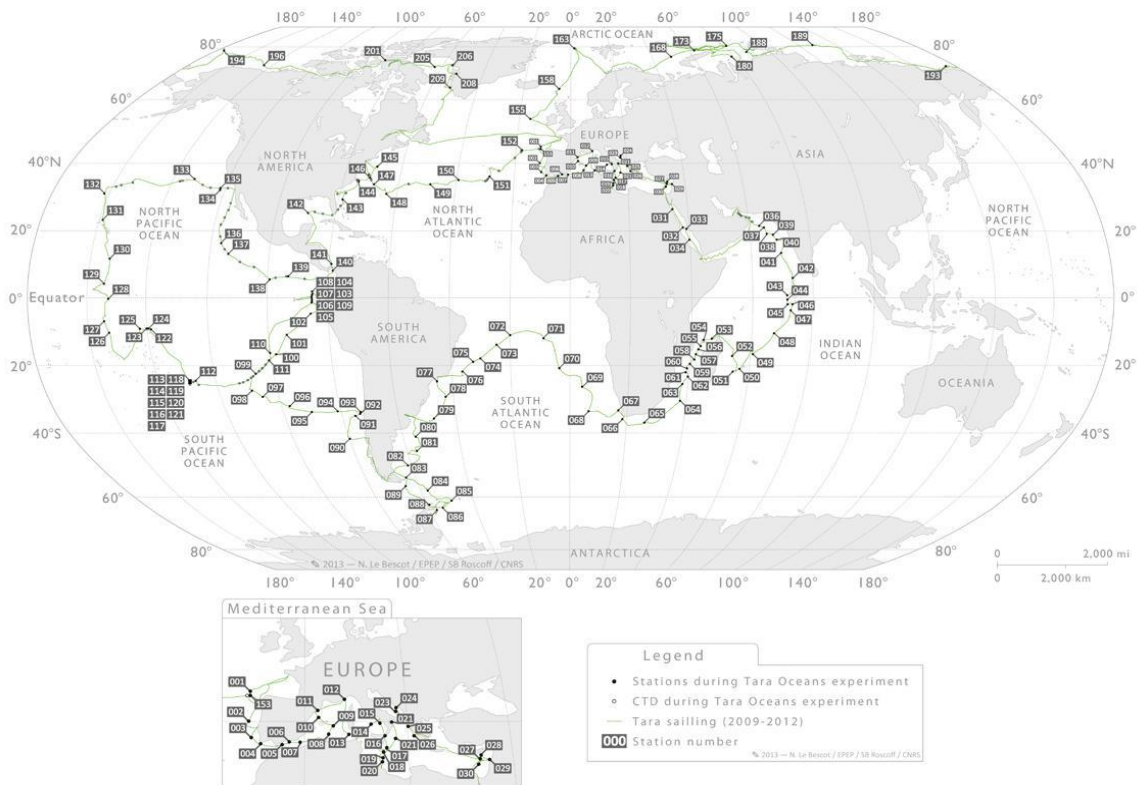


Figure 4. Carte du trajet et des sites d'échantillonnages étudiés lors de l'expédition *Tara Oceans*. À chaque station, l'eau est échantillonnée à plusieurs profondeurs et filtrée, et les paramètres environnementaux sont mesurés ou modélisés. D'après Pesant et al. 2015.

La mise au point d'un protocole d'échantillonnage standardisé favorise la comparaison inter-sites (Alberti et al. 2017; Pesant et al. 2015). Au total, l'expédition récolte environ 35 000 échantillons sur 210 sites et 1 000 profils CTD (conductivité/température/profondeur) mesurés jusqu'à 1 000 m de profondeur. L'eau est échantillonnée à la surface de la zone photique, à la profondeur où la concentration de chlorophylle est maximale (DCM, entre 20 et 100 m de profondeur), ainsi que dans la zone mésopélagique (entre 300 et 400 m, là où il n'y a plus assez de lumière pour supporter la photosynthèse). Pour chaque échantillonnage, on mesure systématiquement des paramètres physico-chimiques tels que les concentrations en nutriments (nitrate, nitrite, ammonium), en oxygène ou en pigments (chlorophylle a), les spectres d'absorption, de diffusion et d'atténuation des particules, la température, la salinité, etc. Ces mesures *in situ* sont complétées par des données satellitaires (chlorophylle a) ou modélisées (métaux traces, nutriments). Des filets de mailles variables sont déployés (Figure 5), ou l'eau est prélevée à l'aide d'une pompe péristatique. Des bouteilles Niskin sont utilisées pour les échantillons profonds. L'eau échantillonnée est ensuite filtrée sur des mailles de taille décroissante, ce qui permet de séparer les organismes planctoniques par taille (Figure 5). Les filtres les plus larges retiennent principalement des métazoaires (20-180  $\mu\text{m}$  et 180-2000  $\mu\text{m}$ ), puis les protistes (20-180  $\mu\text{m}$ , 5-20  $\mu\text{m}$  et 0.8-5  $\mu\text{m}$ ) et les bactéries (0.8-5  $\mu\text{m}$  et 0.22-3  $\mu\text{m}$ ). Enfin, les virus sont collectés dans la fraction < 0.2  $\mu\text{m}$ . Les filtres sont flash-congelés dans de l'azote liquide et rapatriés pour séquençage au Genoscope.

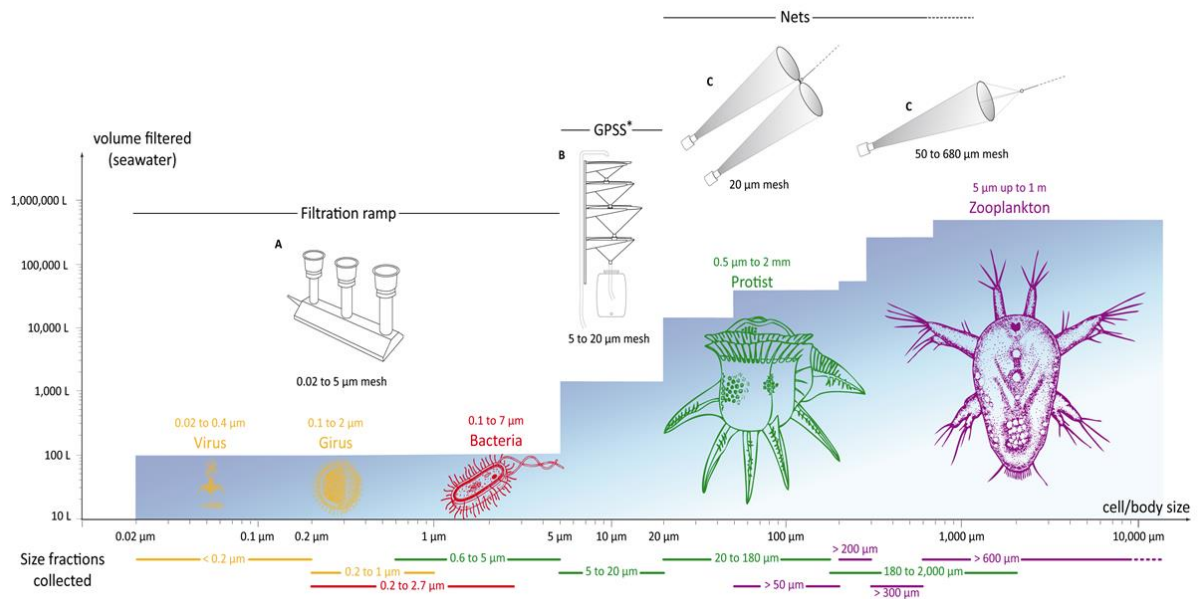


Figure 5. Méthodes de filtration appliquée en fonction de la taille du plancton. Les échantillons d'eau récupérés lors de l'expédition *Tara Oceans* sont filtrés dans des mailles de plus en plus fines afin de récupérer du plancton de toute taille. Le choix des ustensiles d'échantillonnage et de filtration varie selon le type d'organisme ciblé. Le fond bleu indique le volume d'eau indicatif à filtrer pour obtenir une biomasse d'organisme suffisante pour le séquençage. D'après Karsenti et al. 2011.

Parmi les 27 000 échantillons de plancton ainsi récoltés, la moitié est dédiée à des analyses d'imagerie, et l'autre à des analyses génomiques. L'ADN et l'ARN de ces échantillons sont extraits puis amplifiés pour le métabarcoding (ADNr 16SV4 et 18SV9) ou séquencés directement pour des analyses de métagénomique et métatranscriptomique (HiSeq 2000 d'Illumina). L'ADN de cellules uniques est également extrait, amplifié puis séquencé pour fournir des génomes d'espèces non-cultivées. Les séquences sont assemblées, les gènes sont prédits, la taxonomie et la fonction sont assignés, ce qui permet de générer de nombreuses ressources génomiques (Figure 6). Les plus marquantes sont les catalogues de gènes comme OMRGC ou MATOU (Sunagawa et al. 2015; Carradec et al. 2018), les transcriptomes générés à partir des métagénomiques (MGT) (Vorobev et al. 2020), et les génomes assemblés à partir des métagénomiques (MAG) procaryotes et eucaryotes (Delmont et al. 2018; 2022).

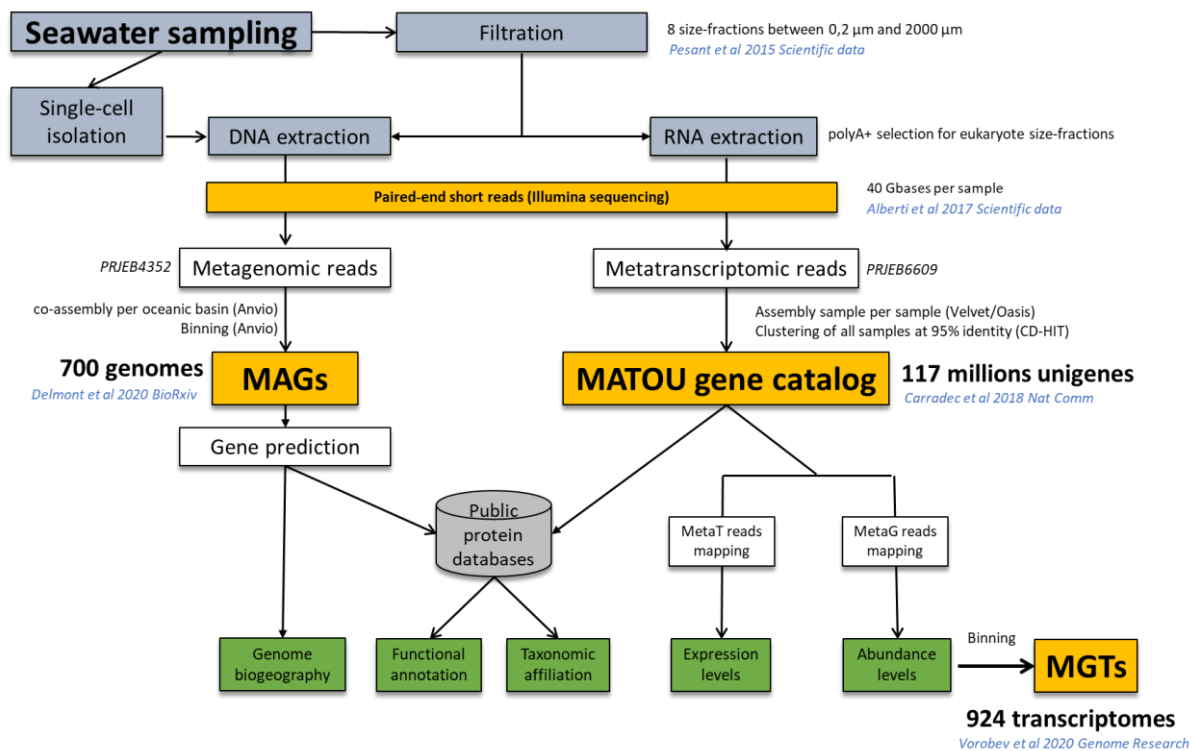


Figure 6. Pipeline simplifié de l'analyse multi-omiques appliquée aux échantillons des fractions de taille eucaryote unicellulaire issus de l'expédition *Tara Oceans*. L'eau échantillonnée est filtrée afin d'isoler les organismes dont la taille est comprise entre 0.2 et 2000 µm. Les acides nucléiques sont extraits puis séquencés avec HiSeq 2000 d'Illumina (Pesant et al. 2015; Alberti et al. 2017). Les lectures sont assemblées, celles de métagénomiques permettent de générer des génomes assemblés à partir de métagénome (MAGs) (Delmont et al. 2022) et celles de métatranscriptomiques permettent d'alimenter le catalogue de gène MATOU (Carradec et al. 2018). La comparaison avec les bases de données de séquences protéiques permet de faire l'affiliation taxonomique et fonctionnelle. Les MAGs peuvent servir à étudier la biogéographie des génomes. Les lectures métatranscriptomiques peuvent être alignées sur les transcrits MATOU ou sur les lectures métagénomiques pour étudier les niveaux d'expressions ou l'abondance. À partir de ce dernier alignement, on peut clusteriser les unigènes pour générer des transcriptomes (MGT). D'après Vorobev et al. 2020.

Cet effort d'échantillonnage est à l'origine d'une ressource génomique unique, qui couvre les trois domaines du vivant dans l'océan. Il a ainsi permis de nombreuses études. Quelles sont la biogéographie et la diversité des communautés virales (Brum et al. 2015) ? Comment la stratification impacte la distribution de la communauté procaryote et quels paramètres environnementaux impactent le plus sa composition (Sunagawa et al. 2015) ? Quel est l'impact des courants sur la dispersion et la physiologie des communautés planctoniques, et comment cela impacte le cycle de l'azote (Villar et al. 2015) ? Peut-on lier ensemble la composition taxonomique et la composition fonctionnelle d'une communauté bactérienne (Louca, Parfrey, and Doebeli 2016) ? Quel est le rôle des micro-organismes dans l'exportation du carbone vers l'océan profond (Guidi et al. 2016) ? Peut-on caractériser la diversité et la multiplicité des rôles trophiques des protistes (Vargas et al. 2015) ? Quelle est la structure de population des copépodes et comment évolue-t-elle (Madoui et al. 2017) ? Les données récoltées par l'expédition *Tara Oceans* permettent de questionner la diversité, la fonction, la biogéographie, les interactions et l'écologie du plancton marin, à de multiples échelles de tailles et de complexité (Figure 7) (Sunagawa et al.

2020).

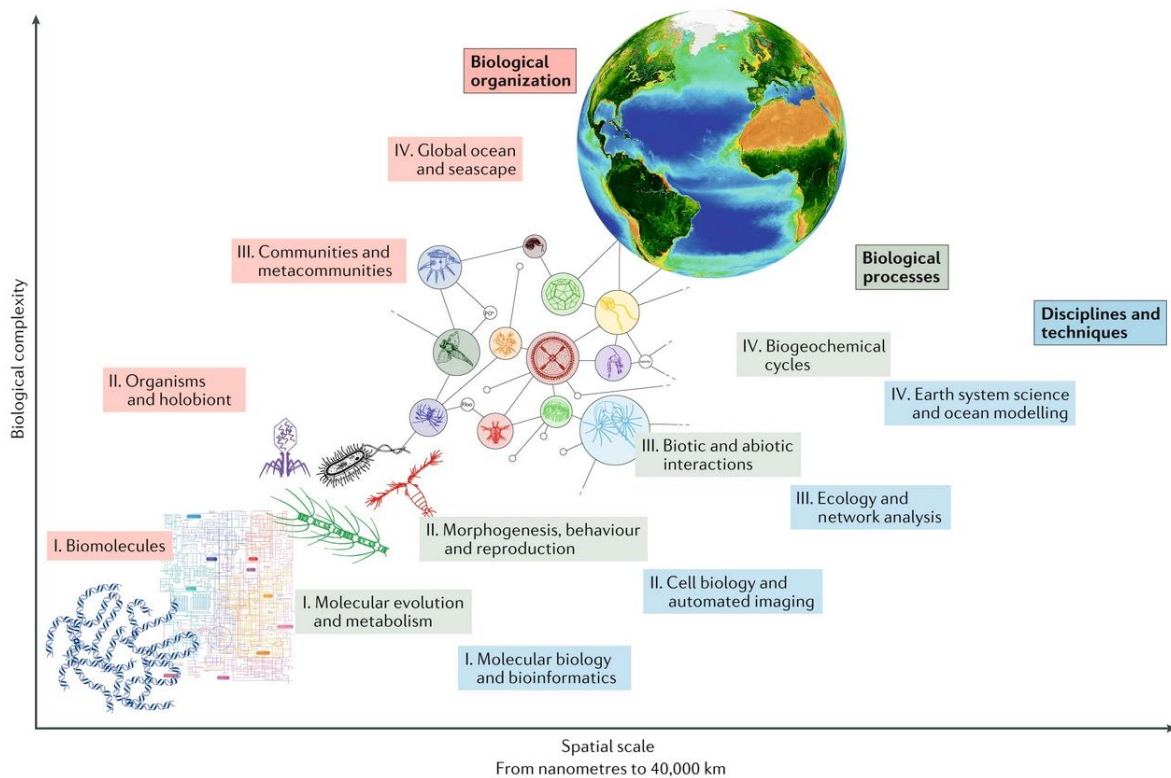


Figure 7. Les données générées par l'expédition *Tara Oceans* permettent d'étudier l'océan à toutes les échelles de taille et de complexité biologique, au travers une grande variété de niveau d'organisation biologique, de processus, et de techniques. D'après Sunagawa et al. 2020.

Les données issues de l'expédition *Tara Oceans* sont remarquables pour en raison de l'attention particulière portée aux protistes. Souvent négligés dans les études en génomique environnementale, ils représentent tout de même 85 % des eucaryotes. La filtration de l'eau entre 0.8 et 5  $\mu\text{m}$  permet de récolter des données inestimables sur le pico-nano plancton eucaryote (Vargas et al. 2015). L'effort de séquençage de barcode lors de l'expédition *Tara Oceans* est suffisant pour approcher la saturation : sur les 150 000 OTU eucaryotes identifiés, on estime que 75 % de la diversité totale est couverte, c'est à dire qu'il n'y a "que" 40 000 OTU manquants. En comparaison avec les 11 200 morphotypes de protiste marin décrits depuis les débuts de l'étude du plancton, c'est une immense fenêtre qui s'ouvre sur l'importance inattendue de certains taxons à l'abondance sous-estimée. Environ un tiers des OTU n'ont pas pu être assignés taxonomiquement, ce qui prouve que de nombreuses espèces restent inconnues. Les eucaryotes hétérotrophes sont les plus abondants, et les autotrophes ne représentent que 20 % des OTU. Cependant, une diversité étonnante de modes trophiques est observée : la proportion d'organismes mixotrophes dans le plancton eucaryote est en réalité importante, et pourrait grandement impacter les flux de carbones et de nutriments. De plus, les interactions cellules/cellules sont très courantes, avec une proportion inattendue de parasites et d'endosymbiotes. Les interactions biotiques sont

des forces évolutives potentielles bien plus grandes que les facteurs environnementaux abiotiques.

## Les limites de la génomique environnementale des eucaryotes

L'étude du plancton eucaryote marin se heurte à un fort biais de composition des bases de données, en faveur des procaryotes, et en faveur des espèces d'intérêt médicales ou agronomiques. En 2017, la majorité des séquences disponibles couvrait déjà principalement des organismes procaryotes (Figure 8) (Keeling and Campo 2017). C'est toujours le cas aujourd'hui. Sur la base de données du Joint Genome Institute (<https://gold.jgi.doe.gov/statistics>), on recense 89 192 projets de génomique étudiant des eucaryotes pour 223 742 projets de génomiques ciblant des bactéries en 2023. Parmi les eucaryotes représentés dans les bases de données, peu sont des espèces marines (Keeling et al. 2014), tandis qu'on trouve une grande quantité de génomes provenant de plantes, d'animaux, de champignons ou de parasites (del Campo et al. 2014). Ce biais persiste en 2023 : sur l'ensemble des projets disponibles sur JGI, 61.8 % concernent des plantes terrestres (Streptophyta), 25 % des champignons (Ascomycètes et Basidiomycètes) et 3.4 % des Chordés (<https://gold.jgi.doe.gov/distribution>). La plupart des espèces pour lesquelles on dispose d'un génome complet sont des espèces cultivables, en général peu abondantes dans l'environnement, et distantes des espèces qui dominent les écosystèmes. Ce biais crée des difficultés dans l'affiliation des séquences environnementales.

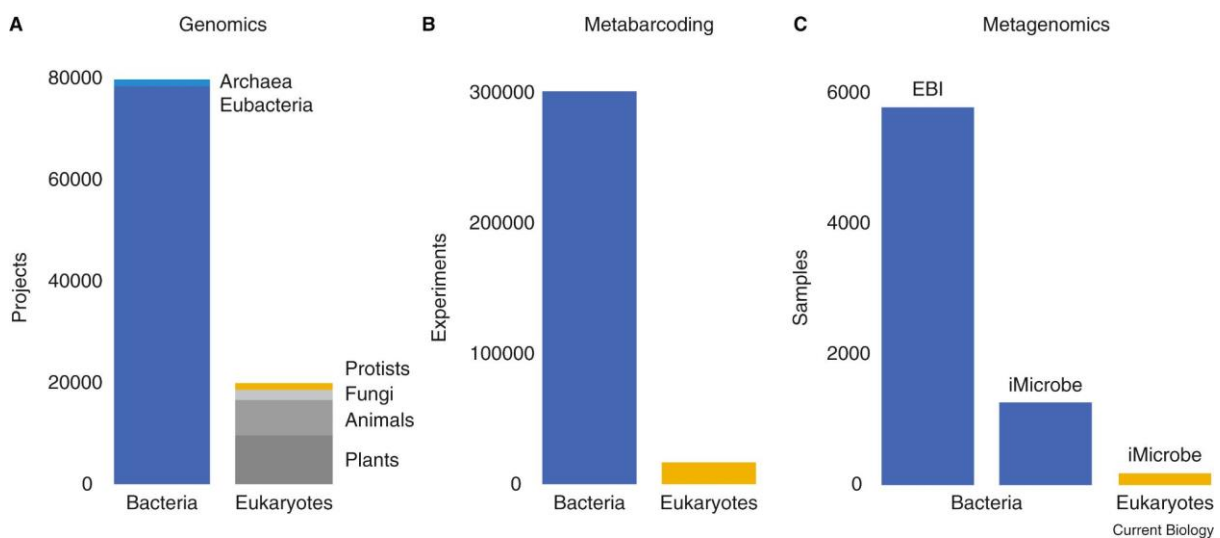


Figure 8. Proportion de données génomiques affiliées à des procaryotes (bleu) ou à des eucaryotes (protiste en orange, autres eucaryotes en échelles de gris) dans les différentes bases de données de génomique (The Genomes OnLine Database), métabarcoding (the Sequence Read Archive) et métagénomique /métagénomique (EBI Metagenomes et iMicrobe). D'après Keeling and Campo 2017.

Il existe encore beaucoup de gènes dont nous ignorons les fonctions. En 2018, un catalogue de gènes eucaryotes est construit à partir des données *Tara Oceans*, et 51.2 % des gènes n'ont aucun homologue dans les bases de données (Carradec et al. 2018). On estime qu'on peut annoter  $\frac{1}{3}$  des gènes avec certitude ;  $\frac{1}{3}$  font l'objet

d'hypothèses ; nous sommes incapables de faire des prédictions pour le tiers restant (Weissenbach and Sghir 2016). La majorité des fonctions convenablement annotées appartient au métabolisme central, comme la photosynthèse, le cycle de Krebs, la glycolyse etc. Les fonctions accessoires sont peu ou pas décrites, rendant difficile ou impossible de les annoter avec confiance (Worden and Allen 2010).

Une difficulté supplémentaire pour l'étude des eucaryotes est la complexité de leurs génomes. Dans le cas des procaryotes, il est possible de décrire les voies biochimiques sur la base de la présence des gènes codant pour les enzymes impliquées. En comparaison, le mode de vie des eucaryotes dépend de nombreux facteurs qui ne sont pas systématiquement déchiffrables d'après le génome. Les modes de vie sont parfois difficiles à identifier sur la seule base des séquences. Si la présence du chloroplaste indique de manière évidente le mode de vie autotrophe de l'organisme, les gènes marqueurs évidents pour l'hétérotrophie ou la mixotrophie sont plus rares. Les symbioses ou le parasitisme sont fréquents dans les protistes marins (Vargas et al. 2015) mais peu d'outils génétiques permettent de les inférer à partir du génome. Les protistes peuvent présenter des organelles complexes avec des fonctions de protection (frustule, thèque), de mouvement (flagelle, cils), de perception de l'environnement (lumière, salinité, etc.) et de formation de colonies ou de prédation (Worden et al. 2015). Ces structures génèrent des comportements particuliers chez ces cellules, dont les mécanismes moléculaires sont souvent inconnus et qui pourtant influencent leur écologie.

Les études environnementales menées grâce à la génomique environnementale souffrent également de plusieurs biais techniques pouvant interférer à chaque étape du protocole (Figure 9A et 9B). Tout d'abord, le choix de la méthode de conservation des échantillons ou d'extraction des acides nucléiques va fortement impacter les résultats d'une analyse (Mäki et al. 2017). Les étapes de PCR et de séquençage peuvent aussi amener leurs lots de biais : le choix du gène marqueur ou de la température d'hybridation, la mise en place ou non de réplicats, le choix de la technologie de séquençage, etc., sont autant de biais pouvant amener à favoriser ou défavoriser certaines séquences par rapport aux autres (van der Loos and Nijland 2021). La nature de la séquence peut également jouer : les séquences longues ou riches en GC sont moins amplifiées lors de la PCR (Krehenwinkel et al. 2017). Ainsi, chaque étape du protocole est susceptible d'amener son lot de biais, ce qui peut mener à des variations drastiques de résultats entre deux analyses menées avec un protocole différent à partir du même échantillon (Figure 9C).

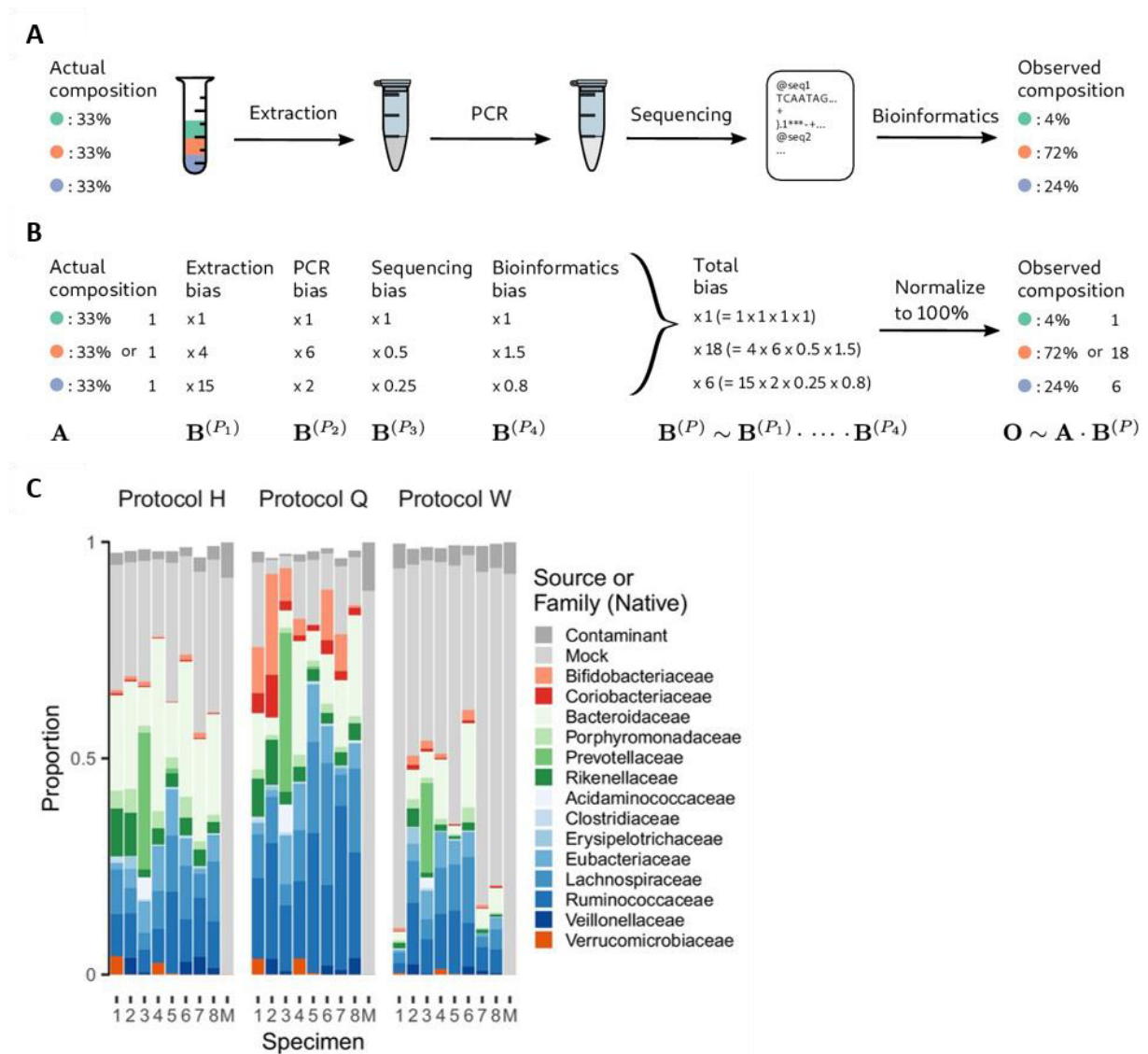


Figure 9. Accumulation des biais technique en génomique environnementale et impact sur l'analyse. A) Protocole standard appliqué à un échantillon environnemental. L'addition des biais inhérents à chaque étape peut amener à une observation très différente de la composition réelle. B) Modèle mathématique des biais du protocole. Les biais d'extraction, d'amplification, de séquençage et d'analyse s'accumulent, créant un biais globale propre à chaque protocole. C) Variation de la composition en espèce de 9 échantillons de microbiote en fonction du protocole appliqué. D'après McLaren, Willis, and Callahan 2019.

Outre ces biais techniques, la quantification des données génomiques pose un véritable défi. Pour le métabarcoding par exemple, le nombre de copies du gène marqueur peut grandement varier selon les espèces et les genres et introduire des erreurs dans l'estimation de l'abondance. C'est le cas pour le plancton marin eucaryote, où le nombre de copie du gène codant pour l'ADNr 18S peut aller d'une à mille (Not et al. 2009; Mäki et al. 2017; Santi et al. 2021). Pour la métagénomique et la métatranscriptomique, les recherches actuelles sont cantonnées à des quantifications relatives : on ne regarde pas l'abondance d'un organisme dans l'absolu, mais relativement aux autres organismes de l'échantillon. De manière équivalente, on étudie l'expression d'un gène par rapport à celles des autres. Ce biais limite fortement les

analyses : impossible de comparer la composition de plusieurs échantillons si leur biomasse totale diffère trop, et difficile de déterminer l'ampleur absolue des variations observées (Vandeputte et al. 2017). La combinaison avec des données de quantifications du nombre de cellules comme par exemple la cytométrie en flux ou le comptage par microscopie peuvent permettre de quantifier précisément les abondances d'ADN et d'ARN (Props et al. 2017). Bien que l'intégration de ces techniques en génomique environnementale est encore rare, la quantification des données de méta-omiques à partir du nombre de cellules ou de la quantité de matériel génétique peut drastiquement impacter les conclusions que l'on peut tirer d'un même échantillon (Figure 10).

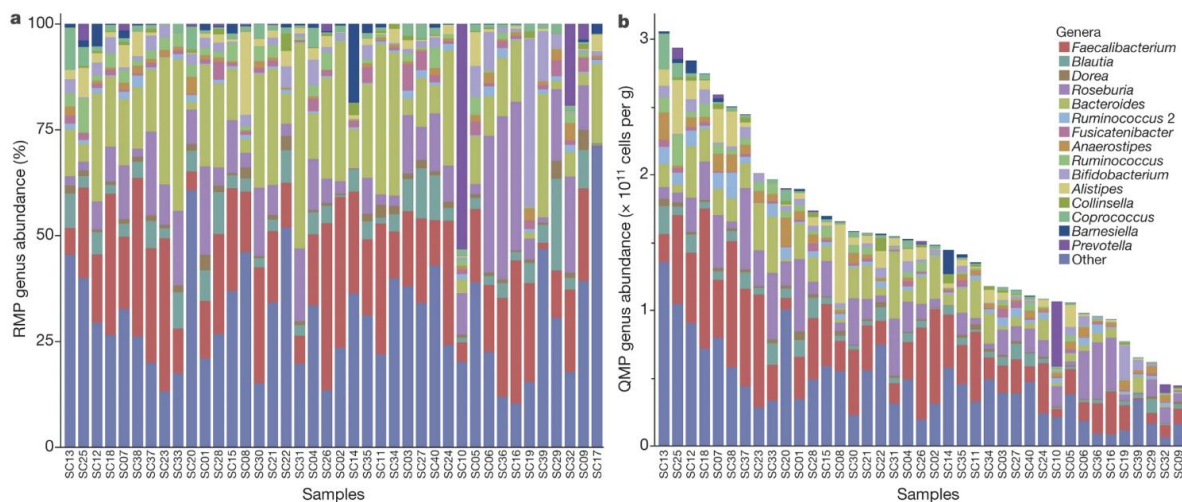


Figure 10. Composition en genre bactérien de 40 échantillons de microbiote intestinal. A) Profil d'abondance relative obtenu avec un protocole standard de séquençage des ADNr 16S. B) Profil d'abondance quantitative obtenu en complétant le compte des séquences avec un compte du nombre de cellule. D'après Vandeputte et al. 2017.

Les méthodes de génomique environnementale ont permis dans les années 2000 de surmonter le biais de la culture. Mais depuis 2015, l'omniprésence des études basées sur l'analyse de séquences est questionnée. Pour compenser les biais de la génomique environnementale, le couplage avec l'expérimentation en laboratoire est de plus en plus plébiscité. En effet, le laboratoire reste le lieu idéal pour tester la fonction des gènes et étudier la physiologie dans des conditions contrôlées. De plus, c'est en séquençant de nombreuses espèces cultivables et en testant expérimentalement les fonctions des gènes qu'il devient possible de combler les lacunes de nos connaissances (Worden and Allen 2010). Inversement, la génomique environnementale est malgré ces biais inhérents un puissant complément aux études de physiologie en laboratoire. Elle permet de capturer la complexité des réponses physiologiques dans l'environnement, telles que les co-limitations de nutriments ou les interactions avec d'autres organismes, alors que les expériences en laboratoire ne permettent pas de saisir ou de simuler la complexité des conditions environnementales (Worden et al. 2015).

En 30 ans de génomique et 20 de métagénomique, l'étude des protistes marins au

laboratoire et dans l'environnement a permis la compréhension de nombreux phénomènes tels que l'origine de la multicellularité, l'évolution des introns, la méthylation, la compaction de la chromatine, le cycle du carbone, le réseau trophique ou les symbioses. Aujourd'hui, il est nécessaire de développer des nouvelles stratégies de travail pour surmonter les freins évoqués précédemment. Worden et al. (2015) propose des pistes pour donner à la génomique environnementale un nouvel élan, en couplant travail de paillasse, travail de terrain et travail d'analyse. Ses suggestions comprennent entre autres :

- Définir de nouvelles espèces modèles, couvrant des branches encore peu étudiées de l'arbre du vivant, et ayant une importance écologique significative.
- Développer des outils de génie génétique pour ces nouvelles espèces modèles.
- Échantillonner de manière à préserver les contacts cellules/cellules.
- Utiliser l'imagerie de manière systématique dans les études environnementales.
- Mener des expériences dans des bioréacteurs afin de simuler de multiples variations de conditions simultanément et se rapprocher des conditions naturelles.

## 2 - Le phytoplancton dans un océan oligotrophe

Un milieu est dit oligotrophe lorsqu'il est pauvre en substance nutritive. La carence d'un élément essentiel à la croissance des organismes peut fortement impacter leur distribution et leur abondance, que cette carence soit temporaire ou durable. Parmi les nutriments essentiels au développement du phytoplancton, on retrouve l'azote (N), le phosphore (P), le silicate (Si), les métaux traces comme le fer (Fe), le manganèse (Mg) ou le cobalt (Co) (Bristow et al. 2017; Browning and Moore 2023). Dans cette thèse, nous nous concentrerons uniquement sur les effets des manques d'azote et de fer sur le phytoplancton eucaryote.

### L'azote, une ressource inégalement distribuée

Le cycle de l'azote marin et son assimilation par le phytoplancton

Parmi les cycles biogéochimiques qui font circuler les éléments au sein de l'écosystème, le cycle de l'azote marin fait l'objet d'un intérêt accru depuis la moitié des années 1990, quand on a commencé à comprendre son fonctionnement global. Aujourd'hui, on découvre encore régulièrement un nouveau processus ou une nouvelle espèce impliquée.

Tout commence avec le diazote atmosphérique ( $N_2$ ), qui entre dans l'eau de mer par diffusion et est fixé sous forme d'ammonium ( $NH_4^+$ ) par les bactéries fixatrices d'azote, dites diazotrophes. Une fois fixé, l'ammonium peut être assimilé par le phytoplancton eucaryote ou bien être nitrifié en nitrite ( $NO_2^-$ ) par les archées oxydatrices d'ammonium, puis en nitrate ( $NO_3^-$ ) par les bactéries oxydatrices de nitrite. Au même titre que l'ammonium, le nitrite et le nitrate sont également des formes d'azote biodisponibles pour le phytoplancton eucaryote. L'azote sous forme d'ammonium, de

nitrate ou de nitrite est ainsi assimilé par le phytoplancton dans la matière organique via la biosynthèse des acides aminés et acides nucléiques. Ces étapes prennent place dans la zone photique. Les formes d'azote organiques générées par le phytoplancton peuvent être exportées vers le fond par précipitation de la matière organique. Elles vont être désaminé en ammonium avant d'être de nouveau nitrifié successivement en nitrite et en nitrate par des micro-organismes diazotrophes. Le nitrate généré en profondeur peut soit remonter vers la surface avec les masses d'eau, soit être dénitrifié. Des micro-organismes vont le métaboliser en oxyde nitrique, puis en oxyde nitreux, et finalement en diazote lors de leur respiration anaérobie, utilisant ces composés comme des accepteurs d'électrons successifs. En fonction de la profondeur à laquelle ces processus ont lieu, le diazote peut être libéré dans l'océan ou dans l'atmosphère, ou bien stocké dans les sédiments (David A. Hutchins and Capone 2022) (Figure 11).

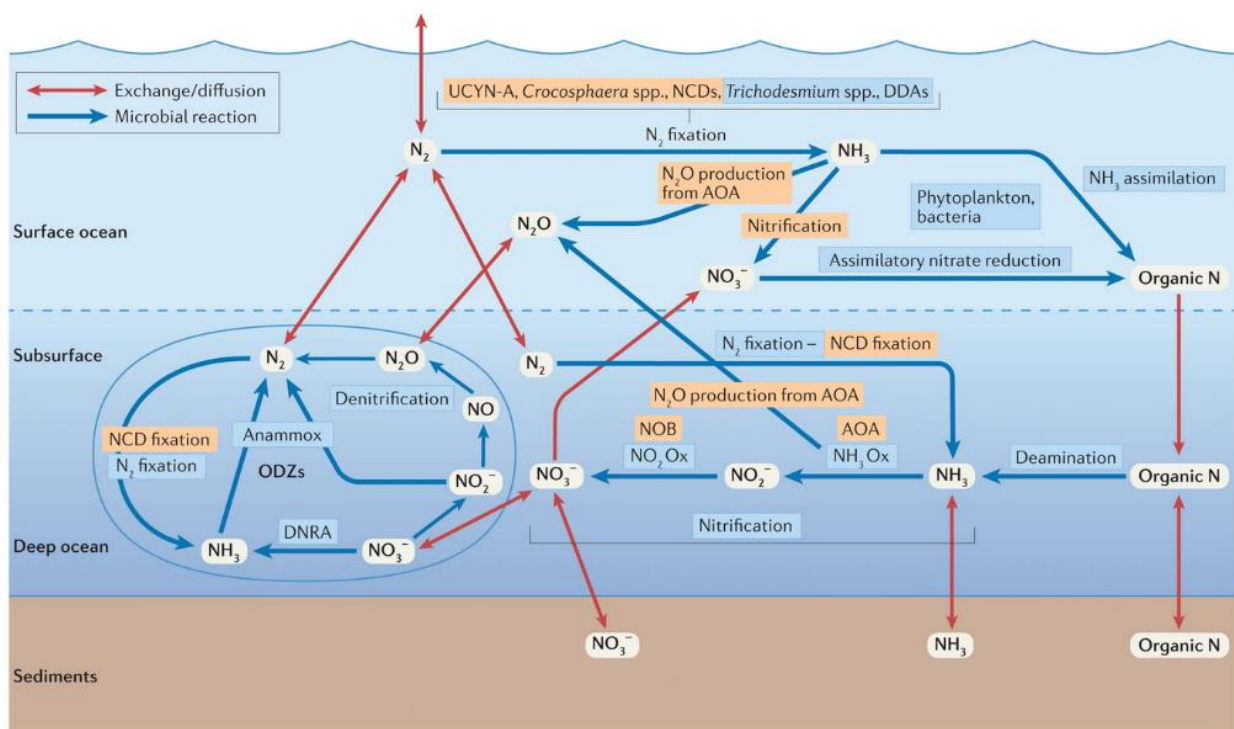


Figure 11. Cycle de l'azote marin. Les flèches rouges désignent les processus d'échange et de diffusion, et les bleus les réactions réalisées par des micro-organismes. Le phytoplancton assimile l'ammonium et le nitrate et relargue de la matière organique dans la zone photique (surface ocean). Des processus de transformation de l'azote ont également lieu dans les sédiments, mais les molécules y restent piégées sur des périodes de temps géologiques. Modifié d'après Hutchins and Capone 2022.

L'assimilation de l'azote inorganique par le phytoplancton dépend de la molécule (Figure 12). Sous la forme de nitrate ou de nitrite, l'assimilation a lieu dans le chloroplaste. Dans le cytoplasme, le nitrate est réduit en nitrite par l'activité de la nitrate réductase, avant d'être transporté dans le chloroplaste. C'est ici que la nitrite réductase va transformer le nitrite en ammonium, la forme d'azote qui va être incorporée dans la matière organique. Pris en charge dans le cycle de la glutamine synthétase/glutamate synthase (GS/GOGAT), l'azote sous forme de groupe amine va entrer dans la glutamine grâce à l'activité de la glutamine synthétase (GS) requérant du glutamate et de l'ATP,

puis il va être transféré à l' $\alpha$ -oxoglutarate et générer 2 glutamates par l'activité de la glutamate synthase (GOGAT). Le glutamate ainsi généré va être soit réintégré dans le cycle GS/GOGAT, soit subir une transamination, menant l'azote à participer à la biosynthèse des acides aminés, nucléiques, des protéines ou d'autres composés azotés (Sanz-Luque et al. 2015). Si la source d'azote puisée dans l'environnement est l'ammonium, le métabolisme d'assimilation sera très similaire, mais aura principalement lieu dans la mitochondrie. Des transporteurs d'ammonium le font entrer dans la cellule puis dans la mitochondrie ou le chloroplaste. Ensuite il sera soit pris en charge par le cycle de la GS/GOGAT (dans le chloroplaste ou la mitochondrie), soit métabolisé avec un ion bicarbonate en carbamyl-phosphate par la carbamyl-phosphate synthétase (CPS), et entrer dans le cycle de l'urée (dans la mitochondrie).

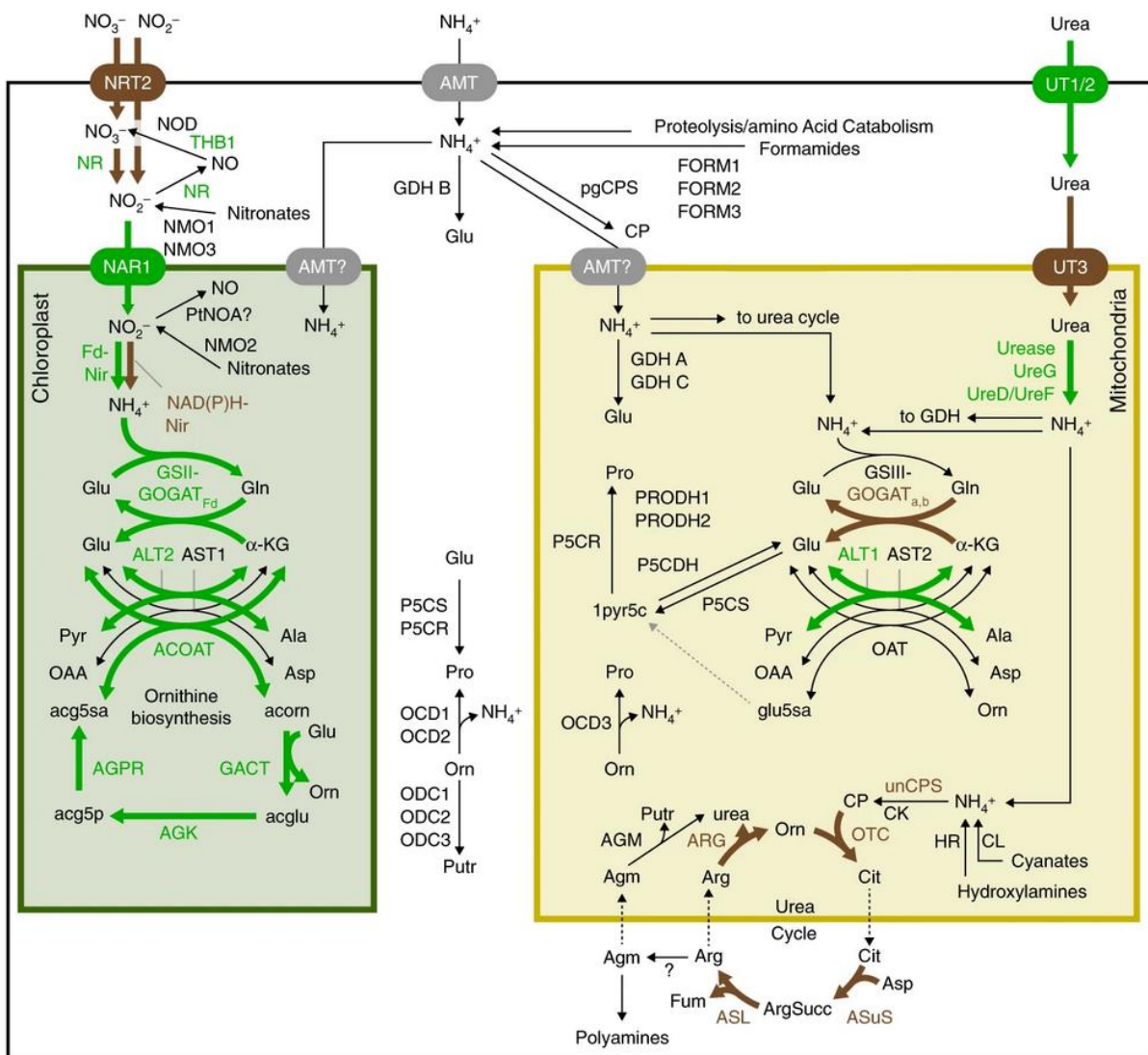


Figure 12. L'assimilation de l'azote chez la diatomée *Phaeodactylum tricornutum*. La couleur des flèches indique l'origine évolutive des voies : en vert, les voies provenant de l'endosymbiote photosynthétique, en marron, celles provenant de l'algue rouge primitive. Les noms des métabolites sont écrits en noirs. Transporteurs. NRT2 : transporteur de nitrate/nitrite; NAR1 : transporteur de nitrite; AMT : transporteur d'ammonium; UT : transporteur d'urée. Voie d'assimilation de l'azote. NR : nitrate réductase; Fd-Nir : nitrite réductase; GS : glutamine synthétase; GOGAT : glutamate synthase. Cycle de l'urée. CPS : carbamyl-phosphate

synthétase; OTC : ornithine transcarbamylyase; AsuS : argininosuccinate synthase; ASL : argininosuccinate lyase; ARG : arginase. Biosynthèse des acides aminés. ALT1/ALT2 : alanine transaminase. Biosynthèse de l'arginine. ACOAT : acétylornithine aminotransférase; AGPR : N-acétyl-gamma-glutamyl-phosphate réductase; GACT : glutamate N-acétyltransférase; AGK : acétylglutamate kinase. D'après Smith et al. 2019.

Le cycle de l'urée n'est pas présent dans l'ensemble du phytoplancton : les espèces issues de l'endosymbiose primaire (par exemple les Chlorophytes) en sont dépourvus (Allen et al. 2011). Le carbamyl-phosphate est transformé en une série de métabolites par une suite d'activités enzymatiques : l'ornithine transcarbamylyase condense le carbamyl-phosphate et l'ornithine pour générer de la citrulline. Diffusée dans le cytoplasme, cette dernière est prise en charge par l'arginosuccinate synthase et condensée avec de l'aspartate pour former de l'arginosuccinate. Ce dernier est métabolisé par l'argininosuccinate lyase, libérant du fumarate et de l'arginine. Finalement, l'arginase hydrolyse cette dernière en ornithine et en urée, bouclant ainsi le cycle. Très similaire au cycle de l'urée permettant aux métazoaires d'éliminer l'azote fixé, celui des diatomées permet de redistribuer le carbone et l'azote dans la cellule, notamment en réponse à des variations d'azote disponible (Allen et al. 2011).

La variabilité spatiale et temporelle de la disponibilité de l'azote marin

Le phytoplancton est une composante centrale du cycle de l'azote, car il permet d'incorporer l'azote inorganique (sous forme d'ammonium, de nitrate ou de nitrite) dans la matière organique, le rendant biodisponible pour les échelons trophiques supérieurs. Réciproquement, l'azote est essentiel à la croissance du phytoplancton car il participe à de nombreux processus biologiques tels que la photosynthèse, la biosynthèse d'acides aminés et d'acides nucléiques. Ainsi, la disponibilité du nitrate et de l'ammonium peut affecter la production primaire, et leur distribution régie celle du phytoplancton (Moore et al. 2013).

On a longtemps pensé que l'azote fixé ne pouvait pas être un élément limitant la croissance du phytoplancton dans l'environnement. On considérait que le diazote atmosphérique formait un réservoir infini pour les bactéries diazotrophes qui pouvaient le fixer constamment en ammonium (Zehr and Kudela 2011). Cependant, la fixation de l'azote dépend de nombreux autres paramètres, notamment la concentration de métaux traces. Ainsi, la plupart des eaux de surface sont en réalité déplétées en azote inorganique (Figure 13). L'ammonium est une source d'azote très labile, présentant un renouvellement rapide, généré soit par fixation du diazote, soit par relargage par la biomasse planctonique. On le qualifie de réservoir d'azote "régénéré". Le nitrate en revanche, est qualifié de réservoir d'azote "exogène". C'est la forme d'azote inorganique la plus abondante dans les océans, allant d'une concentration non-délectable jusqu'à 35  $\mu\text{M}$  (Voss et al. 2013). Très rapidement assimilé par le phytoplancton, il peut être supplémenté par la nitrification de l'ammonium, ou bien par des apports exogènes comme les remontées d'eau profonde riche en nutriments, les dépôts atmosphériques, le lavage des côtes par ruissellements des eaux pluviales, les apports anthropogéniques, etc. (Zehr and Kudela 2011).

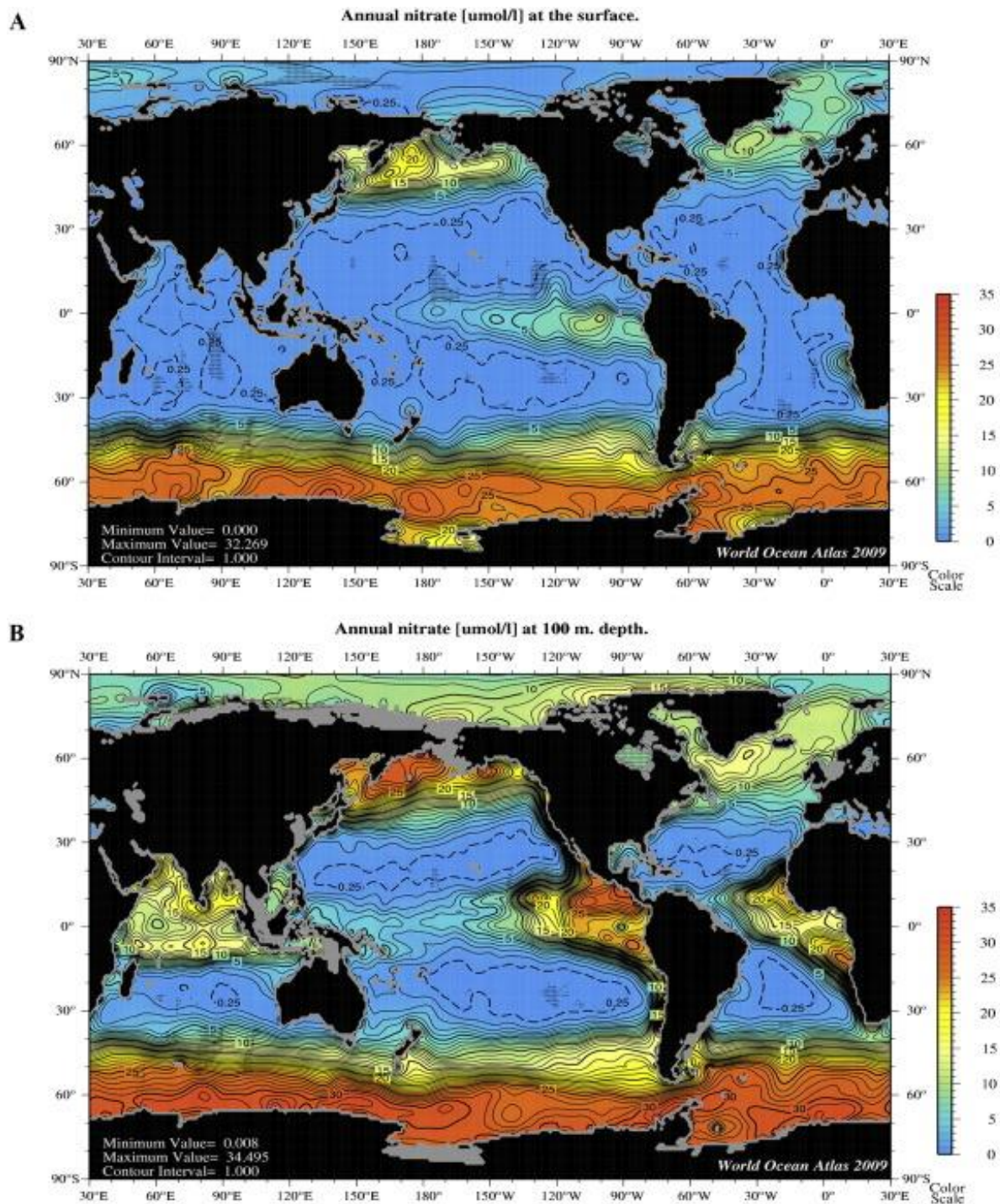


Figure 13. Concentration annuelle moyenne de nitrate dans A) les eaux de surface et B) à 100 m de profondeur. La couleur indique la concentration de nitrate en  $\mu\text{M}$ . Basée sur les données du World Ocean Atlas 2009. D'après Rogato et al. 2015.

La concentration des nutriments (nitrate, nitrite, ammonium, phosphate) dans l'eau de mer suit une distribution verticale marquée : ils sont faiblement concentrés dans la zone photique, et leur concentration augmente avec la profondeur (Figure 14). La remontée des eaux profondes est l'un des plus importants apports de nitrate vers les eaux de surface (Voss et al. 2013). Comme évoqué précédemment, une grande part de l'azote organique précipite (mort cellulaire, décomposition, fèces, etc.) et est

successivement désaminé et nitrifié dans les strates inférieures. Les masses d'eaux présentant des propriétés chimiques différentes (température, salinité, oxygénation, densité) forment naturellement des couches distinctes, et les transferts entre elles sont rares : c'est la stratification. Des phénomènes de mélanges peuvent abolir cette stratification, tels que des courants ascendants (upwelling), des tourbillons (eddies), ou encore des variations de température de l'eau de surface, saisonnières ou exceptionnelles.

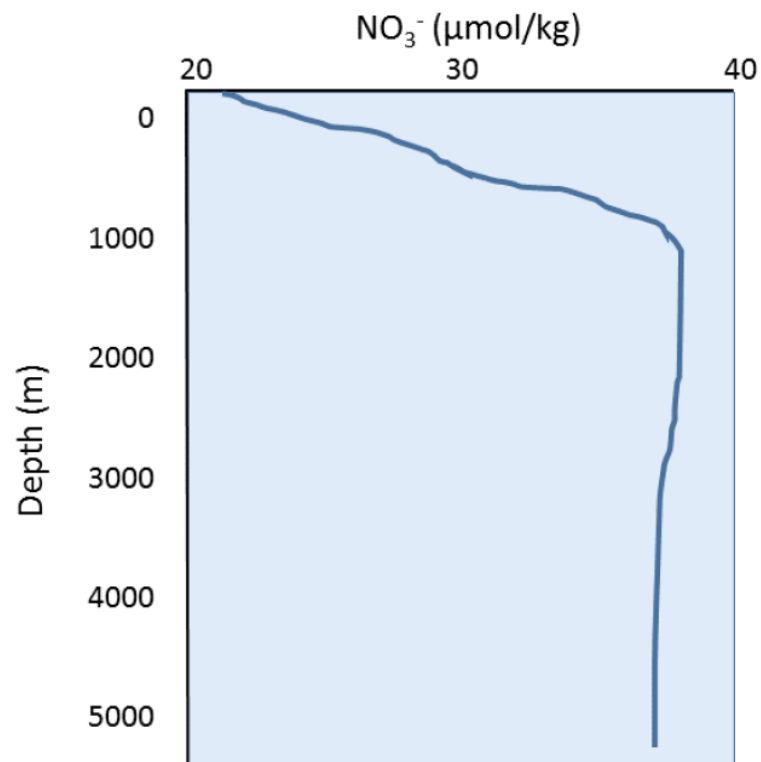


Figure 14. Variation de la concentration de nitrate dans la colonne d'eau. Les profils pour les autres nutriments (phosphate, silicate) sont similaires. La zone photique se trouve dans les 150 premiers mètres. D'après Webb 2021

La stratification permet aussi d'expliquer la variabilité spatiale et temporelle de la concentration des nutriments. Celle-ci varie grandement en fonction du bassin océanique. En général, l'océan pélagique est oligotrophe, c'est-à-dire que les concentrations de nutriments y sont très faibles, particulièrement dans la zone photique. Les exceptions sont les eaux polaires (la faible température de l'eau de surface favorise la remontée des eaux profondes et donc l'apport des nutriments), les zones d'upwelling (de forts courants marins font remonter les eaux profondes), et les zones côtières, où les apports terrigènes et anthropogéniques sont courants (Levitus et al. 1993). De manière saisonnière, il est naturel qu'en hiver les eaux de surface refroidissent, abolissant la stratification et permettant la remontée des nutriments. Au contraire, l'augmentation de la température des eaux de surface en été limite ce mélange en renforçant la stratification.

## L'acclimatation du phytoplancton au manque d'azote fixé

La majorité des eaux de surface pélagiques étant appauvries en azote inorganique, le phytoplancton doit développer des stratégies pour s'acclimater aux variations de disponibilité du nitrate et de l'ammonium. Une diminution de la taille est un avantage majeur pour faire face au manque d'azote inorganique, et de nutriments en général. En effet, les cellules de petite taille présentent un meilleur ratio surface-volume, ce qui facilite la diffusion au travers de la membrane plasmique (Raven 1998), et les besoins nutritifs de la cellule sont proportionnels à son volume (Shuter 1978; Litchman et al. 2007). En général, le picophytoplancton est plus abondant dans les eaux pauvres en nutriments que le nano- ou le micro-phytoplancton (Cabré et al. 2016), ce qui en fait un acteur essentiel de l'assimilation du nitrate dans les zones oligotrophes.

Une autre stratégie du phytoplancton pour faire face aux déplétions en azote inorganique est l'optimisation de la captation de l'ammonium et du nitrate en augmentant l'expression des transporteurs dédiés. Les diatomées présentent 2 formes de transporteurs : à faible affinité et à forte affinité. Les transporteurs à forte affinité pour l'ammonium (AMT) ou pour le nitrate (NRT2) sont particulièrement efficaces pour capter l'azote inorganique quand il est faiblement concentré. Chez les diatomées, NRT2 est surexprimé en réponse à la carence en nitrate (Mock et al. 2008; Allen et al. 2011; Ashworth et al. 2013; S. Bender et al. 2014). En revanche, l'expression des AMT est plus variable selon les espèces. Des études menées en laboratoires ont montré qu'ils étaient généralement surexprimés chez *Phaeodactylum tricornutum*, *Cylindrotheca fusiformis*, *Fragilariopsis cylindrus*, *Pseudo-nitzschia multiseriis* lors de privation ou de carence en nitrate, mais sous-exprimés chez *Thalassiosira pseudonana* (Hildebrand 2005; S. J. Bender, Parker, and Armbrust 2012; S. Bender et al. 2014; Levitan et al. 2015). La régulation des transporteurs de nitrate à faible affinité (NPF) semble plutôt dépendre de l'intensité lumineuse (Santin et al. 2021).

L'allocation parcimonieuse des ressources en azote est une autre stratégie courante pour faire face aux privations. La photosynthèse, le métabolisme central du carbone et la biosynthèse des protéines sont réduits afin de limiter les besoins en azote. Chez *T. pseudonana*, les protéines impliquées dans la photosynthèse et dans la biosynthèse des protéines (protéines ribosomiques, ARNt synthétase) sont réduites quand le nitrate est limité (Hockin et al. 2012). Chez cette même diatomée ainsi que chez *F. cylindrus* et *P. multiseriis*, la limitation en nitrate provoque une diminution de l'expression des gènes impliqués dans la fixation du carbone (cycle de Calvin) et le métabolisme des carbohydrates et des acides gras (S. Bender et al. 2014). Chez *P. tricornutum*, c'est aussi la fixation et le métabolisme du carbone qui sont réduits (McCarthy et al. 2017) lorsque le nitrate n'est plus assimilé. La photosynthèse et la biosynthèse des protéines ribosomiques sont aussi régulées à la baisse lorsque la concentration de nitrate diminue (Remmers et al. 2018). Le métabolisme du carbone est réorienté en faveur du stockage de l'énergie, et l'activité photosynthétique est réduite (Scarsini et al. 2022).

Le phytoplancton est également capable de recycler l'azote intracellulaire. Par exemple,

les diatomées sont capables de dégrader leurs acides aminés et leurs protéines afin de récupérer l'azote de ces molécules. Chez *P. tricornutum*, un manque de nitrate dans le milieu se traduit par une diminution de la quantité d'acides aminés intracellulaires (notamment de l'acide glutamique, de l'acide aspartique, de l'histidine et de la sérine) et par une augmentation du nombre de protéines impliquées dans le catabolisme des acides aminés (aminotransférase) et des protéines (protéases) (Hockin et al. 2012). Chez *T. pseudonana*, *F. cylindrus* et *P. multiseriis*, les gènes codant pour des splicéosomes ou impliqués dans le métabolisme des protéines et des acides aminés représentent respectivement 8 %, 4 % et 4 % des transcrits surexprimés dans des conditions de privations de nitrate. Chez *P. tricornutum*, l'augmentation de l'abondance des transcrits de gènes codant pour des désaminases, des acétamidases et des formamidases indiquent un catabolisme des acides aminés dans des conditions de privation d'azote (Alipanah et al. 2015). L'azote récupéré par ce recyclage est réorienté vers l'assimilation par le cycle de l'urée ou le cycle GS/GOGAT, et le carbone fixé par photosynthèse est dirigé vers la biosynthèse des lipides (Levitan et al. 2015). Cela permet aux cellules de stocker l'azote et le carbone puis de les remobiliser rapidement, lorsque les conditions nutritionnelles redeviennent favorables (Scarsini et al. 2022).

Beaucoup d'organismes seraient capables d'accumuler l'azote inorganique, construisant des stocks pour faire face aux périodes de privation. Les enzymes GS/GOGAT sont considérées comme une forme de stockage de l'azote dans le glutamate (Hockin et al. 2012). Le stockage intracellulaire du nitrate dans des structures dédiées a été prouvé chez un certain nombre de phyla phytoplanctoniques, notamment chez les Chlorophytes, les Haptophytes, les diatomées et les dinoflagellés (Kamp et al. 2015). Le nitrate peut être directement stocké dans des vacuoles chez les diatomées : le nitrate capté, mais non-assimilable par un *P. tricornutum* mutant privé de nitrate réductase s'accumule dans des vacuoles (McCarthy et al. 2017). Chez les espèces de dinoflagellés symbiotiques du genre *Symbiodinium*, la formation de cristaux de guanine intracellulaire peut servir de système de stockage de l'azote dans des situations de privation (Kopp et al. 2013). Il a été montré récemment que ce mécanisme pourrait être bien plus répandu dans le phytoplancton que ce que l'on pensait initialement (Pilátová et al. 2022).

Une dernière stratégie pour faire face au manque d'azote inorganique est de se tourner vers des sources d'azote organiques dissoutes, ou DON (Kumar and Bera 2020). La DON forme un réservoir d'importance primordiale, dont le phytoplancton est aussi la source principale. Elle est produite soit par libération active et exsudation, soit par la diffusion passive de métabolites à travers les membranes, soit par des phénomènes biotiques (broutage, lyse virale, mort cellulaire, décomposition). Il est estimé que le picophytoplancton génère plus de la moitié de la DON dans les zones oligotrophes (Varela et al. 2005). Sa composition précise n'est que partiellement caractérisée (beaucoup de molécules sont inconnues), mais elle est composée de 2 fractions : l'une réfractaire, l'autre labile (Bronk et al. 2007). Cette dernière est biodisponible et contribue à la croissance du phytoplancton et à la productivité primaire, notamment

dans les environnements côtiers et estuariens. La DON labile contient un grand nombre de composés tels que l'urée, les acides aminés et les acides nucléiques (Sipler and Bronk 2015). Il a été montré que ces composés peuvent soutenir la croissance du phytoplancton dans des conditions de faible teneur en azote.

L'urée représente seulement 6 à 19 % de l'ensemble du stock de DON (Sipler and Bronk 2015) mais il s'agit d'une source d'azote organique de première importance pour le phytoplancton. Beaucoup d'espèces peuvent l'incorporer grâce aux transporteurs d'urée et le métaboliser en ammonium grâce à l'uréase. L'urée peut être d'origine exogène ou intracellulaire, et l'activité de l'uréase a lieu dans le cytosol (Kumar and Bera 2020). Elle peut soutenir la croissance d'espèces formant des proliférations d'algues toxiques telles qu'*Aureococcus anophagefferens* (Berg et al. 1997). Les transporteurs d'urée sont surexprimés dans des conditions de carence en azote chez les diatomées *P. tricornutum*, *T. pseudonana*, *F. cylindrus* et *P. multiseriens* (S. Bender et al. 2014; Smith et al. 2019; Kumar and Bera 2020) et des études d'amendements en nitrate *in situ* montrent qu'elles expriment les gènes leur permettant de privilégier l'urée quand le nitrate est très faiblement concentré (Lampe et al. 2019).

Les acides aminés et les acides nucléiques forment un réservoir très labile de DON. Les premiers peuvent être désaminés par l'aminooxydase ou transaminés par les alanine aminotransférases ou les aspartate aminotransférases (Kumar and Bera 2020). Les seconds sont principalement représentés par les purines. Une série d'enzymes permettent de les dégrader en un ensemble de sous-produits tels que la xanthine, l'hydroxanthine, l'urate et l'urée, à partir desquels il est possible de générer de l'ammonium (Gann et al. 2022). Il a été montré qu'*E. huxleyi* peut utiliser de l'acétamide ou du formamide, ainsi que de l'urée, de l'hydroxyurée, de l'hydroxanthine ou des purines comme source d'azote organique (Palenik and Henson 1997).

## **Le fer, un micronutriment rare**

### Le fer dans l'océan

Le fer est un micronutriment essentiel au phytoplancton. Il entre dans la composition de nombreuses enzymes et transporteurs d'électrons impliqués dans la photosynthèse, la respiration, la réduction des composés nitrés ou sulfurés, et la fixation de l'azote (John A. Raven 2013). Le fer joue un rôle catalytique dans des enzymes tels que le photosystème I (composé de 12 atomes de fer), le cytochrome *b6f* (6 atomes), le cytochrome *c6* (1 atome), et les superoxydes dismutases à fer (Geider and La Roche 1994; Morrissey and Bowler 2012). Ainsi, un accès réduit à cet élément se traduit par une diminution du taux de croissance et de l'activité photosynthétique, et une augmentation du stress oxydatif (Morrissey and Bowler 2012).

Jusque dans les années 1990, l'importance du fer dans le métabolisme phytoplanctonique était sous-estimée. On observait dans l'océan de larges zones riches en nutriments mais pauvres en chlorophylle (HNLC). Lors d'expériences menées

dans les années 1990, du fer a été déversé dans ces zones, résultant en un regain de croissance du phytoplancton (Boyd et al. 2007). Ainsi, l'enrichissement en fer pouvait provoquer le doublement de la biomasse phytoplanctonique, le triplement de la concentration de chlorophylle, et la multiplication par 4 de la production primaire (Coale et al. 1996). On estime que la croissance du phytoplancton est limitée sur 30 à 40 % de la surface océanique à cause du manque de fer, alors que la quantité d'autres nutriments est suffisante (Boyd and Ellwood 2010).

En effet, le fer est un élément trace, c'est-à-dire que ces concentrations sont naturellement très faibles, de l'ordre du nanomolaire (Figure 15A). Il est plus abondant le long des côtes, où les apports proviennent principalement des sédiments transportés par les rivières (625 à 962 Mt/an), de la fonte des glaciers (34 à 211 Mt/an) ou des rejets anthropogéniques. Dans les zones pélagiques, la concentration y est si faible qu'il est difficile de la mesurer *in situ*. Le fer provient notamment d'apports aériens, arrivant sous forme de poussières transportés par les vents depuis les zones arides (16 Mt/an) (Jickells et al. 2005). Ainsi la Méditerranée, la mer Rouge, le Golfe Persique et la mer d'Arabie sont particulièrement enrichis en fer grâce à la proximité du Sahara. Au contraire, le Pacifique et l'océan Austral font partie de ces zones HNLC où le fer limite la croissance du phytoplancton (Figure 15B).

Le fer dissous peut être trouvé sous plusieurs formes, dont la biodisponibilité va principalement dépendre de sa chélation. Ainsi, le fer inorganique non chélaté  $Fe'$  sera plus facilement assimilable que le fer chélaté à un ligand organique  $FeL$ . De plus, le fer peut présenter différents niveaux d'oxydation. Ainsi, on le retrouve (lié à un ligand ou non) sous forme d'ions ferreux  $Fe^{2+}$  ou d'ions ferrique  $Fe^{3+}$ . Le premier est thermodynamiquement instable dans l'eau de mer, mais serait plus facilement assimilable par le phytoplancton. À l'inverse, le second est stable, mais serait moins biodisponible (Shaked and Lis 2012).

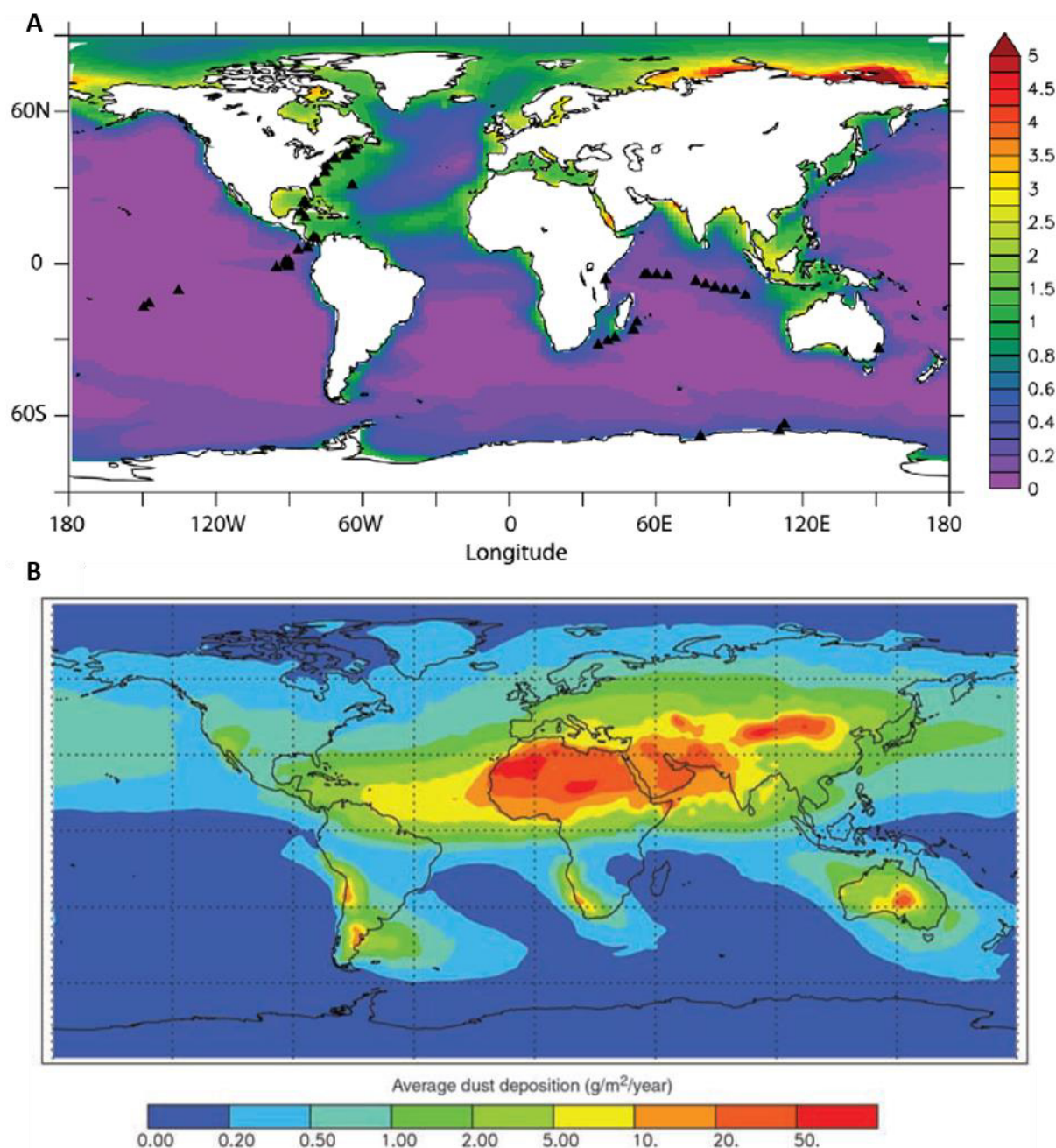


Figure 15. Cartes de la disponibilité et du dépôt du fer dans les océans. A) Concentration annuelle moyenne de fer entre 0 et 100 mètres modélisée avec NEMO-PISCES. Les triangles noirs représentent les sites d'échantillonnage du Global Ocean Sampling. D'après Toulza et al. 2012. B) Flux de dépôts de fer éoliens dans les océans, basés sur 3 modélisations. D'après Jickells et al. 2005.

### L'acclimatation du phytoplancton au manque de fer

Le phytoplancton eucaryote a mis en place de nombreux mécanismes pour faire face aux variations de fer disponibles. Les espèces côtières ou habituées à des milieux relativement riches en fer sont moins tolérantes à la limitation, et leur physiologie sera fortement impactée par le manque : diminution de la croissance et de l'activité photosynthétique, formation d'agrégat, etc. En revanche, les espèces pélagiques peuvent tolérer de plus fortes variations de disponibilité, et sont moins affectées

lorsque la concentration en fer est très faible (Morrissey and Bowler 2012; John A. Raven 2013; Bottebol et al. 2017).

Pour s'adapter ou s'acclimater à la limitation en fer, plusieurs stratégies existent. La première consiste à optimiser la prise du fer dans l'environnement. Comme évoqué précédemment au sujet de l'azote, la taille est une première adaptation, car une taille réduite facilite la diffusion des éléments. Des cellules adaptées à de très faibles concentrations de fer sont généralement plus petites que leurs comparses de la même espèce ou du même genre cultivées dans un milieu riche (John A. Raven 2013; Bottebol et al. 2017). Il est également possible de déployer des chélateurs de fer dans l'environnement pour capter les atomes biodisponibles plus efficacement. Cette stratégie est très courante chez les procaryotes et les plantes vasculaires, capables de déployer des sidérophores afin de capter le fer, mais on ne connaît aucune microalgue eucaryote capable d'en produire. La diatomée *P. tricornutum* serait capable de capter et d'utiliser les sidérophores bactériens (Kazamia et al. 2018). L'acide domoïque pourrait jouer un rôle équivalent, facilitant l'entrée du fer dans les cellules de *Pseudo-nitzschia spp.* (John A. Raven 2013). Produire plus de transporteurs dans des situations de carences en fer permet aussi d'optimiser la captation (Figure 16A). Chez la diatomée pélagique *P. tricornutum*, une carence en fer amène à la surexpression de gènes codant pour des réductases ferriques, permettant de réduire le  $Fe^{3+}$  en  $Fe^{2+}$ , ainsi que des transporteurs de métaux de la famille ZIP. Chez la diatomée côtière *T. pseudonana*, des perméases de fer vont être exprimées en plus des réductases ferriques (Morrissey and Bowler 2012). Chez *Ostreococcus tauri*, la stratégie diffère : le fer précédemment capté dans l'environnement peut être stocké et réparti à l'intérieur de la cellule grâce à la ferritine (Lelandais et al. 2016). Cette enzyme est répandue chez de nombreuses espèces de microalgues, mais sa capacité à stocker le fer sur le long terme est questionnée, car son expression dans les données métatranscriptomiques environnementales ne corrèle pas avec la concentration de fer (Caputi et al. 2019).

Parmi les gènes impliqués dans la captation et le transport du fer, les "iron starvation-induced proteins" sont spécifiquement régulés par le manque de fer (Figure 16B). Identifiées chez *Phaeodactylum tricornutum* (Allen et al. 2008; Morrissey et al. 2015), l'expression de ISIP1, ISIP2A et ISIP3 augmente fortement dans des conditions pauvres en fer. ISIP1 a longtemps été envisagé comme un transporteur permettant l'endocytose de fer lié à un sidérophore, mais des travaux récents sur *P. tricornutum* indiquent qu'il s'agirait plutôt d'un transporteur intracellulaire, amenant le fer capté là où il sera réduit (Behnke and LaRoche 2020; Kazamia et al. 2022). ISIP2A, ou la phytotransferrine, est un transporteur carbonate-dépendant qui concentre et transporte le fer non réduit  $Fe^{3+}$  (Behnke and LaRoche 2020; Kazamia et al. 2022; McQuaid et al. 2018). Finalement, ISIP3 pourrait être une protéine de stockage du fer, mais son rôle exact n'est pas encore élucidé (Behnke and LaRoche 2020). Chez les diatomées, l'abondance et l'expression des gènes codant pour ces protéines sont négativement corrélés avec la concentration en fer dans l'environnement, et ils sont surexprimés dans des eaux pauvres en fer (Caputi et al. 2019).

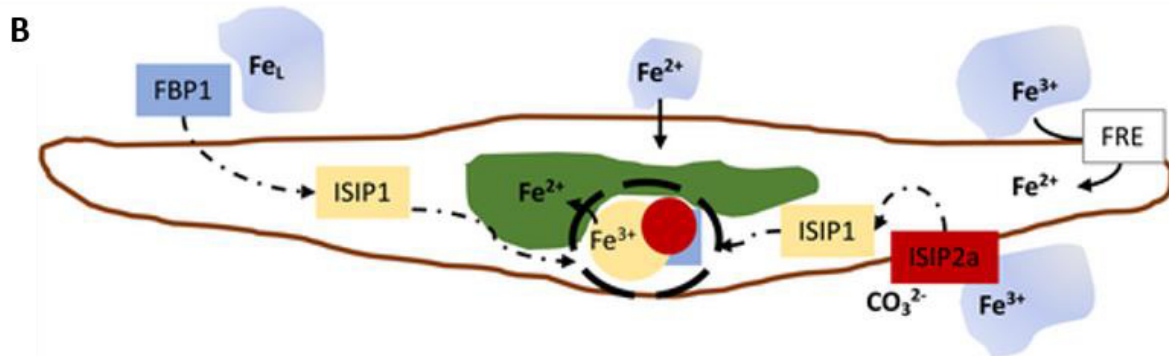
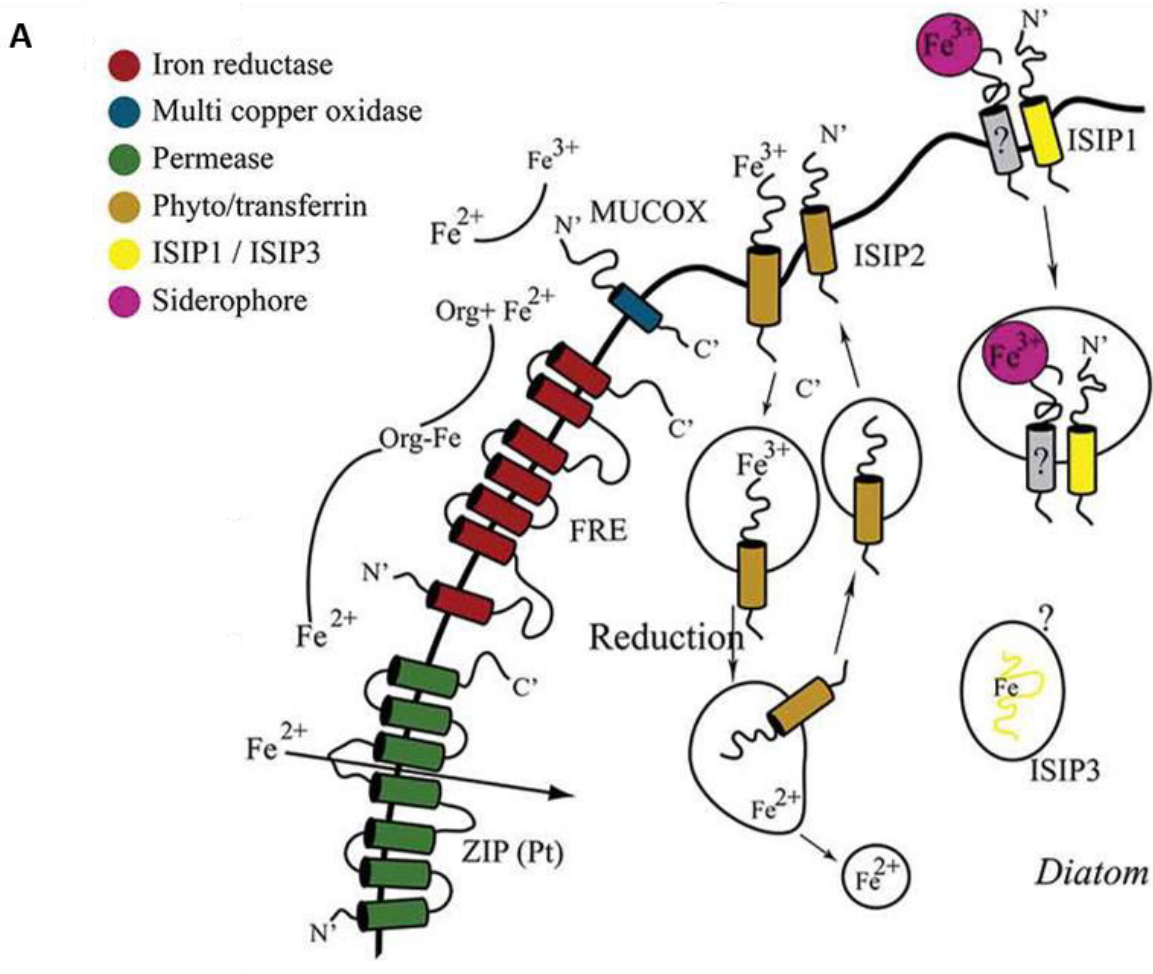


Figure 16. Transport du fer chez les diatomées. A) Aperçu des transporteurs de fer à haute affinité chez les diatomées. Org: ligand organique; ISIP: protéine induite par la privation de fer; FRE: réductase de fer; ZIP: protéine régulée par le zinc et le fer; MUCOX: oxydase multicuivre. Le nom phyto-transferrine désigne l'ISIP2A. Modifié d'après Behnke and LaRoche 2020. B) Fonctionnement des ISIP2A et ISIP1 chez *P. tricornutum*. ISIP2A, ou la phyto-transferrine, fait entrer le fer non réduit Fe<sup>3+</sup>. ISIP1 transporte le fer dans la cellule à l'endroit où il sera réduit en Fe<sup>2+</sup>. La réductase de fer (FRE) réduit le Fe<sup>3+</sup> en Fe<sup>2+</sup> pour permettre sa diffusion dans la cellule. D'après Kazamia et al. 2022.

La stratégie d'acclimatation au manque de fer la plus répandue chez le phytoplancton eucaryote marin est la modulation des besoins. Les cellules sont capables de diminuer le recours aux protéines contenant du fer en limitant certains processus cellulaires. *P. tricornutum* par exemple réduit la photosynthèse et l'assimilation du nitrate quand le

fer est insuffisant (Morrissey and Bowler 2012). Également, une stratégie très courante est de remplacer les protéines contenant du fer par des homologues non-ferreux. Ce remplacement peut être une adaptation génomique, avec des organismes vivants dans des zones faibles en fer qui ont un plus grand nombre de copies de gènes codant pour des protéines ne nécessitant pas de fer (Caputi et al. 2019). Il peut aussi s'agir d'une acclimatation transcriptomique : l'expression des gènes codant pour ces protéines et leurs homologues non-ferreux est modulée en fonction la disponibilité du fer. Par exemple, la ferrédoxine, une protéine permettant le transport d'électron lors de la photosynthèse peut être remplacée par la flavodoxine, une protéine remplissant la même fonction de manière moins efficace mais sans nécessiter de fer. Il a été observé par des analyses de métatranscriptomiques et métagénomiques *in situ* que l'expression des gènes codant pour les flavodoxines chez les diatomées est significativement plus importante dans les eaux pauvres en fer, alors que le nombre de copies génomique ne varie pas (Carradec et al. 2018). Le cytochrome c6 peut être remplacé par la plastocyanine, qui contient un atome de cuivre au lieu d'un atome de fer, chez la diatomée *T. oceanica*, et les superoxydes dismutases à fer peuvent être remplacées par celles à cuivre, à zinc, au manganèse ou à nickel (Morrissey and Bowler 2012). La fructose-biphosphate aldolase (FBA), une enzyme impliquée dans la glycolyse, la néoglucogénèse et le cycle de Calvin, se présente sous une forme ferreuse, la FBA II, et une forme non-ferreuse, la FBA I. Ces enzymes font également l'objet d'une régulation transcriptomique : elles présentent un nombre de copies constant au travers des métagénomés *Tara Oceans*, mais la FBA II est bien plus exprimée dans les eaux riches en fer, et la FBA I dans les eaux pauvres (Caputi et al. 2019).

## **La colonne d'eau, un environnement hétérogène**

Le phytoplancton est distribué dans l'ensemble de la colonne d'eau, dans la zone photique. Aussi appelée zone épipélagique, elle va de la surface jusqu'à la profondeur où il n'y a plus assez de lumière pour permettre la photosynthèse. Cette profondeur varie en fonction de la turbidité et de la latitude. Dans cette zone photique, la distribution du phytoplancton n'est pas homogène. En effet, les conditions environnementales varient avec la profondeur, structurant verticalement la communauté planctonique (Figure 17). La quantité de lumière diminue progressivement avec la profondeur alors que les nutriments (nitrate, ammonium, phosphate) sont apportés par la remontée des courants profonds. Leurs concentrations augmentent donc avec la profondeur. Le phytoplancton va particulièrement proliférer à la profondeur où la co-concentration de ces deux éléments sera optimale : la profondeur de chlorophylle maximum (DCM). Généralement située entre 50 et 120 m de profondeur, la DCM est définie par la concentration en chlorophylle a, qui y est maximale par rapport au reste de la colonne d'eau, indiquant l'abondance et la forte activité d'organismes photosynthétiques.

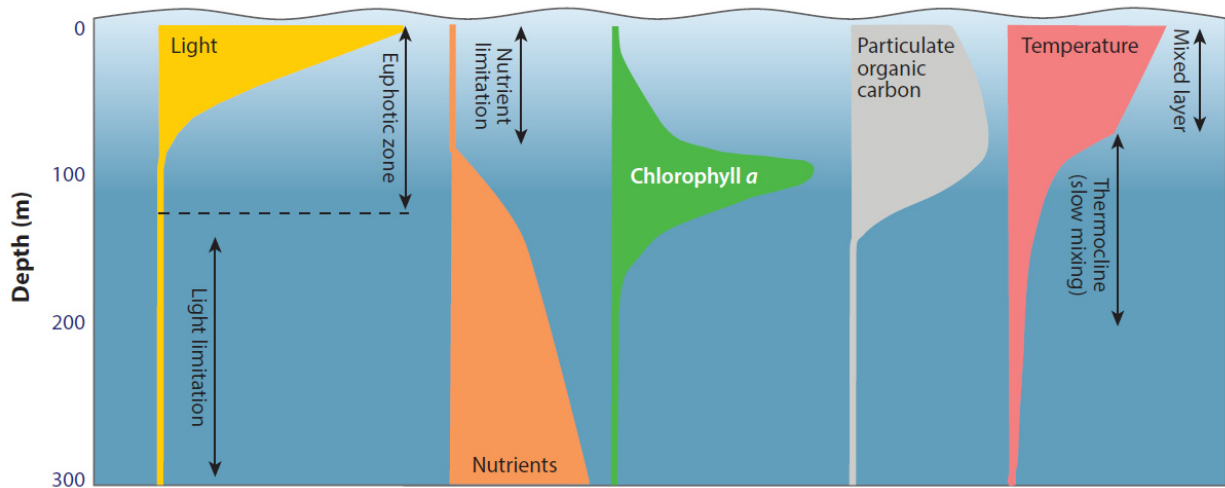


Figure 17. Distribution dans la colonne d'eau de la luminosité, la concentration en nutriments, en chlorophylle a, en carbone organique particulaire et la température. D'après Pierella Karlusich, Ibarbalz, and Bowler 2020.

Cette distribution inégale des nutriments et de la luminosité impacte la composition de la communauté phytoplanctonique le long de la colonne d'eau. En effet, il y a plus de variabilités dans la composition en espèces sur quelques dizaines de mètres de profondeur plutôt que sur des centaines de kilomètres carrés de surface (Estrada et al. 2016). Bien que très variable selon les régions, on retrouvera plutôt des coccolithophores et des dinoflagellés en surface, et plutôt des Pelagophyceae, des diatomées et d'autres genres de dinoflagellés à la DCM (Estrada et al. 2016; Giner et al. 2020).

Chaque environnement présente des contraintes qui lui sont propres. La surface représente un environnement stressant, combinant une forte irradiation et une raréfaction des nutriments inorganiques. Il serait tentant de l'envisager comme une sorte de désert océanique, mais ça serait oublier la grande adaptabilité dont sont capables les micro-organismes. Par exemple, les symbioses peuvent permettre de faire face à la pénurie de nutriments. Ainsi, les diatomées et les Haptophytes, associées respectivement avec des cyanobactéries diazotrophes et des diazotrophes non-photosynthétiques *Candidatus Atelocyanobacterium thalassa* (UCYN-A), peuvent compter sur la fixation d'azote de ces dernières pour compenser les faibles concentrations en nitrate (Thompson et al. 2012; Tuo, Chen, and Chen 2014; Turk-Kubo et al. 2021). Également, beaucoup d'organismes sont capables de migrer verticalement : les flagellés peuvent se déplacer, les diatomées et les coccolithophores peuvent moduler leurs flottaisons, ainsi il leur est possible de descendre récupérer des nutriments puis de remonter profiter d'une luminosité plus importante (Kemp and Villareal 2018). Cette migration verticale, qui n'est pas nécessairement synchronisée avec le cycle jour-nuit, pourrait impacter grandement la productivité primaire (Wirtz et al. 2022). Finalement, on pense que la profondeur peut impacter le mode trophique. Ainsi, les organismes mixotrophes (capables d'utiliser aussi bien la matière organique qu'inorganique comme source de nutriments), occuperaient plutôt la surface, laissant

la DCM plus riche en nutriments inorganiques aux autotrophes obligatoires (Poulton et al. 2017). À la DCM, ce sont plutôt les interactions biotiques qui vont générer de la compétition entre les organismes, avec une intensification de la prédation par broutage ou de la lyse virale par exemple.

## **Les impacts du changement climatique sur l'océan et le phytoplancton**

Le groupe d'experts intergouvernemental sur l'évolution du climat (Shulka et al. 2019) prédit que le changement climatique aura de nombreuses conséquences sur les écosystèmes marins, les plus importantes étant l'augmentation de la température de l'eau de surface, l'acidification et la désoxygénation (Figure 18A). Ces trois événements sont déjà en train d'impacter le climat, le phytoplancton, la totalité du réseau trophique et l'ensemble de la biogéochimie marine. Cependant, ces événements vont avoir des effets synergiques complexes, rendant parfois complexes de prédire leurs effets sur le phytoplancton à l'échelle globale.

Un premier constat est la biomasse phytoplanctonique a déjà diminué de 1 % par an depuis les années 1970 (Boyce, Lewis, and Worm 2010) et la surface des zones oligotrophes a augmenté de 0,8 à 4,3 % par an depuis 1998 (Polovina, Howell, and Abecassis 2008). La richesse en espèce phytoplanctonique devrait particulièrement diminuer dans les zones subtropicales et tempérées, et augmenter dans les zones polaires (Henson et al. 2021).

L'augmentation de la température des eaux de surface devrait amener à une intensification de la stratification, qui est un des principaux mécanismes expliquant la diminution de la biomasse phytoplanctonique. L'apport en nutriments (ammonium, nitrate, nitrite, phosphate) depuis les eaux profondes vers la surface devrait réduire de 9 à 14 % d'ici à 2100 (Shulka et al. 2019). Ce phénomène devrait être particulièrement intense dans les zones tempérées et tropicales, et devrait amener à un déclin de la production primaire nette de 7 à 16 % d'ici à 2100 (Figure 18B) (Polovina, Howell, and Abecassis 2008; Boyce, Lewis, and Worm 2010; Shulka et al. 2019). En réponse aux variations de la production primaire, l'exportation du carbone vers le fond de l'océan devrait diminuer de 9 à 16 % d'ici à 2100 (Shulka et al. 2019). Dans les zones polaires, le réchauffement des eaux de surface pourrait amener au contraire à l'intensification du mélange dans certaines zones, provoquant une augmentation locale de l'apport en nutriment et de la biomasse phytoplanctonique (Ardyna and Arrigo 2020).

La limitation en nutriments dans les eaux chaudes et tempérées favorisera les espèces de petites tailles, du fait de besoins nutritifs plus faibles et d'une diffusion plus efficace (Shuter 1978; Raven 1998; Litchman et al. 2007; Henson et al. 2021). L'augmentation de la température de l'eau et la diminution de la concentration en nutriments vont également grandement participer à la diminution de la taille du phytoplancton (Figure 18C). La taille des microfossiles de diatomées est corrélée avec la température (Van de

Waal and Litchman 2020). Expérimentalement, la chaleur induit une réduction en nutriments qui amène à un changement de la communauté planctonique en faveur d'espèces plus petites (Peter and Sommer 2013). Ainsi, il est prédit que la structure de la communauté phytoplanctonique sera impactée, avec une diminution du nombre de grandes espèces en faveur de petites espèces (Henson et al. 2021).

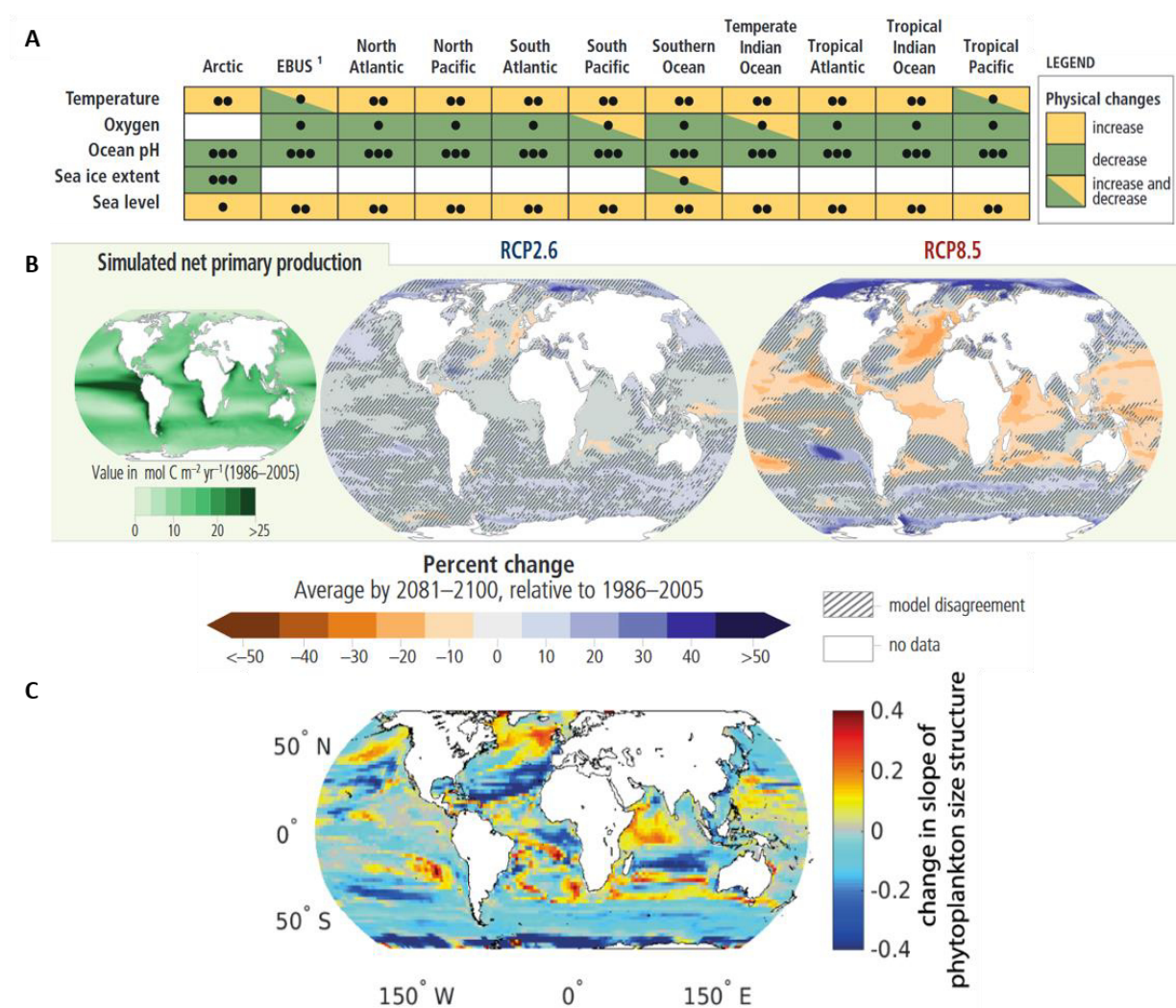


Figure 18. Impacts du changement climatique sur l'océan et le phytoplancton. A) Observation des variations de température, concentration d'oxygène, d'acidité, de surface de mer et de niveau de la mer dans les océans du monde faites ces 20 dernières années. La couleur de la case indique la direction du changement observé. Le nombre de points dans la case indique le niveau de confiance attribué à l'observation. Un point : confiance faible ; deux points : confiance moyenne ; trois points : confiance élevée. EBUS : abréviation pour Eastern Boundary Upwelling System, qui désigne les quatre principaux courants de remontée d'eau : le courant de Benguela, le courant des Canaries, le courant de Californie et le courant de Humboldt. D'après Shulka et al. 2019. B) Carte de la production primaire nette mondiale moyenne modélisée entre 1986 et 2005, et le taux de changement sous les scénarios de réchauffement RCP2.6 (le moins intense) et RCP8.5 (le plus intense) d'ici à 2100 par rapport à 1986-2005. Il est prédit que la production primaire nette diminue dans les zones en rouge et augmente dans les zones en bleus. Les zones en gris ou hachurées désignent les aires où la prédiction est impossible. D'après Shulka et al. 2019. C) Prédiction du changement de la structure de taille de la communauté phytoplanctonique par rapport à la structure de taille actuelle. Il est prédit que le petit phytoplancton sera plus abondant que le grand dans les zones en bleu, et le grand sera plus abondant dans les zones en jaune/rouge. D'après Henson et al. 2021

L'acidification pourrait également impacter le cycle de l'azote marin, mais d'une manière différente. La fixation de l'azote par les bactéries diazotrophes pourrait être augmentée de 29 % à cause de l'acidification de l'eau de mer, ce qui amènerait à une augmentation de la concentration d'ammonium disponible (Wannicke et al. 2018). À l'inverse, l'augmentation du pH a un effet inhibant sur l'activité de la nitrogénase, l'enzyme permettant la fixation de l'azote (Van de Waal and Litchman 2020). De plus, des expériences ont montré que l'acidification peut provoquer une réduction des taux de nitrification de 3 à 44 %, empêchant la génération de nitrite et de nitrate à partir d'ammonium, et amenant à la production d'un surplus d'oxyde nitreux, un gaz à effet de serre (Beman et al. 2011).

L'augmentation de la température, ainsi que l'augmentation des concentrations de CO<sub>2</sub> pourraient aussi stimuler la fixation d'azote. En effet, la cyanobactérie diazotrophe *Trichodesmium erythraeum* présentait une augmentation permanente et constitutive de son taux de fixation d'azote après avoir été cultivée dans un milieu enrichi en CO<sub>2</sub> (Walworth, Fu, et al. 2016; Walworth, Lee, et al. 2016). Simultanément, le stress azoté pourrait permettre une meilleure résistance au stress thermique. Les diatomées cultivées dans un milieu riche en azote développent une meilleure tolérance aux fortes températures que des lignées cultivées dans un milieu pauvre (David A. Hutchins and Capone 2022). Ainsi, l'augmentation de la fixation de l'azote engendrée par une température plus élevée pourrait compenser la diminution de l'apport de l'azote par le mélange des couches d'eau. Les effets de l'acidification et du réchauffement sur les différentes composantes du cycle de l'azote marin sont nombreux et ont des effets compensatoires. La tendance semble plutôt s'accorder sur un changement dans la distribution de l'azote en faveur de l'ammonium (grâce à l'augmentation de la fixation de l'azote avec l'acidification et le réchauffement) et en défaveur du nitrate et du nitrite (à cause de l'intensification de la stratification).

Comme tous les cycles biogéochimiques, celui du fer marin devrait aussi être modifié à cause du changement climatique. Il est prédit que les apports de fer augmenteront, particulièrement dans les eaux tropicales (D. A. Hutchins and Boyd 2016). La fonte des glaciers et de la glace de mer provoquée par l'augmentation de la température devrait intensifier les apports de fer d'origine glaciaire ou hydrothermale dans les régions côtières. L'intensification de l'aridification des sols terrestres pourrait amener des apports de fers éoliens supplémentaires. Les méga feux qui sont de plus en plus courants pourraient également enrichir l'océan en fer pyrogénique, fortement biodisponible. On estime que ce sont les grands incendies de 2019-2020 en Australie qui ont fertilisé l'océan Australe, provoquant une efflorescence phytoplanctonique sans précédent dans une zone habituellement peu productive (Tang et al. 2021). Cependant, les conséquences de ces modifications sont difficiles à prévoir. En effet, l'acidité et la température influencent la spéciation chimique du fer, réduisant le temps de demi-vie du fer oxydé le plus assimilable Fe<sup>2+</sup> ce qui pourrait atténuer les effets de l'enrichissement des apports en fer précédemment évoqués (D. A. Hutchins and Boyd 2016). Également, il a été montré expérimentalement que l'augmentation de la

température de l'eau atténue les effets délétères de la limitation en fer. Même sans apport supplémentaire, la diatomée *Pseudo-nitzschia* présente une meilleure croissance dans une eau plus chaude (Jabre et al. 2021). Ainsi, l'impact du changement climatique sur le cycle du fer, bien que certain, est imprévisible.

Les métabolismes du fer et de l'azote étant très intriqués, les zones oligotrophes devraient être particulièrement impactées par les modifications évoquées jusqu'ici. La stratification et le réchauffement devraient modifier le réservoir d'azote inorganique assimilable, avec une diminution du nitrate et une augmentation de l'ammonium. Cela pourrait avoir des conséquences sur les besoins en fer du phytoplancton, avec une diminution des besoins en nitrate réductase, l'enzyme ferreuse impliquée dans l'assimilation du nitrate, mais une augmentation des besoins en nitrogénase, l'enzyme ferreuse impliquée dans la fixation de l'azote (D. A. Hutchins and Boyd 2016). L'augmentation des apports de fer dans l'océan pourrait lever une des plus importantes limitations à la croissance du phytoplancton, provoquant une augmentation de l'assimilation des nutriments, de la fixation de l'azote et de la production primaire. Cela pourrait amener à terme à une consommation rapide des nutriments disponibles dans l'eau (Jickells et al. 2005). Finalement, si on peut prévoir des tendances régionales, il est difficile de prévoir l'impact global, car l'ensemble des effets vont s'accumuler ou se compenser.

### **3 - Comprendre les picoeucaryotes photosynthétiques pour comprendre le phytoplancton**

#### **Le picophytoplancton, un rôle écologique amené à prendre de l'importance**

Le picophytoplancton regroupe l'ensemble des organismes photosynthétiques unicellulaires dont la taille est inférieure à 3  $\mu\text{m}$  (Fuller et al. 2006; Vaulot et al. 2008). Ne comprenant pas les étapes de dispersion d'espèces de macro-algues benthiques, on retrouve dans cette catégorie des cyanobactéries et des eucaryotes photosynthétiques de petite taille vivant sous forme libre, c'est-à-dire non-symbiotique et non-parasitaire. En 2008, 71 espèces de picoeucaryotes photosynthétiques (PPE) marins étaient répertoriées (Figure 19) (Vaulot et al. 2008). Parmi ces espèces, on retrouve une trentaine d'Ochrophytes, dont une vingtaine de diatomées issues de genres tels que *Skeletonema*, *Chaetoceros* ou *Thalassiosira*, et quatre Pelagophyceae : *Aureococcus anophagefferens*, *Aureoumbra lagunensis*, *Pelagomonas calceolata* et *Pelagococcus subviridis*. On y compte également une vingtaine de Chlorophytes, dont les Mamiellophyceae *Micromonas*, *Bathycoccus* et *Ostreococcus*. Une dizaine de Prymnesiophytes appartenant entre autres aux genres *Chrysochromulina* et *Phaeocystis* ainsi que le Cryptophytes *Hillea marina* font aussi partie de cette liste. Ce court résumé de diversité ne couvre que les espèces décrites il y a 15 ans, et la génomique a permis de décrire de nombreuses autres espèces depuis.

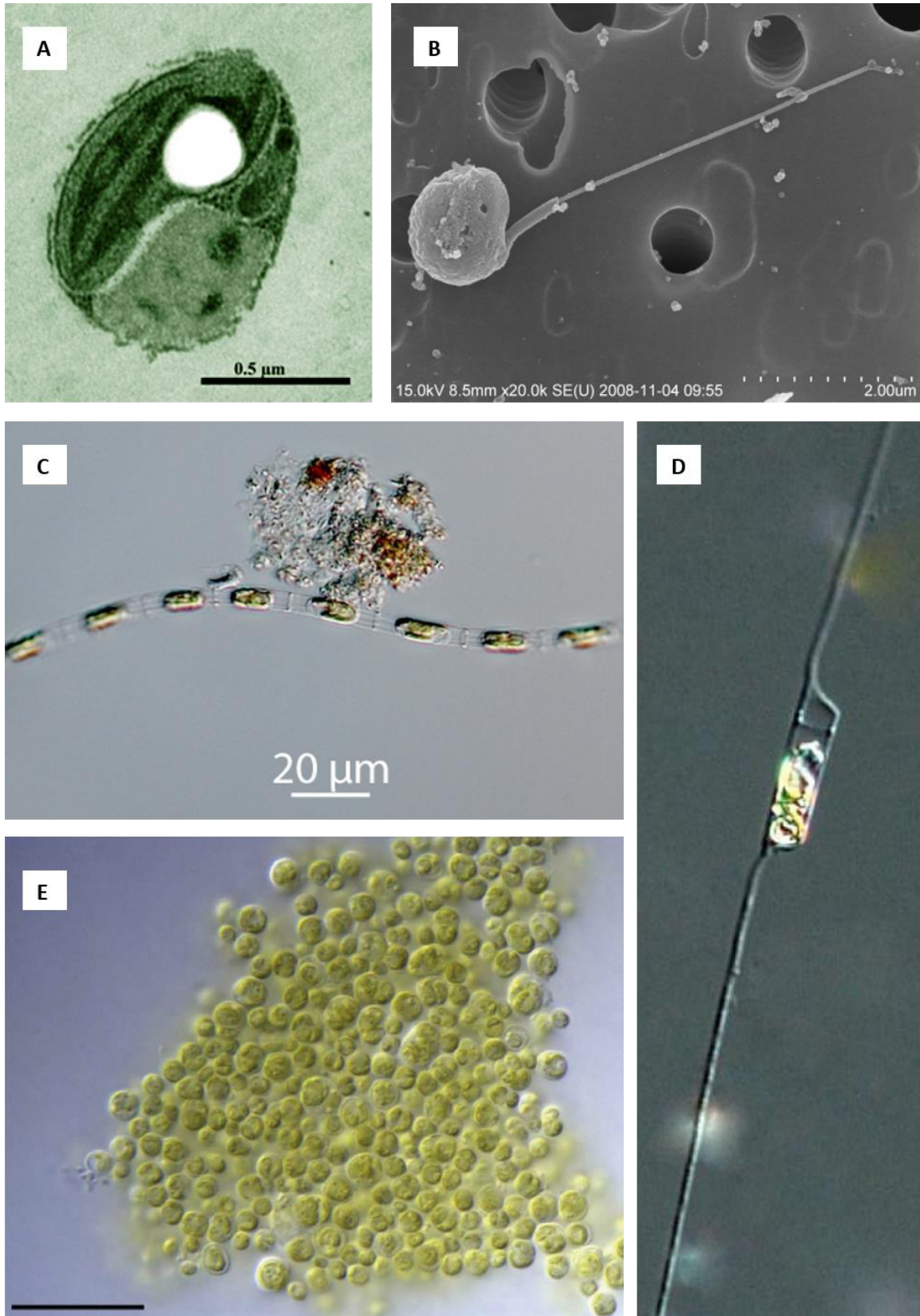


Figure 19. Quelques exemples de picoeucaryotes photosynthétiques (PPE). A) *Ostreococcus tauri* RCC1115 par H. Moreau, Laboratoire Arago (<http://phycocosm.jgi.doe.gov>). B) *Micromonas pusilla* RCC835 par E. Foulon, Station biologie de Roscoff (<http://roscoff-culture-collection.org>). C) Chaîne de *Skeletonema marinoi* (<http://nordicmicroalgae.org>). D) *Chaetoceros minimus* (<http://nordicmicroalgae.org>). E) Colonie de *Prasinoderma coloniale* d'après Jouenne et al. 2011. Échelle : 20 μm.

Les PPE ne sont pas les organismes les plus abondants, mais ils ont un rôle essentiel dans la production primaire et les flux de carbone. Dans l'upwelling chilien, où le nanophytoplancton (3-20  $\mu\text{m}$ ) est le plus abondant, le picophytoplancton représentait tout de même 55 à 92 % du total de chlorophylle a, et plus de 60 % du carbone fixé par la photosynthèse, dominant ainsi la production primaire (Rii et al. 2016). On estime que les PPE, bien que moins abondants que les cyanobactéries *Prochlorococcus* et *Synechococcus*, contribuent au moins de manière équivalente à la biomasse et à la production primaire (Worden and Not 2008; Rii et al. 2016). Les PPE sont donc particulièrement abondants et forment un compartiment essentiel du phytoplancton marin en termes de productivité.

La petite taille des PPE est un avantage majeur dans l'océan et explique en grande partie le succès écologique de ces organismes dans les eaux oligotrophes (Raven 1998). Une taille de cellule réduite permet d'avoir :

- Une efficacité photosynthétique accrue (théorisé sous le nom de "package effect").
- Une diffusion membranaire facilitée, se traduisant par une meilleure efficacité à capter les nutriments faiblement concentrés.
- Un besoin réduit de ribulose 1,5-bisphosphate carboxylase/oxygénase (ou Rubisco), une enzyme complexe à synthétiser.

Ainsi, les PPE devraient présenter un meilleur taux de croissance dans les milieux oligotrophes, et être plus abondants dans les eaux pauvres en nutriments ou en lumière, que les plus grandes cellules qui seront moins performantes sur les plans nutritifs et énergétiques (Raven 1998).

Cette diminution de la taille a amené chez les espèces de PPE décrites une importante réduction de la taille de leur génome. Ce phénomène est en général associé avec le passage vers un mode de vie parasitaire ou symbiotique. Pourtant, les PPE demeurent autonomes. Cette réduction de génome se traduit par une diminution du nombre de gènes impliqués dans les voies de biosynthèses des acides aminés, la réparation et la recombinaison de l'ADN, et la biogénèse des ARN messager et des ribosomes (Derilus et al. 2020). L'hypothèse a été faite que ces pertes de fonctions, notamment la perte des gènes de la voie de biosynthèses des métabolites solubles tel que les acides aminés, ont lieu car les organismes sont capables de mutualiser leurs métabolites (Morris, Lenski, and Zinser 2012; Derilus et al. 2023). Bien que ce partage des métabolites secondaires et des nutriments reste à prouver, il pourrait participer à expliquer l'importance des PPE dans l'océan. De plus, si la réduction du génome a été observée chez les PPE pour lesquels nous avons le génome complet, il est fort possible qu'un contre-exemple existe.

L'acclimatation transcriptomique pourrait être un des leviers majeurs du succès écologique des PPE. L'acclimatation se définit comme un changement physiologique temporaire en réponse à un changement environnemental. Par exemple, l'analyse des données de métatranscriptomique de *Bathycoccus* montre une forte modulation de

l'expression des gènes en fonction des conditions environnementales, suggérant que la large distribution de cet organisme est rendue possible par l'acclimatation de cette espèce (Vannier et al. 2016). Par son caractère temporaire, elle se différencie de l'adaptation, qui désigne l'ensemble des changements génétiques qui s'accumulent au cours des générations.

Alors que les PPE jouent un rôle écologique essentiel et que leurs biomasses devraient augmenter dans les prochaines décennies à cause de l'agrandissement des zones oligotrophes, leur étude est encore assez limitée. En effet, la plupart des études sur la physiologie et la réponse transcriptomique aux changements d'environnement portant sur les PPE se concentrent principalement sur les *Mamiellophyceae* et les diatomées. Si ces espèces ont l'avantage d'être abondantes et bien connues, elles ne reflètent pas toute la diversité des PPE. Pour comprendre comment les PPE vont s'acclimater aux changements environnementaux engendrés par le réchauffement climatique, il faut un organisme modèle qui puisse être étudié aussi bien en laboratoire que dans l'environnement, et représentatif des zones pélagiques oligotrophes.

## **Les Pelagophyceae, des picoeucaryotes photosynthétiques peu étudiés**

Caractéristique, abondance et distribution des Pelagophyceae

Les Pelagophyceae ou Pélagophytes sont des picoeucaryotes photosynthétiques marins appartenant aux Ochrophytes (Straménopiles). Les organismes de cette classe sont caractérisés par une thèque perforée en matière organique dont l'épaisseur varie entre 12 à 80 nm selon les genres. Cette thèque présente des micropores (8 à 14 nm) et des macropores (18 à 25 nm) et peut être formée par 1 à 5 couches, en général 4 ou plus chez les Pelagomonadales (sauf chez le genre *Pelagomonas* qui n'en présente qu'une) et 2 chez les Sarcinochrysidales (Wetherbee et al. 2021; 2023). Flagellés ou coccoïdes, sous forme libre ou formant des colonies et des mucilages, la majorité des espèces sont côtières ou sédimentaires, à part 2 espèces de Pelagomonadales, *Pelagococcus subviridis* et *Pelagomonas calceolata*, qui sont les seules espèces pélagiques (Wetherbee et al. 2021; 2023).

Que ce soit en métabarcoding ou en métatranscriptomique, de nombreuses études montrent que ces organismes sont parmi les PPE les plus abondants, principalement dans les eaux chaudes et oligotrophes (Dupont et al. 2015; Keeling and Campo 2017). Cependant, leur richesse spécifique est faible avec un indice de Shannon inférieur à 2 (Pierella Karlusich, Ibarbalz, and Bowler 2020). Leur abondance est corrélée positivement à la température et négativement avec la concentration en fer et chlorophylle a, ce qui peut s'expliquer par le fait que les Pelagophyceae sont plus abondants dans les zones où il y a moins de biomasse phytoplanctonique (Pierella Karlusich, Ibarbalz, and Bowler 2020). On les retrouve en plus grande abondance à la DCM qu'à la surface (Choi et al. 2020), mais ils peuvent dominer la communauté phytoplanctonique à toutes les profondeurs (Dupont et al. 2015; Choi et al. 2020).

Les études de métatranscriptomiques semblent indiquer que les Pélagophytes ont une forte capacité d'acclimatation par modulation de l'expression des gènes. L'acclimatation transcriptomique des Pélagophytes a particulièrement été étudiée dans la réponse *in situ* au manque de fer (Caputi et al. 2019; Carradec et al. 2018). L'expression des gènes codant pour les ISIP est très fortement modulée selon la concentration en fer, alors que le nombre de copies de ces gènes dans le génome ne varie pas. C'est également vrai pour les couples de protéines ferrédoxine/flavodoxine et FBA I/FBA II: le nombre de copies reste stable, mais la version ferreuse (ferrédoxine ou FBA II) est plus exprimée dans les stations riches en fer et l'homologue non-ferreux (flavodoxine ou FBA I) dans celles pauvres en fer.

### Diversité des Pelagophyceae

Les Pélagophytes sont répartis dans 2 ordres : les Pelagomonadales et les Sarcinochrysidales (Wetherbee et al. 2023). La plupart des espèces décrites ne le sont que sur base morphologique ou génétique avec le séquençage des gènes marqueurs codant pour l'ADNr 18S, pour les photosystèmes *psaA*, *psaB*, *psbA*, *psbC* et pour la sous-unités *rbcl* de la Rubisco. Nous allons brièvement revenir sur chaque espèce de Pélagophytes décrites, afin de dresser un portrait exhaustif de cette classe qui est assez peu étudiée (Figure 20).

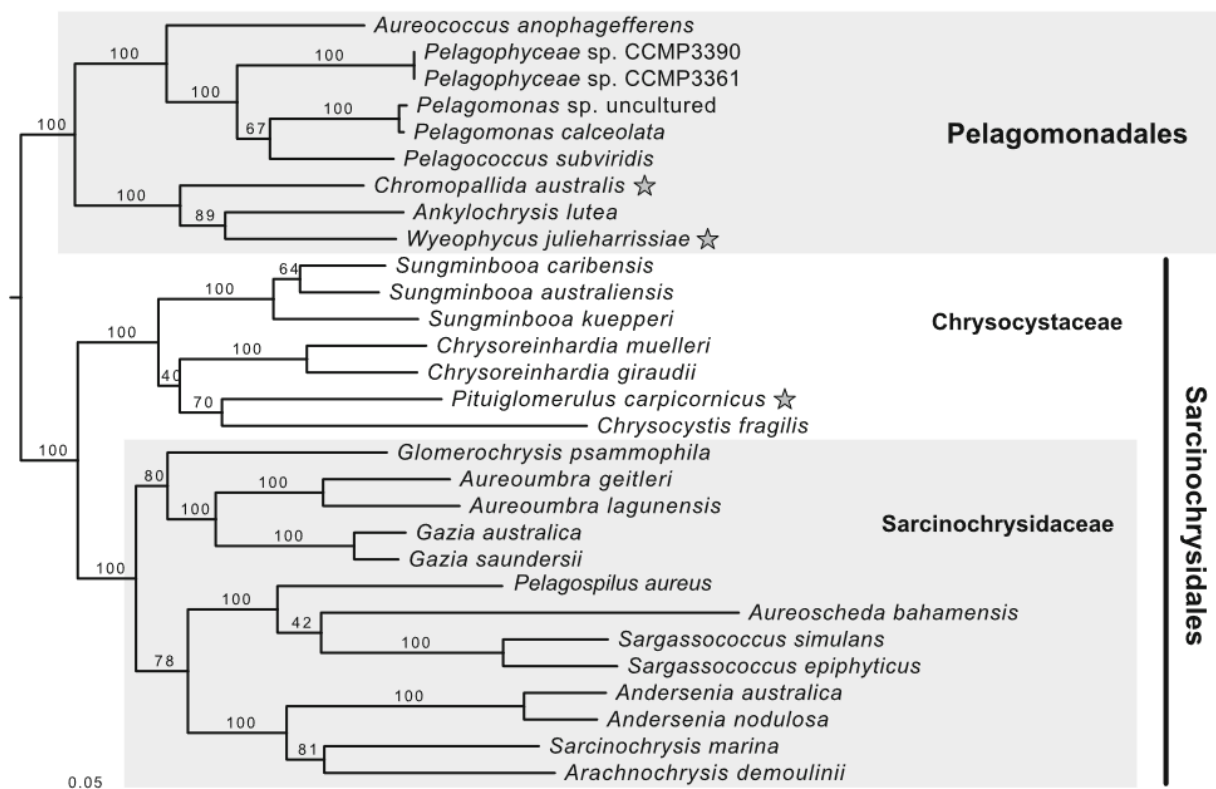


Figure 20. Arbre phylogénétique des Pelagophyceae construit à partir de l'alignement de l'ADNr 18S et des gènes marqueurs chloroplastiques *psaA*, *psaB*, *psbA*, *psbC*. Les étoiles indiquent les dernières espèces décrites. D'après Wetherbee et al. 2023.

Dans l'ordre des Sarcinochrysidales, on retrouve 2 familles : les Chrysocystaceae et les Sarcinochrysidaceae. La première famille, les Chrysocystaceae, regroupe 4 genres. Aucun génome complet n'a été séquencé pour les espèces de cette famille, mais toutes ont bénéficié d'un séquençage des gènes marqueurs. Le genre *Chrysoreinhardia* contient deux espèces : *C. giraudii* et *C. muelleri*. La première a été décrite par Derbès et Sollier en 1856, puis de nouveau en 2000, avant d'avoir un séquençage de ces gènes marqueurs en 2018 (Derbès and Solier 1856; Hoffmann et al. 2000; Han et al. 2018). Elle forme des mucilages, mais peut également exister sous forme libre flagellée (Han et al. 2018). *C. muelleri* en revanche n'a été décrite que récemment, sur base morphologique et séquençage des gènes marqueurs (Han et al. 2018). La coloniale *Pituiglomerulus capricornicus* aussi n'a été décrite que très récemment (Wetherbee et al. 2023). Le genre *Sungminbooa* comprend trois espèces : *S. kuepperi*, *S. caribensis* et *S. australiensis*. Aucune ne semble capable de former des colonies ou des mucilages (Han et al. 2018). Décrit dans le Pacifique en 1995, *Chrysocystis fragilis* est le seul représentant de son genre. Il possède 2 formes : une coloniale filamenteuse et une libre motile (Lobban et al. 1995). Il est très peu étudié, mais une efflorescence de cette espèce a été reportée sur la Grande Barrière de Corail en 2004 (Schaffelke et al. 2004).

La famille des Sarcinochrysidaceae contient 10 genres, dont la plupart ne sont décrits que sur base morphologique et par séquençage des gènes marqueurs codant pour l'ADNr 18S, pour les photosystèmes psaA, psaB, psbA, psbC et pour la sous-unités rbCL de la Rubisco. Décrites en 2015, les espèces filamenteuses *Andersenia australica* et *A. nodulosa* se développent dans le sable (Wetherbee et al. 2015; Han et al. 2018). Le genre *Gazia* contient 2 espèces, toutes deux sédimentaires : *G. australica* et *G. saundersii*. Cette dernière est coloniale, contrairement à la première (Wetherbee et al. 2021). Seul représentant de son genre, *Glomerochrysis psammophila* est aussi une espèce coloniale sédimentaire (Wetherbee et al. 2021). *Arachnochrysis demoulinii* et *Pelagospilus aureus* ont toutes les deux été décrites sur base morphologique et des gènes marqueurs récemment (Han et al. 2018). *Ankylochrysis lutea*, tout d'abord décrite en 1970 sous le nom de *Ankylonoton lutem*, a bénéficié d'une description taxonomique plus poussée en 1994, puis d'un séquençage de ces gènes marqueurs en 2018 (van der Veer 1970; Honda and Inouye 1995; Han et al. 2018). Également capable de former des mucilages, *Aureoscheda bahamensis* a été décrite en 2014 puis a bénéficié d'un séquençage de ces gènes marqueurs (Wynne et al. 2014; Han et al. 2018). Décrite en 1930, seul le séquençage des gènes marqueurs a permis de faire rentrer *Sarcinochrysis marina* dans la classe des *Pelagophyceae* (Geitler 1930; Han et al. 2018). Les polysaccharides produits par cette espèce ont été étudiés pour des applications biomédicales en tant qu'anti-cancéreux (Sun et al. 2013). Une seule espèce de la famille des Sarcinochrysidaceae a bénéficié d'un séquençage de son transcriptome en 2022. Le genre *Sargassococcus* est représenté par deux espèces, toutes capables de former des petites colonies : *S. epiphyucus* ou *S. simulans*. Cette dernière serait une espèce symbionte du corail et son transcriptome séquencé a révélé une grande quantité d'antennes collectrices de lumière, d'enzyme oxydoréductrice et d'enzymes impliqués dans l'assimilation et l'utilisation du carbone organique. Ces transcrits permettent

respectivement une acclimatation à une faible luminosité, au stress oxydatif et une capacité à la mixotrophie (Han et al. 2018, 20; Varasteh et al. 2022).

L'ordre et la famille des Pelagomonadales sont un peu plus étudiés sur le plan génomique et physiologique que les Sarcinochrysidales. On compte 7 genres décrits dans cette famille. Seules représentantes de leurs genres respectifs, la toxique *Aureococcus anophagefferens* et la pélagique *Pelagomonas calceolata* sont les deux espèces les mieux décrites. Nous nous attarderons sur leurs caractéristiques plus bas. Également espèces uniques au sein de leurs genres, la flagellée *Chromopallida australis* et la sédimentaire *Wyeophycus julieharrissiae* n'ont été décrites que très récemment, de ce fait seules leurs descriptions taxonomiques et les séquences de leurs gènes marqueurs sont disponibles (Wetherbee et al. 2023).

Le genre *Aureoumbra* contient deux espèces : *A. geitleri*, et *A. lagunensis*. La première est la moins étudiée, car elle n'a été décrite que récemment sur base morphologique et barcoding (Han et al. 2018; Wetherbee et al. 2023). Décrite en 1997, *A. lagunensis* est capable de former des efflorescences dans les estuaires (DeYoe et al. 1997). Adaptée à des faibles concentrations d'azote, cette espèce peut utiliser l'ammonium, le nitrite, l'urée et le glutamate comme source d'azote, mais pas le nitrate (DeYoe and Suttle 1994; Muhlstein and Villareal 2007; Agostoni and Erdner 2011). Elle semble avoir une préférence pour l'ammonium et l'urée, qui représente 90 % de l'azote assimilé par cette espèce pendant ces efflorescences. De plus, des enrichissements en ammonium peuvent stimuler sa croissance (Kang, Koch, and Gobler 2015). Elle peut s'acclimater à une faible luminosité (Gobler and Sunda 2012). *A. lagunensis* peut s'acclimater à de très fortes variations de salinité, pouvant croître dans de l'eau comprise entre 10 et 90 PSU, avec une croissance optimale dans un milieu allant de 20 à 70 PSU (Buskey, Wysor, and Hyatt 1998). À titre de comparaison, la salinité de l'eau de mer est de 34-35 PSU. Également, elle est capable de se protéger contre le broutage par les copépodes grâce à un mucilage en polysaccharide qui recouvre sa surface extracellulaire (Liu and Buskey 2000). Son génome chloroplastique a été séquencé en 2010, son génome mitochondrial en 2021, et une étude sur sa réponse transcriptomique aux variations d'azote, phosphore et quantité de lumière a été menée en 2021 (Ong et al. 2010; Sibbald, Lawton, and Archibald 2021; Kang et al. 2021). La diminution de l'expression des ferrédoxines et flavodoxines en réponse à la privation en azote semble indiquer que *A. lagunensis* serait capable diminuer ces besoins photosynthétiques pour s'acclimater aux limitations de nutriments (Kang et al. 2021).

Seul représentant de son genre, *Chrysophaeum taylorii* a été identifié la première fois en Floride en 1941. Elle est distribuée dans les eaux tropicales de l'Amérique du Nord, Pacifique et Atlantique, dans les Caraïbes, et plus récemment dans la mer Méditerranée, elle est donc considérée comme une espèce envahissante (Aktan and Topaloğlu 2011). *C. taylorii* est capable de former des mucilages sur les surfaces rocheuses peu profondes, ce qui lui permettrait de se propager plus efficacement et de faire face aux manques de nutriments (Caronni et al. 2015; 2016; 2017). Cette espèce est étudiée pour

ces métabolites secondaires d'intérêt biomédical. En 1986 et 1989, il a été montré que *C. taylorii* pouvait synthétiser deux composés cytotoxiques pouvant être utilisés pour traiter des cellules cancéreuses *in vitro* : l'hormothamnione et la 6-desmethoxyhormothamnione (William H Gerwick et al. 1986; W. H. Gerwick 1989). Également, *C. taylorii* peut synthétiser des chrysosphaentine, un nouveau type d'antibiotique prometteur qui bloque la division cellulaire en inhibant l'activité de la protéine FtsZ, homologue bactérien de la tubuline (Keffer et al. 2013; Davison and Bewley 2019). Le génome de cette espèce a été séquencé très récemment, afin d'étudier plus amplement son potentiel métabolique (Davison et al. 2023).

*Pelagococcus subviridis* est l'une des deux seules espèces pélagiques de la classe des Pelagophyceae. Décrite en 1977, mieux détaillée par microscopie électronique en 1987, cette espèce a une composition en pigment légèrement différente des autres Pélagophytes : un mélange de 19'-butnoyloxyfucoxanthine (le pigment marqueur des Pélagophytes), de fucoxanthine, de diadinoxanthine et de diatoxanthine (Lewin et al. 1977; Veski and Jeffrey 1987). Son transcriptome séquencé en 2021 pour étudier sa réponse au manque de lumière, d'azote et de phosphate, montre que cette espèce était peu impactée par les privations en nutriments. En revanche, en faible luminosité, un gène codant pour une lysophospholipase était surexprimé, suggérant un clivage des lysophospholipides pour restructurer la membrane cellulaire (Kang et al. 2021).

Évidemment, beaucoup d'espèces de Pelagophyceae restent à découvrir, particulièrement dans les zones polaires (Figure 21). Ainsi, dans une étude de 2020, jusqu'à 9 groupes de Pélagophytes non-décrits sont identifiés sur base des séquences codantes pour l'ADNr 16S chloroplastiques (Choi et al. 2020). Parmi eux, PEC-I serait une espèce du genre *Pelagomonas* occupant les eaux subarctiques du Pacifique et PEC-II un *Pelagococcus* propre aux eaux du Pacifique Sud. PEC-III pourrait appartenir à un nouveau genre, particulièrement abondant dans les eaux de surface très stratifiées et pauvres en nutriments. Proches de *A. anophagefferens*, PEC-IV et PEC-V pourraient aussi former un nouveau genre, distribué dans les eaux du Pacifique Nord-Est et de l'Antarctique respectivement. Appartenant à l'ordre des *Sarcinochrysidales*, les groupes PEC-VI, VII, VIII et IX formeraient de nouveaux genres associés respectivement à la fonte des neiges arctiques, de la glace de mer en Alaska, aux eaux du Pacifique nord et à la mer du Nord. Également en 2021, une étude visant à identifier la diversité taxonomique du phytoplancton eucaryote dans la péninsule Fildes à la pointe sud du Chili identifie 3 variants de séquence d'amplicon (ASV) de l'ADNr 16S chloroplastique appartenant à la classe des Pélagophytes. Le premier correspond à 100 % à la référence de *P. calceolata* et les deux autres sont très abondants, mais ne correspondent à aucune référence (Trefault et al. 2021). En 2022, la réponse du Pélagophyte arctique CCMP2097 aux variations de salinité montre que cette espèce associée à la glace de mer présente une forte acclimatation transcriptomique. Face à de fortes salinités, des complexes de canaux  $K^+$  sont surexprimés, et lorsque la salinité diminue, l'expression des gènes est modulée en faveur d'antiporteurs  $Na^+/H^+$  et de symporteurs  $Na^+/Pi$ . Les auteurs s'intéressent également aux gènes impliqués dans l'adaptation au froid : une protéine

de liaison à la glace, des protéines antigel et une acyl-estérase (Freyria et al. 2022). Les Pelagophyceae polaires restent donc peu décrits, alors qu'ils semblent nombreux, diversifiés et que leurs rôles écologiques pourraient être importants.

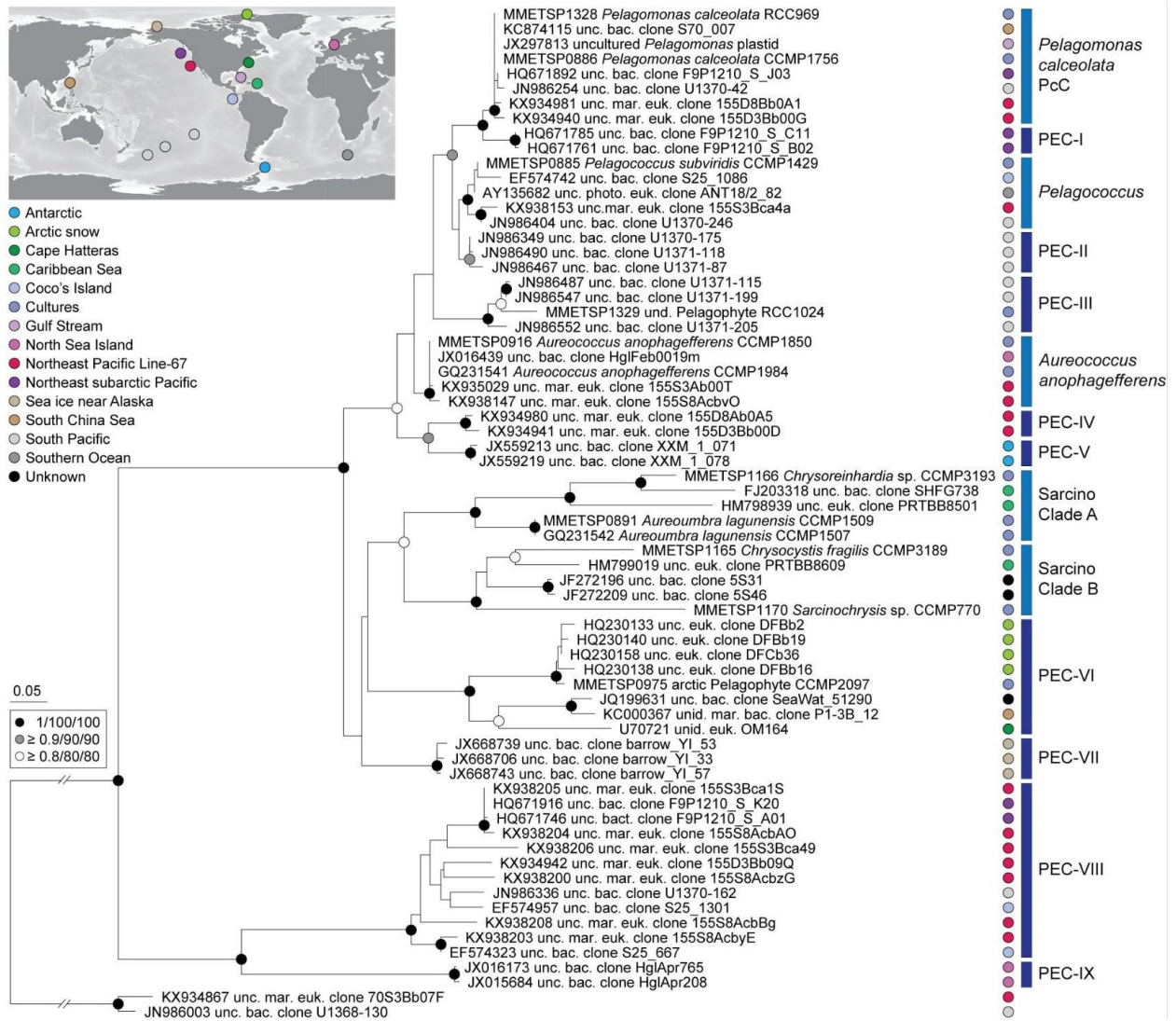


Figure 21. Arbre phylogénétique des Pelagophyceae construit à partir de l'alignement des ADN<sub>r</sub> chloroplastique 16S de 56 clones environnementaux et 9 espèces cultivées, montrant l'existence d'un ensemble de Pélagophytes non-décrits. Cinq clades connus ont été élucidés et neuf clades inconnus ont été identifiés : les clades PEC-I à PEC-IX. Deux séquences de provenant de Dictyophytes servent d'extra-groupe. L'arbre a été construit avec la méthode du maximum de vraisemblance (Maximum Likelihood, RAxML) sous le modèle d'évolution GTR (General Time Reversible) corrigé, avec 1 000 répliquats bootstrap. Des reconstructions additionnelles ont été faite avec les méthode PhyML et MrBayes. La couleur indique la région océanique d'où provient l'échantillon (voir la carte en haut à gauche). D'après Choi et al. 2020.

### *Aureococcus anophagefferens* et ces blooms toxiques en milieu oligotrophe

Le genre le plus étudié sur le plan physiologique et métabolique est le genre *Aureococcus*. La seule espèce, *Aureococcus anophagefferens* (Figure 22), est capable de former des blooms toxiques qui ont des effets destructeurs sur les zostères (*Zostera marina*) et les bivalves filtrants, notamment les coquilles Saint-Jacques (*Argopecten*

*irradians*) et les moules (*Mytilus edulis*) (Bricelj and MacQuarrie 2007). De ce fait, elle est le centre d'intérêt d'un certain nombre d'études ayant pour objectif de comprendre les mécanismes déclenchant et régissant ces blooms. *A. anophagefferens* prolifère dans des estuaires en Amérique du Nord, en Afrique du Sud et en Chine, où la salinité est élevée (> 25 ‰) et le temps de résidence long (C. J. Gobler, Lonsdale, and Boyer 2005).

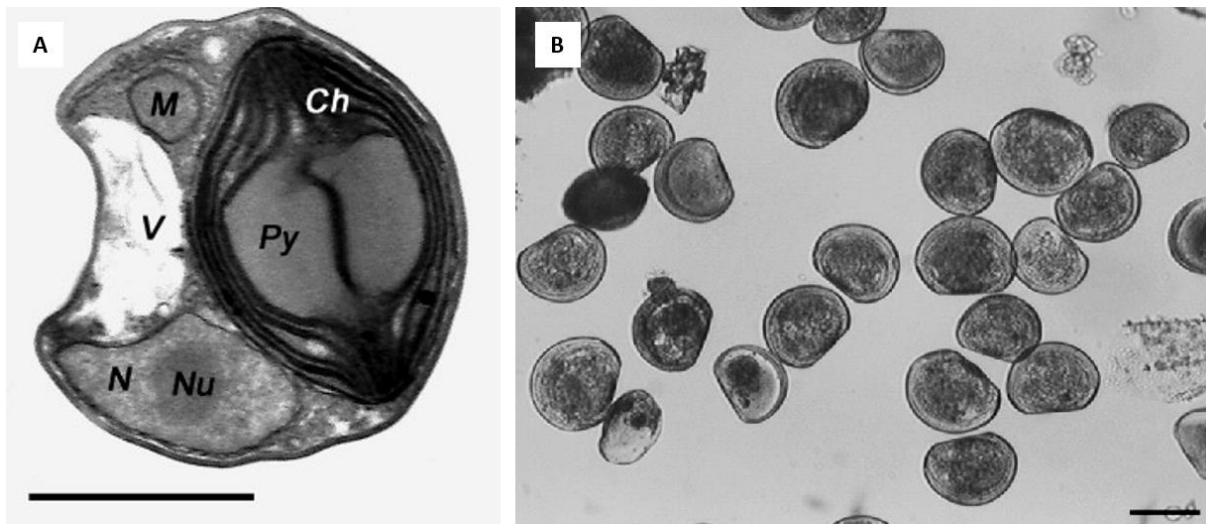


Figure 22. *Aureococcus anophagefferens* A) Structure intracellulaire en microscopie électronique en transmission. N : noyau ; Nu : nucléole ; Ch : chloroplaste ; Py : pyrénioïde ; M : mitochondrie ; V : vacuole. Échelle : 1 µm. D'après Gobler and Sunda 2012. B) Culture de *A. anophagefferens*. Échelle : 100 µm. D'après Bricelj and MacQuarrie 2007.

Alors que les blooms sont généralement déclenchés par un apport en nutriments, *A. anophagefferens* prolifère malgré une faible luminosité et une faible concentration en carbone et azote inorganique (Christopher J. Gobler, Renaghan, and Buck 2002; Christopher J. Gobler et al. 2004). Une première adaptation à la faible concentration d'azote inorganique disponible est sa petite taille. Avec un diamètre moyen de 2 µm, la cellule est plus efficace pour fonctionner malgré une faible luminosité et quantité de nutriments (Gobler et al. 2011).

Également, *A. anophagefferens* peut utiliser une grande diversité de source d'azote organique pour compenser le manque de nitrate et d'ammonium. Lors d'une efflorescence, on estime que 70 % de l'azote assimilé par *A. anophagefferens* est d'origine organique. Ainsi, la part d'azote assimilée est composée à 58-64 % d'urée, 18-26 % d'ammonium, 11-16 % d'acide aminés et 5-8 % de nitrate (Berg et al. 1997). Elle peut donc déployer un large panel d'enzymes pour avoir recours à cette DON, avec un rendement parfois plus efficace que celles de certaines bactéries hétérotrophes. L'uréase, l'acétamidase, la peptidase et la désaminase permettant d'hydrolyser respectivement l'urée, l'acétamide, les peptides et les purines (Gry Mine Berg, Repeta, and Laroche 2002). Des études de transcriptomique ont montré qu'en cas de carence en azote les gènes surexprimés codent pour des transporteurs d'ammonium, des acétamidases, des formamidases, des peptidases et des transporteurs de purines, confirmant l'utilisation de source d'azote organique en cas de raréfaction du nitrate et de l'ammonium (Wurch et al. 2011; Wurch, Gobler, and Dyhrman 2014). D'ailleurs,

l'urée semble être la source d'azote qui convient le mieux à *A. anophagefferens*, même en présence d'azote inorganique. En effet, elle semble avoir une meilleure efficacité pour assimiler l'urée, qui permet d'améliorer son taux de croissance et son efficacité photosynthétique par rapport au nitrate ou à l'ammonium (Ou et al. 2018).

Son génome complet, séquencé en 2011, permet d'avoir une vision plus précise des gènes impliqués dans cette forme de mixotrophie azotée (Gobler et al. 2011). Parmi ces 56 Mbp, *A. anophagefferens* présente 62 gènes codant pour des antennes collectrices de lumière (ou LHC "light harvesting complex"), ce qui représente jusqu'à 3 fois plus d'antennes que chez d'autres PPE. Ceci peut expliquer sa capacité à se développer sous une faible luminosité. Un très grand nombre de gènes permettant l'assimilation de la DON ont été identifiés : des transporteurs d'oligosaccharides, des enzymes de dégradation des polysaccharides, des carbohydrates, des acides aminés, des peptides ou des acides nucléiques. Également, *A. anophagefferens* serait capable de produire des composés antibactériens, anti-prédateurs ou anti-compétitions (Gobler et al. 2011). Elle présente alors le plus grand sélénoprotéome connu avec 59 sélénoprotéines. Les variations de sélénium dissous pendant l'efflorescence ainsi que des expériences d'enrichissement, montre que ces protéines auraient un rôle important pour protéger contre le stress oxydatif et pour catalyser plus efficacement la biosynthèse des protéines (Gobler et al. 2013).

### ***Pelagomonas calceolata*, un Pélagophyte d'intérêt**

Décrite en 1993, la description de *Pelagomonas calceolata* (Figure 23) a nécessité la création d'une nouvelle classe, les Pelagophyceae, ainsi que de nouveaux ordre, famille et genre : les Pelagomonadales, les Pelagomonadaceae et les *Pelagomonas*. Elle est à ce jour la seule espèce appartenant à ce genre. D'une taille de 2 à 3 µm de longueur et environ 1.5 µm de largeur, son nom fait référence à sa forme de chausson (Andersen et al. 1993). La description de l'espèce se fait à partir de la souche RCC100 (aussi nommée CCMP1214), échantillonnée en 1973 dans le gyre central du Pacifique nord, à 30°50'N, 136°50'W. Elle est uniflagellée, sans corps basal secondaire ni racine microtubule ou striée. Sa fine thèque en matière organique, caractéristique des Pelagophyceae, ne sera décrite plus précisément que 30 ans plus tard (Wetherbee et al. 2023). La cellule présente un seul chloroplaste, entouré de lamelle périphérique, et une seule mitochondrie dense avec des crêtes tubulaires. L'appareil de Golgi se situe sous le flagelle. Une organelle éjectant un corps cylindrique (probablement en réponse à un stress) et plusieurs vacuoles stockant des glucides ont été observées, sans que leurs fonctions ne soient établies. Lors de sa description en 1993, *P. calceolata* est décrit sur base morphologique et son ADNr 18S est séquencé. Son génome chloroplastique a été séquencé en 2012, se révélant très similaire à celui d'*Aureococcus anophagefferens* (105/106 protéines en commun) (Worden et al. 2012). Un premier transcriptome a été publié en 2014, puis un second en 2021, pour étudier la réponse à des limitations en azote, en phosphate et en luminosité (Keeling et al. 2014; Kang et al. 2021).

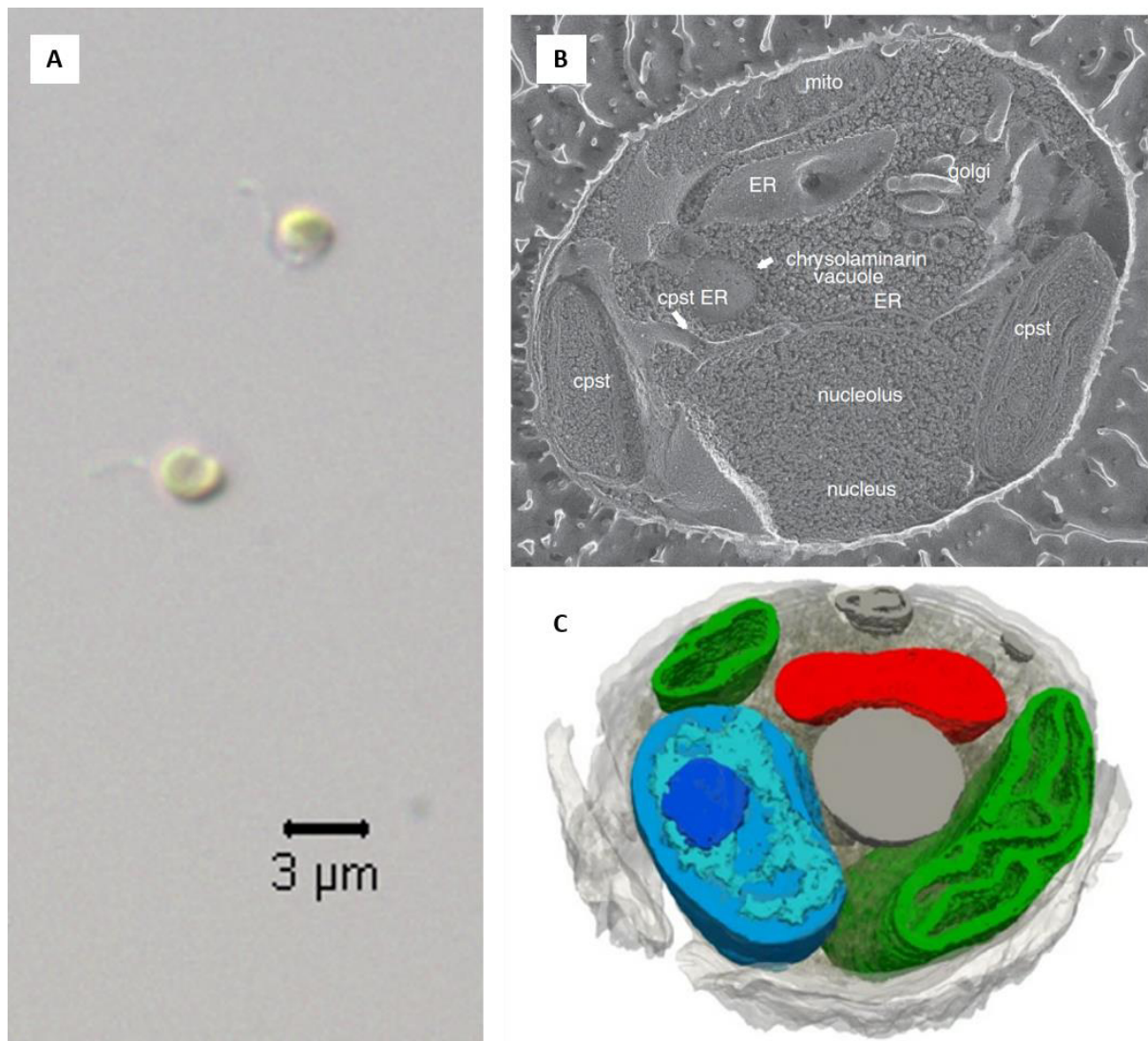


Figure 23. *Pelagomonas calceolata*. A) Cellules de *P. calceolata* RCC100 en microscopie optique par N. Guérin, réalisé à la Station Biologie de Roscoff. B) Structure intracellulaire de *P. calceolata* CCMP1756 en microscopie électronique à transmission. cpst : chloroplaste ; ER : réticulum endoplasmique ; golgi : appareil de golgi ; mito : mitochondrie. D'après Dupont et al. 2015. C) Modélisation des compartiments cellulaires de *P. calceolata* à partir d'imagerie en microscopie électronique FIB-SEM. En bleu le noyau, en rouge la mitochondrie, en vert les chloroplastes. D'après Uwizeye et al. 2021.

Plusieurs études ont estimé l'abondance de *P. calceolata* à partir de l'analyse métagénomique des gènes marqueurs chloroplastiques et de l'ARNr 18S. Dans les zones pélagiques, la richesse en espèces des Pelagophyceae est assez faible, avec un indice de diversité de Shannon compris entre 0 et 1.5 (Pierella Karlusich, Ibarbalz, and Bowler 2020). Puisque les seules espèces pélagiques dont les séquences 18SV9 sont connues sont *Pelagomonas calceolata* et *Pelagococcus subviridis*, ce sont probablement elles que l'on détecte par métabarcoding. En 2015, il était montré lors d'une étude de métagénomique *in situ* que la majorité des Pelagophyceae étaient similaires à la souche CCMP1756 de l'espèce *Pelagomonas calceolata* (Dupont et al. 2015). Une étude de métabarcoding du 16S chloroplastique estimait que dans le Pacifique, plus de la moitié des amplicons assignés à la classe des Pelagophyceae appartenaient à l'espèce

*P. calceolata* (Choi et al. 2020). Ainsi, *P. calceolata* semble être le Pélagophyte majoritaire dans les données environnementales, représentant sa classe dans les eaux oligotrophes.

*Pelagomonas calceolata*, contrairement à la plupart des autres Pélagophytes, est cosmopolite et abondant. On la retrouve dans la majorité des stations échantillonnées lors du Global Ocean Sampling en analysant l'abondance des séquences codantes pour l'ADNr 18S (Worden et al. 2012). Une étude plus récente de métabarcoding de l'ADNr 16S chloroplastique fait cette même conclusion (Choi et al. 2020), et les analyses de fluorescences par hybridation *in situ* estiment que *P. calceolata* contribue significativement à la biomasse phytoplanctonique (Worden et al. 2012). Particulièrement abondant à la DCM, *P. calceolata* est aussi présent dans les eaux de surface, où son profil d'expression laisse penser qu'il est essentiel à l'assimilation du nitrate et qu'il domine la production de matière organique (Dupont et al. 2015).

Malgré son abondance, le rôle écologique de *Pelagomonas calceolata* reste peu étudié. Quelques études nous laissent penser que l'abondance et l'ubiquité de cette espèce sont possibles grâce à sa forte capacité d'acclimatation. En effet, *P. calceolata* peut supporter d'importantes déplétions en nutriments et luminosité avant que sa physiologie et sa croissance ne soit impacté. En 2005, une expérience en laboratoire suivait la croissance et l'efficacité photosynthétique de plusieurs microalgues à des concentrations décroissantes de  $\text{PO}_4^{3-}$ , de  $\text{NH}_4^+$  et de fer (Timmermans et al. 2005). Seule une privation intense d'un de ces nutriments impacte le taux de croissance de *P. calceolata*. L'efficacité photosynthétique n'est affectée que très légèrement, et seulement aux concentrations les plus basses. L'effet de l'intensité lumineuse sur le taux de croissance a également été testé, et seuls des niveaux inférieurs à  $10 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  amenaient à une diminution des taux de croissance et de l'efficacité photosynthétique. Ainsi, *P. calceolata* pourrait présenter une forte résistance physiologique à la déplétion en nutriments et en privation de lumière.

La photoacclimatation de *P. calceolata* a été étudiée plus en détail. En 2009, Dimier et al. ont voulu comprendre l'impact des changements de cycle lumineux sur la physiologie de *P. calceolata* (Dimier et al. 2009 a; 2009 b). Trois cycles ont été testés : le premier simulant un cycle jour/nuit normal, le second simulant un mixage normal et le troisième simulant un mixage intense. Cette étude a montré que *P. calceolata* se divise de nuit au rythme de 0,87 division/jour. Le taux de croissance, la taille des cellules et l'efficacité photosynthétique ne sont pas affectés par le changement de rythme lumineux. Elle semble faire preuve d'une forte capacité d'acclimatation aux changements de rythme lumineux, avec une activation rapide du cycle xanthophylle et de l'extinction non-photochimique (non-photochemical quenching, NPC). L'intensité du mixage impacte le type d'acclimatation. Dans le cas d'un mixage normal, *P. calceolata* adopte une stratégie d'acclimatation de type  $\sigma$ , c'est-à-dire que la taille des antennes collectrices augmente. Dans le cas d'un mixage intense, la stratégie adoptée est de type  $n$ , c'est-à-dire qu'il y a augmentation du nombre d'unités

photosynthétiques. Cette stratégie a un fort coût nutritionnel (Dimier et al. 2009 a). Il a été montré que *P. calceolata* est adapté à des luminosités faibles et variables, grâce à une grande quantité et diversité de pigment caroténoïdes (Dimier et al. 2009 b).

Dans des eaux de surface du Pacifique Nord, une étude a comparé l'expression des gènes d'un transcriptome de *Pelagomonas* dans 3 stations de DCM riches en nitrate contre 5 stations (1 à la DCM, 4 en surface) pauvre en nitrate (Dupont et al. 2015). Les transcrits les plus abondants dans les premières codent pour des gènes de transports des nitrates et des nitrites, des carbamyl-phosphate synthétases (enzyme du cycle de l'urée) et des glutamate synthases, indiquant une assimilation importante d'azote inorganique. Également, *P. calceolata* exprime des protéines ribosomales, indiquant une croissance, et des dynéines, suggérant une forte motilité. En revanche, les fonctions les plus exprimées dans les stations pauvres en nitrate correspondent à des protéines de stress, des protéines chaperonnes, des histidine kinase, des régulateurs de transcription, et une protéine de détection du nitrate appelées NIT-sensing, décrite chez les bactéries. Avec 90 % des transporteurs de nitrate de l'ensemble du jeu de données provenant de *P. calceolata*, cette espèce pourrait avoir un rôle essentiel dans l'assimilation du nitrate dans les eaux oligotrophes (Dupont et al. 2015).

Quatre transcriptomes de *Pelagomonas* cultivés en condition standard ou de faible nitrate, phosphate, lumière ont été étudié en 2021 (Kang et al. 2021). Cette étude a été réalisée sans réplicats biologique. Lors d'une privation en phosphate, il semblerait que les gènes codant pour les pyrophosphatases soient surexprimés. Lors d'une privation en ammonium, ce sont des enzymes permettant la dégradation de molécules organiques azotées telles que les cathépsines, les uréases, les arginases et les aminotransférases qui semblent surexprimés, indiquant potentiellement une dégradation des protéines et des acides aminés. Dans les deux situations, cela semble indiquer que lors d'une privation en nutriments, *P. calceolata* pourrait être capable de dégrader ces molécules organiques pour y récupérer de l'azote ou du phosphore.

Plus récemment dans le cadre d'une thèse de doctorat, des expériences en mésocosme d'enrichissement de l'eau de mer ont permis de mieux caractériser la réponse transcriptomique de *P. calceolata* aux changements de concentration de nutriments. Ainsi, elle s'est montrée très réceptive aux enrichissements en nitrate et en fer menant à une augmentation du ratio N:Fe. Son abondance augmente, et un ensemble de fonctions est différentiellement exprimée : les transcrits codant pour des peroxyosomes et des enzymes d'assimilation de l'ammonium sont surexprimés, montrant une augmentation de la biosynthèse des acides aminés et gras, et de l'assimilation de l'azote. Ceux codant pour des facteurs de régulation de la transcription, des enzymes de biosynthèse de la spermidine, de dégradation des histidines ou des marqueurs de la privation en phosphate sont sous-exprimés, indiquant une levée de la privation en phosphore et une forte régulation de l'expression génique (Groussman 2022).

## 4 - Hypothèse générale et objectifs de la thèse

Dans cette thèse, nous faisons l'hypothèse que l'abondance de *P. calceolata* pourrait augmenter avec le réchauffement global des océans grâce à ses capacités d'acclimatation transcriptomiques aux environnements pauvres oligotrophes. Cette hypothèse se base sur les études existantes sur *P. calceolata* qui suggèrent que cette espèce peut faire preuve d'une forte capacité d'acclimatation transcriptomique, mécanisme probablement responsable du succès écologique des PPE.

C'est au travers de 3 chapitres que nous étudierons cette microalgue en couplant la métagénomique et métatranscriptomique environnementale avec des expérimentations au laboratoire pour tenter de répondre aux problématiques suivantes :

### 1) Quelle est l'abondance et la niche écologique de *P. calceolata* ?

Dans un premier temps, j'ai voulu comprendre quelle était la niche écologique de *P. calceolata* et quelles étaient les fonctions qui lui permettaient d'occuper ces environnements et d'être cosmopolite malgré les variations de conditions environnementales. Son génome est une ressource unique, que j'ai exploitée de plusieurs façons. D'abord, j'ai comparé son contenu en gène avec les génomes d'autres PPE, afin de découvrir son potentiel génétique et de constater ces ressemblances et différences avec d'autres espèces. Ensuite, j'ai aligné son génome sur les données métagénomiques de *Tara Oceans* afin d'étudier son abondance et sa distribution dans l'océan. Finalement, j'ai aligné le transcriptome de référence de *P. calceolata* sur les données métatranscriptomiques de *Tara Oceans* afin de comprendre quels étaient les gènes impliqués dans son succès écologique. Ce travail a donné lieu à un premier article paru en 2022 dans le journal *Communications Biology* (Chapitre 1).

### 2) Quels sont les mécanismes d'acclimatation de *P. calceolata* face aux conditions pauvres en nitrate ?

J'ai voulu comprendre comment *P. calceolata* s'acclimatait aux changements de disponibilités d'azote. En effet, une des conséquences du réchauffement climatique est la diminution de la quantité de nutriments disponible, et les études de Dupont et al. et de Kang et al. suggèrent que *P. calceolata* a une forte capacité d'acclimatation transcriptomique aux changements de disponibilités des nutriments (Dupont et al. 2015; Kang et al. 2021). J'ai comparé les gènes différentiellement exprimés dans les données environnementales *Tara Oceans* avec ceux différentiellement exprimés lors d'expériences de changements de quantité et de source d'azote. La comparaison entre les gènes différentiellement exprimés *in situ* et *in vitro* m'ont permis de mieux comprendre les mécanismes d'économie et de recyclage de l'azote mis en place par *P. calceolata*. Le couplage de ces deux méthodes m'a permis de lever le voile sur le rôle de la cyanate lyase dans la réponse au manque de nitrate. Cette étude sera présentée sous forme d'un second article en préparation (Chapitre 2).

- 3) Comment la profondeur et la quantité de lumière affecte l'expression des gènes de *P. calceolata* ?

Pour mon troisième chapitre, j'ai voulu mieux comprendre comment *P. calceolata* s'acclimatait au changement de profondeur. Bien que relativement plus abondante à la DCM, *P. calceolata* est également très présente en surface, indiquant qu'elle est capable de moduler l'expression de ces gènes pour occuper ces deux environnements qui sont pourtant extrêmement différents. J'ai encore une fois analysé les gènes différentiellement exprimés dans les données *Tara Oceans*, selon la profondeur cette fois-ci, afin de comprendre quelles étaient les physiologies adoptées par *P. calceolata* selon les contraintes inhérentes à ces deux environnements.

- 4) La microalgue *P. calceolata* est-elle un bon modèle pour l'étude des PPE pélagiques ?

Au regard des caractéristiques de *P. calceolata* et des résultats présentés dans cette thèse, nous estimons que faire de cette espèce un organisme modèle représentatif des PPE et des Pélagophytes serait possible. Dans mon dernier chapitre, je prendrai le temps de discuter les conclusions et limites de ce travail, et d'envisager l'usage de *P. calceolata* en tant qu'organisme modèle, d'un point de vue scientifique et technique.



# Chapitre 1 - Genomic adaptation of the picoeukaryote *Pelagomonas calceolata* to iron-poor oceans revealed by a chromosome-scale genome sequence

---

## Préambule au Chapitre 1

Dans ce premier chapitre, publié sous forme d'article en 2022 dans *Communications Biology*, le premier objectif est de caractériser le génome de *P. calceolata*. En effet, il s'agit du premier génome de référence complet pour cette espèce, et du deuxième génome de Pélagophytes avec celui de *Aureococcus anophagefferens* (Gobler et al. 2011). Les ressources génomiques étant rares pour les PPE, l'analyse grâce à la génomique comparative permet de comprendre les similitudes et différences entre les espèces.

Le deuxième objectif est de caractériser l'abondance et la niche écologique de *P. calceolata*. En alignant les lectures métagénomiques issues de l'expédition *Tara Oceans* sur le génome de référence, nous avons pu comprendre la distribution de *P. calceolata* à l'échelle globale, et modéliser sa distribution future.

Finalement, le troisième objectif est de comprendre comment *P. calceolata* peut s'acclimater aux différents milieux qu'elle occupe, particulièrement les milieux pauvres en fer et pauvres en nitrates. L'analyse du contenu en gènes a permis d'identifier les fonctions pouvant jouer un rôle dans la réponse au manque de fer et de nitrate. En alignant les lectures métatranscriptomiques issues de l'expédition *Tara Oceans* sur le génome de référence, nous avons pu constater comment l'expression de ces gènes variait avec la concentration en fer ou en nitrate. Nous avons ainsi pu acquérir une idée précise des mécanismes existant chez *P. calceolata* permettant l'acclimatation aux milieux oligotrophes.












Ce premier chapitre éclaire donc sur l'abondance et la niche écologique de *P. calceolata*, ainsi que sur les mécanismes moléculaires de l'acclimatation aux manques de fer et de nitrate exprimés dans son environnement.

Ce travail a été publié dans la revue *Communications Biology* en Septembre 2022 : <https://doi.org/10.1038/s42003-022-03939-z>

Dans cette partie, j'ai réalisé l'analyse de génomique comparative ainsi que les analyses de métagénomiques et de métatranscriptomiques environnementales.



## Genomic adaptation of the picoeukaryote *Pelagomonas calceolata* to iron-poor oceans revealed by a chromosome-scale genome sequence

Nina Guérin<sup>1,2</sup>, Marta Ciccarella <sup>1</sup>, Elisa Flamant<sup>1,2</sup>, Paul Frémont <sup>1,2</sup>, Sophie Mangenot<sup>1,2</sup>, Benjamin Istace<sup>1</sup>, Benjamin Noel <sup>1</sup>, Caroline Belser <sup>1</sup>, Laurie Bertrand<sup>1,2</sup>, Karine Labadie <sup>2,3</sup>, Corinne Cruaud<sup>2,3</sup>, Sarah Romac <sup>4</sup>, Charles Bachy <sup>4,5</sup>, Martin Gachenot<sup>5</sup>, Eric Pelletier<sup>1,2</sup>, Adriana Alberti<sup>1,2,6</sup>, Olivier Jaillon <sup>1,2</sup>, Patrick Wincker <sup>1,2</sup>, Jean-Marc Aury <sup>1</sup> & Quentin Carradec <sup>1,2</sup> ✉

The smallest phytoplankton species are key actors in oceans biogeochemical cycling and their abundance and distribution are affected with global environmental changes. Among them, algae of the Pelagophyceae class encompass coastal species causative of harmful algal blooms while others are cosmopolitan and abundant. The lack of genomic reference in this lineage is a main limitation to study its ecological importance. Here, we analysed *Pelagomonas calceolata* relative abundance, ecological niche and potential for the adaptation in all oceans using a complete chromosome-scale assembled genome sequence. Our results show that *P. calceolata* is one of the most abundant eukaryotic species in the oceans with a relative abundance favoured by high temperature, low-light and iron-poor conditions. Climate change projections based on its relative abundance suggest an extension of the *P. calceolata* habitat toward the poles at the end of this century. Finally, we observed a specific gene repertoire and expression level variations potentially explaining its ecological success in low-iron and low-nitrate environments. Collectively, these findings reveal the ecological importance of *P. calceolata* and lay the foundation for a global scale analysis of the adaptation and acclimation strategies of this small phytoplankton in a changing environment.

<sup>1</sup> Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France. <sup>2</sup> Research Federation for the Study of Global Ocean Systems Ecology and Evolution, R2022/Tara Oceans GO-SEE, 3 rue Michel-Ange, 75016 Paris, France. <sup>3</sup> Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, 2 Rue Gaston Crémieux, 91057 Evry, France. <sup>4</sup> Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR7144, Place Georges Teissier, 29680 Roscoff, France. <sup>5</sup> Sorbonne Université, CNRS, FR2424, Station Biologique de Roscoff, 29680 Roscoff, France. <sup>6</sup> Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France. ✉email: [qcarrade@genoscope.cns.fr](mailto:qcarrade@genoscope.cns.fr)

Marine phytoplankton accounts for more than 45% of photosynthetic primary production on Earth and play an essential role in supplying organic matter to marine food webs<sup>1</sup>. They are key global actors in CO<sub>2</sub> uptake and provide gaseous oxygen to the atmosphere. A global decline of phytoplankton biomass has been reported over the past century (1% of chlorophyll-a concentration per year) leading to a decrease of net primary production in many oceanic regions<sup>2</sup>. This decline is probably a consequence of global ocean warming which drives water column stratification, reducing the nutrient supply to surface waters. Temperature-driven reductions in phytoplankton productivity in tropical and temperate regions are likely to have cascading effects on higher trophic levels and ecosystem functioning<sup>3</sup>.

Photosynthetic picoeukaryotes (PPEs), defined by a cell diameter <3 µm, belong to different phyla, including Chlorophyta, Cryptophyta, Haptophyta, and Stramenopiles<sup>4</sup>. Present in all oceans, PPEs are the dominant primary producers in warm and oligotrophic regions<sup>5</sup>. Ocean warming and expansion of oligotrophic regions in the next decades may extend the ecological niche of PPEs, and a global shift from large photosynthetic organisms toward smaller primary producers is expected<sup>3,6</sup>. For example, sea ice melting in the Canadian Arctic Basin has been associated with an increase in the abundance of PPEs such as *Micromonas* at the expense of larger algae<sup>7</sup>. In the laboratory, this alga has the capacity to change its optimum temperature for growth in only a few hundred generations, suggesting that it will be less affected by global warming than many larger organisms<sup>8</sup>. In addition, the larger cell surface-to-volume ratio of PPEs compared to larger phytoplankton cells is advantageous for resource acquisition and growth in nutrient-limited environments<sup>9,10</sup>.

Iron is one key compound required for the activity of the respiratory chain, photosynthesis and nitrogen fixation<sup>10</sup>. Because bioavailable iron is extremely low in more than one-third of the surface ocean, small phytoplankton has developed several strategies to optimize iron uptake and reduce iron needs<sup>11</sup>. In diatoms, reductive and non-reductive iron uptake mechanisms involve many proteins, including phytoferritins, transmembrane ferric reductases, iron permeases, and siderophore-binding proteins<sup>12</sup>. The iron needs can be modulated by the variation of gene expression levels between iron-required proteins and their iron-free equivalent. These protein switches include electron transfer (flavodoxin/ferredoxin), gluconeogenesis (fructose-bisphosphate aldolase type I or type II) and superoxide dismutases (Mn/Fe-SOD, Cu/Zn-SOD or Ni-SOD)<sup>13–15</sup>.

PPE growth is also limited by nitrogen (N) availability in large portions of the global ocean<sup>16</sup>. Ammonium (NH<sub>4</sub><sup>+</sup>), nitrate (NO<sub>3</sub><sup>-</sup>) and nitrite (NO<sub>2</sub><sup>-</sup>) are the primary source of inorganic N for PPEs, however, several studies have shown that dissolved organic N, like urea, can be metabolized in N-limited environments<sup>17</sup>. For example, several membrane-localized urea transporters in the diatom *Phaeodactylum tricorutum* are maximally expressed in nitrogen-limited conditions<sup>18</sup> and the harmful algal blooms of the pelagophyceae *Aureococcus anophagefferens* may be fueled by urea<sup>19</sup>.

Despite their large taxonomic distribution, most molecular studies on the ecological role of PPEs and their adaptation to the environment are restricted to a few species. PPEs are suspected of possessing highly developed acclimation/adaptation capacities, but the underlying molecular mechanisms remain poorly characterized due to the lack of reference genomic data.

Among PPEs, *Pelagomonas calceolata* was the first described member of the Pelagophyceae class<sup>20</sup>. It has since been identified in many oceanic regions using its 18 S rRNA sequence and chloroplast genome<sup>21–23</sup>. Several studies have demonstrated the capacity of *P. calceolata* to adapt to different environmental conditions. In the laboratory, *P. calceolata* has been shown to exhibit a high degree of

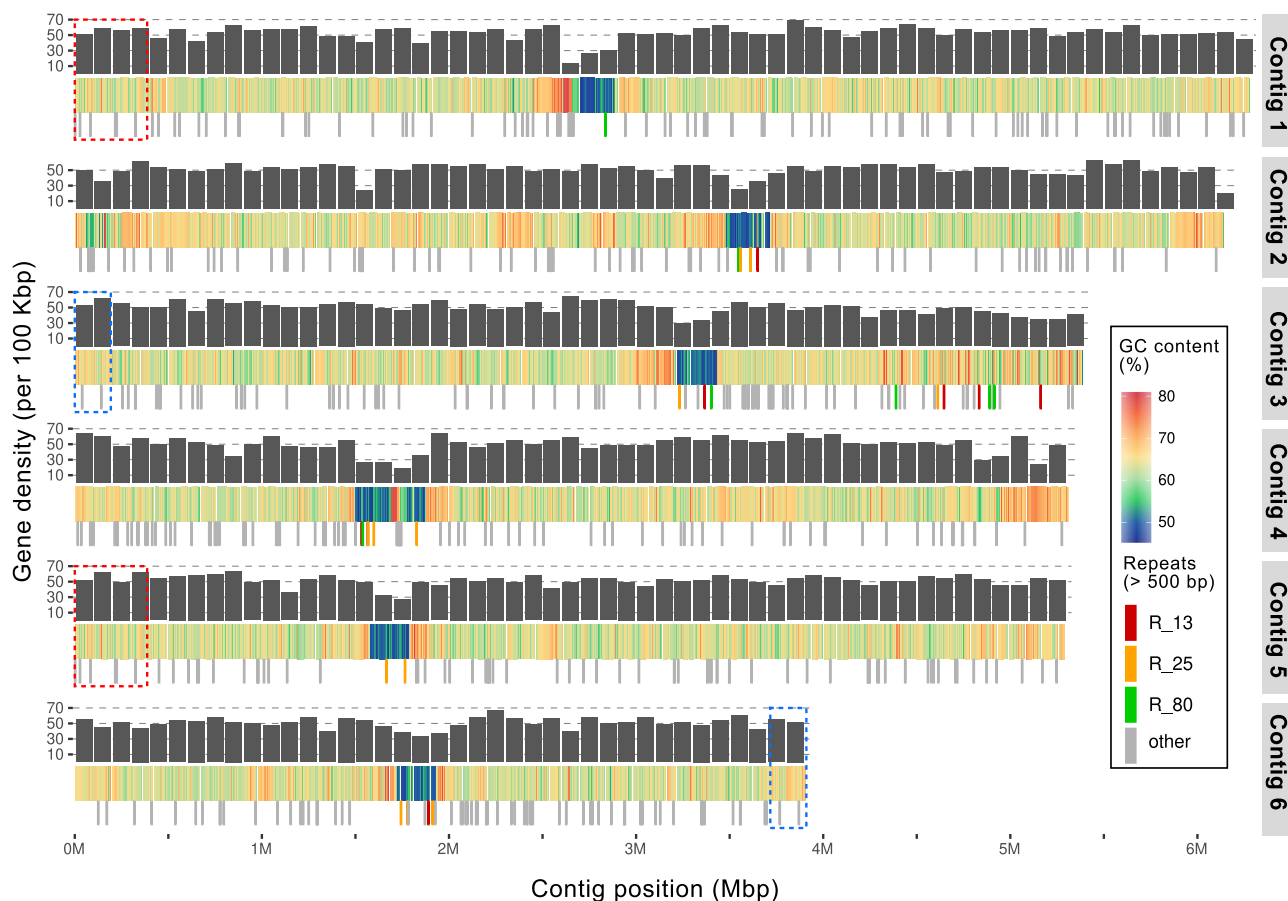
acclimation to light fluctuations with rapid activation of the photoprotective xanthophyll cycle and non-photochemical quenching<sup>24</sup>. In the Marquesas archipelago, *P. calceolata* is one of the most responsive species to iron fertilization with upregulation of genes involved in photosynthesis, amino acid synthesis and nitrogen assimilation<sup>13</sup>. A global-scale analysis of pelagophyte genes revealed that they are adapted to low-iron conditions<sup>14</sup>. In the subtropical Pacific, *P. calceolata* expresses stress genes in surface samples and genes involved in nitrogen assimilation are overexpressed in the deep-chlorophyll maximum<sup>25</sup>. A laboratory study suggests that *P. calceolata* also has the ability to increase the transcription levels of organic-nitrogenous compound cleavage enzymes (cathepsin, urease, arginase) under low nitrogen concentration<sup>26</sup>. Thus, gene expression appears to be controlled according to the nitrogen source and quantity. Taken together, this apparent adaptive plasticity may explain the presence of *P. calceolata* in many different oceanic environments, however, an exhaustive analysis of the genetic capacity of this species and the in situ characterization of its ecological niche is lacking.

Here we sequenced, assembled and annotated the *Pelagomonas calceolata* genome, with a combination of long- and short reads. We examined its genomic structure and gene content relative to other unicellular phytoplankton. We used this genome to detect *P. calceolata* in environmental datasets of *Tara* expeditions across all oceans to characterize its ecological niche and to identify the environmental conditions controlling its relative abundance. Finally, environmental expression levels of genes involved in nitrogen compounds and iron uptake and metabolism were studied.

## Results

**Chromosome-scale assembly and annotation of the *P. calceolata* genome.** To measure the abundance of *P. calceolata* in the oceans and study its genetic capacity to grow in different environmental conditions, we sequenced and assembled the genome of *P. calceolata* RCC100 using long reads of Oxford Nanopore Technologies (ONT) and Illumina short reads. Using the k-mer distribution of short reads, the genome was estimated to be homozygous with a size of 31 Mb (Supplementary Fig. S1a). The ONT long reads were assembled with Flye into six nuclear contigs for a total of 32.4 Mb, 1 plastid circular contig (90 Kb) and 1 mitochondrial circular contig (39 Kb) (Fig. 1, Supplementary Fig. S1b, c, and Supplementary Data 1). Two large and highly similar duplicated regions (>99% of identity) were detected at the extremity of contig 1 and 5 (393 Kb) and at the extremities of contig 3 and 6 (192 Kb; Supplementary Note 1 and Supplementary Fig. S1d). (TTAGGG)<sub>n</sub> telomeric repeats were detected at both ends of contigs 2, 3, 4, and 6, indicating that these four contigs represent complete chromosomes (Supplementary Fig. S1d). For contig 1 and contig 5, telomeric sequences were identified at only one extremity, the other extremity ending in the duplicated region. We used the Hi-C long-range technology to validate the assembly of *P. calceolata* genome sequence. The interaction map revealed a high number of contacts within contigs and very few across contigs (Supplementary Fig. S2 and Supplementary Note 2). No chimeras or fragmentations were detected. This result confirms that the six contigs correspond to six chromosomes of *P. calceolata*.

A total of 16,667 genes were predicted on the *P. calceolata* genome (see “Methods”), which is a high number for a PPE (Table 1 and Supplementary Data 2). There was an average of 0.45 intron per gene, and the distribution of their lengths reveals a peak at around 210 bp, which is the characteristic length of Introner Elements described in *A. anophagefferens*<sup>27</sup> (Supplementary Fig. S3, Supplementary Data 3, and Supplementary Note 3). In all, 9812 *P. calceolata* predicted proteins (58%) are homologous with at least



**Fig. 1** *Pelagomonas calceolata* nuclear genome. Representation of the 6 nuclear contigs of *P. calceolata*. The top layer indicates the number of genes per 100 Kb (black bars), the middle layer represents the GC content in percentage over a window of 200 Kb and the bottom layer is the position of DNA repeats of more than 500 bases repeated at least five times over the entire genome. Red, orange, and yellow bars indicate three different repeats in low-GC regions present in at least three different contigs. Dashed red and blue rectangles are duplicated chromosomal regions.

**Table 1** Genome characteristics of several unicellular photosynthetic eukaryotes.

Phylum/class	Species	Genome size (Mb)	Number of chromosomes	Predicted genes	GC%	Cell size	References
Pelagophyceae	<i>Pelagomonas calceolata</i>	32.4	6	16,667	63.6	2 $\mu$ m	This study
Pelagophyceae	<i>Aureococcus anophagefferens</i>	56.0	Unknown	11,520	67.4	2 $\mu$ m	89
Eustigmatophyceae	<i>Nannochloropsis oceanica</i>	29.3	32	7730	54.0	3 $\mu$ m	90
Diatom	<i>Phaeodactylum tricornutum</i>	27.0	33	10,402	48.8	11 $\mu$ m	91
Diatom	<i>Thalassiosira pseudonana</i>	34.5	24	11,776	46.9	5 $\mu$ m	92
Chlorophyta	<i>Micromonas pusilla</i>	21.9	17	10,575	65	$\leq$ 2 $\mu$ m	93
Chlorophyta	<i>Ostreococcus lucimarinus</i>	13.2	21	7651	60	1.3 $\mu$ m	94
Chlorophyta	<i>Bathycoccus prasinos</i>	15.1	19	7847	48	1–2 $\mu$ m	95
Haptophyta	<i>Emiliania huxleyi</i>	141.7	Unknown	38,549	64.5	4–5 $\mu$ m	96

one gene in a stramenopile genome, including 2631 (16%) only shared with the pelagophyte *A. anophagefferens* (Supplementary Fig. S4). A conserved functional domain (Pfam, KO or InterProScan) was found in 11,240 proteins (67%). Even if gene completeness estimations are imprecise for species distant from model organisms, we obtained 94.0% of completeness with BUSCO<sup>28</sup> (88% single-copy and 6% duplicated genes), showing that our genome is near complete (Supplementary Data 4).

**GC content, centromeres, and meiosis in *P. calceolata*.** A remarkable feature in the *P. calceolata* genome is the distribution of GC content along *P. calceolata* chromosomes (Fig. 1). While the average GC content of the nuclear genome is 63%, one large region

in each contig (259 Kb on average) is 52% GC. These unique large troughs in GC content in each chromosome suggest that these regions encompass centromeres. The Hi-C result confirms centromere positions with the presence of contacts between low-GC regions across chromosomes, suggesting the physical proximity of these regions in the nucleus (Supplementary Fig. S2). Interestingly, we did not observe an accumulation of repeated elements or transposons in these low-GC regions and only a slight decrease of gene density. Genomic specificities and gene content of low-GC regions are detailed in Supplementary Data 5, S6 and Supplementary Note 4. Low-GC patterns in centromeres could be explained by the inhibition of recombination<sup>29,30</sup>, suggesting that *P. calceolata* is capable of meiosis and recombination. Among 23 meiosis-specific

genes characterized in other species<sup>31–33</sup>, 18 homologs are present in the *P. calceolata* genome (Supplementary Data 7). These genes include the double-strand DNA break (DSB) initiator *SPO11*; *RAD50*, *RAD52*, and *MRE11* to bind DSBs; *HOP2*, *MND1*, *DMC1*, and *RAD51* to ensure pairing and invade the homologous strand; *MSH2*, *MSH3*, *PMS1*, and *MSH6* genes involved in the synthesis-dependent strand annealing pathway and *MUS81* necessary for non-interfering (class II) crossing over. The five missing genes are absent in many organisms capable of meiosis, suggesting that they are not required to perform genetic recombination (Supplementary Note 5). Taken together, the genomic structure and the genetic content of *P. calceolata* strongly suggests that this species performs meiosis.

**Relative abundance of *P. calceolata* across oceanic basins.** To estimate the relative abundance of *P. calceolata* across all oceans, we first used the abundance of the V9 region of the 18 S rRNA sequenced from all samples of the Tara Oceans expedition<sup>34</sup>. The most abundant *P. calceolata* Operational Taxonomic Unit (OTU) in the 0.8–5  $\mu\text{m}$  size fraction is on average 0.80% for the 104 surface samples and 1.23% for the 61 deep-chlorophyll maximum (DCM) samples (Supplementary Data 8). According to this abundance estimation method, *P. calceolata* is the third most abundant eukaryote OTU of the 0.8–5  $\mu\text{m}$  size fraction after two Dinophyceae OTUs affiliated to *Ankistrodinium* and an unknown Gymnodiniaceae.

To estimate *P. calceolata* abundance independently from PCRs and 18 S rRNA copy number bias, we used the mapping of metagenomic reads on the *P. calceolata* genome. For the 0.8–5  $\mu\text{m}$  size fraction, the percentage of sequenced reads aligned on the genome is 1.39% ( $n = 93$ ,  $sd = 1.5$ ) in surface samples and 2.67% ( $n = 55$ ,  $sd = 1.6$ ) in DCM samples. In the 0.8–2000  $\mu\text{m}$  size fraction, *P. calceolata* represents 1.01% ( $n = 80$ ,  $sd = 1.2$ ) of all reads in surface samples and 1.56% ( $n = 39$ ,  $sd = 1.3$ ) in DCM samples. A maximal relative abundance of 6.7% in the 0.8–5  $\mu\text{m}$  size fraction was observed in the North Indian Ocean (station TARA\_38) at the DCM (Fig. 2a). In the Indian Ocean, Red Sea and Mediterranean Sea, *P. calceolata* is significantly more abundant in the DCM than at the surface (Fig. 2b). In cold waters (below 10 °C), *P. calceolata* is not detected above our threshold of 25% of horizontal genomic coverage. Important variations between and within each oceanic basin are observed, suggesting that many biotic or abiotic factors influence *P. calceolata* abundance.

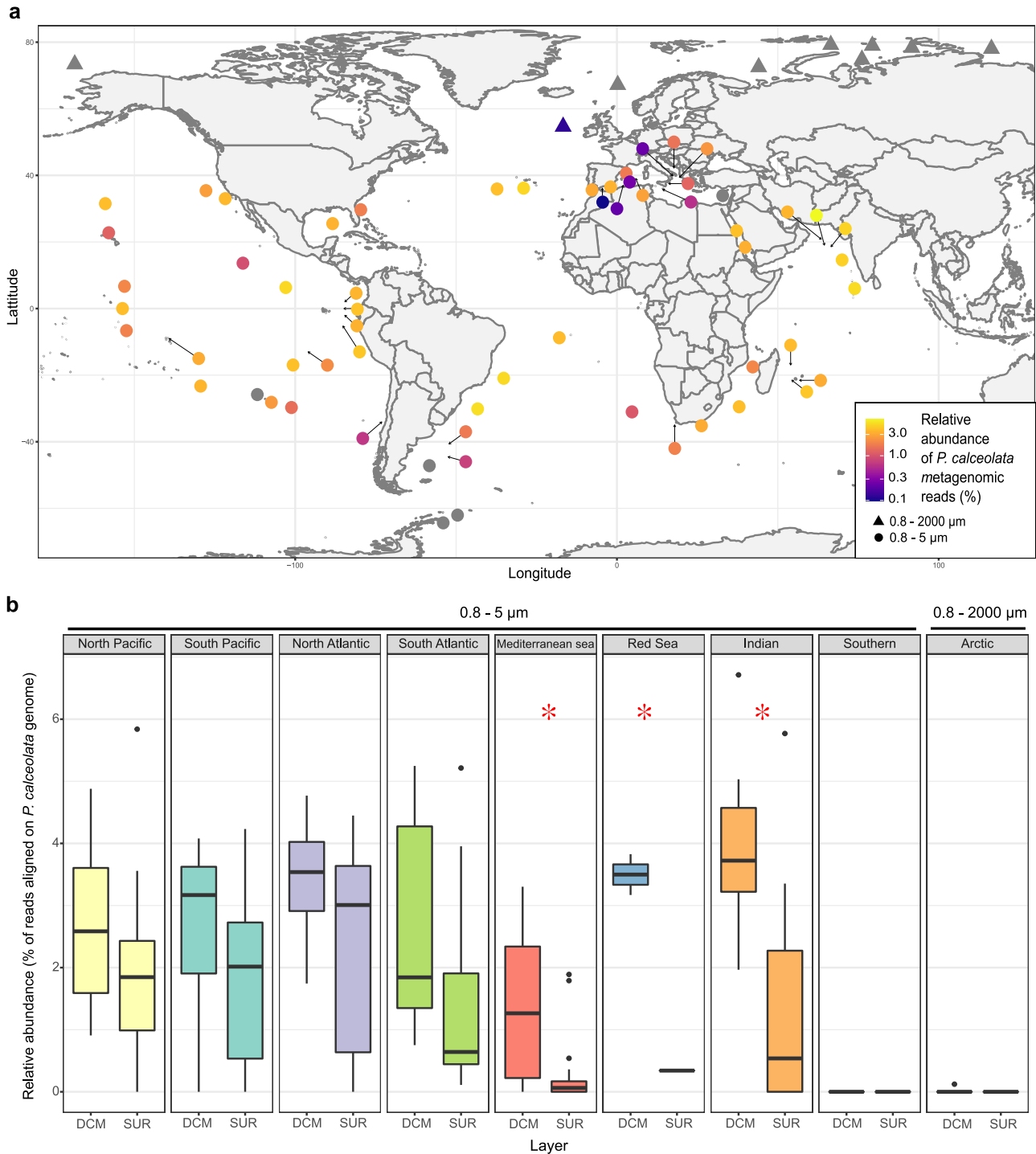
Finally, we compared the two methods of abundance estimations (Supplementary Fig. S5). The metagenomic-based relative abundance is strongly correlated to the metabarcoding-based relative abundance (Pearson correlations of 0.91 and 0.70 for the 0.8–2000 and 0.8–5  $\mu\text{m}$  size fractions, respectively). However, the metabarcoding-based abundance is on average 2.3 lower in the 0.8–5  $\mu\text{m}$  size fraction and 3.1 lower in the 0.8–2000 size fraction compared to the metagenomic-based abundance (Supplementary Fig. S5).

**High relative abundance of *P. calceolata* in temperate, low-light, and iron-poor regions.** In order to identify factors controlling *P. calceolata* abundance in the oceans, we used physical-chemical parameters available for each oceanic station (see “Methods”). Principal component analysis revealed a positive relation between metagenomic-based *P. calceolata* abundance, the temperature, and the coast distance and a negative relation with iron concentration (Fig. 3a, b). This result was consistent over the 2 size fractions containing *P. calceolata* cells (0.8–5  $\mu\text{m}$  and 0.8–2000  $\mu\text{m}$  size fractions; Supplementary Fig. S6). Despite the numerous factors potentially influencing *P. calceolata* abundance, we observed a weak but significant Pearson’s positive correlation with the

temperature, a negative correlation with Photosynthetically Active Radiation (PAR, mean of 30 days) and a negative correlation with iron concentrations (Table 2). In the 0.8–5  $\mu\text{m}$  size fraction, the relative abundance of *P. calceolata* is higher in low-iron conditions (<0.2 nmol/l, 54 samples) with on average 2.3% of metagenomic reads than in high-iron environments (>0.2 nmol/l, 88 samples) with on average 1.7% of metagenomic reads (Wilcoxon test,  $P$  value = 0.02). In the 0.8–2000  $\mu\text{m}$  size fraction, we observe the same tendency with a relative abundance of 1.9% of metagenomic reads on average in low-iron waters (49 samples) and a lower relative abundance of 0.78% of metagenomic reads on average in high-iron environments (59 samples) (Wilcoxon test,  $P$  value =  $9.6e^{-7}$ ). In addition, *P. calceolata* relative abundance is weakly correlated with the 9’butanoyloxyfucoxanthin concentration, a signature pigment for pelagophytes (Pearson 0.22,  $P$  value = 0.02 and Pearson 0.41,  $P$  value =  $4.82e^{-05}$  in the 0.8–5  $\mu\text{m}$  and 0.8–2000  $\mu\text{m}$  size fraction, respectively). We used a general additive model to estimate the contribution of temperature, PAR and iron concentration to *P. calceolata* relative abundance (Table 2). The three factors explain 32.3% of the variations of *P. calceolata* abundance in the 0.8–5  $\mu\text{m}$  size fraction and 56.8% in the 0.8–2000  $\mu\text{m}$  size fraction.

Finally, we projected the relative abundance of *P. calceolata* at the end of the century following Frémont et al. methodology<sup>35</sup>. We modeled the ecological niche of *P. calceolata* using the World Ocean Atlas (WOA18) datasets at the time and location of sampling or using the projected climatology in 2099 using the RPC8.5 scenario (see “Methods”). We used four machine-learning techniques: Generalized Additive Models (GAM), Neural Networks (nn), Random Forest (rf) and Gradient Boosted Trees (bt) and evaluated their performances with two parameters. The Pearson correlation coefficient indicates the correlation between the model and in situ measurements of *P. calceolata* abundance. The four machine-learning tools have similar performances based on Pearson’s correlations (nn = 0.676; gam = 0.621; bt = 0.683; rf = 0.694). The second parameter is the root mean square error (rmse) and reflects the magnitude of the errors in the models (the number of standard deviations from the mean). Using this metric the GAM approach is less good (rmse = 1.04) than the three other tools (nn = 0.964; bt = 0.952; rf = 0.941). These results indicate that we have enough in situ data to capture the global trends on the relative abundance of *P. calceolata* but these models could be imprecise on the amplitude of abundance variations. In addition, the predictions in the tropical waters are uncertain because this environment in 2099 is out of the range of the training dataset. Because the performances of the four models are similar, we combined them to obtain the most accurate projection (Fig. 3c and Supplementary Fig. S7). Despite these limitations, we projected an increase of up to 1.12% of *P. calceolata* relative abundance from latitude 40° to latitude 50° in the North and South hemispheres and a decrease in temperate and tropical waters (–0.8% maximum).

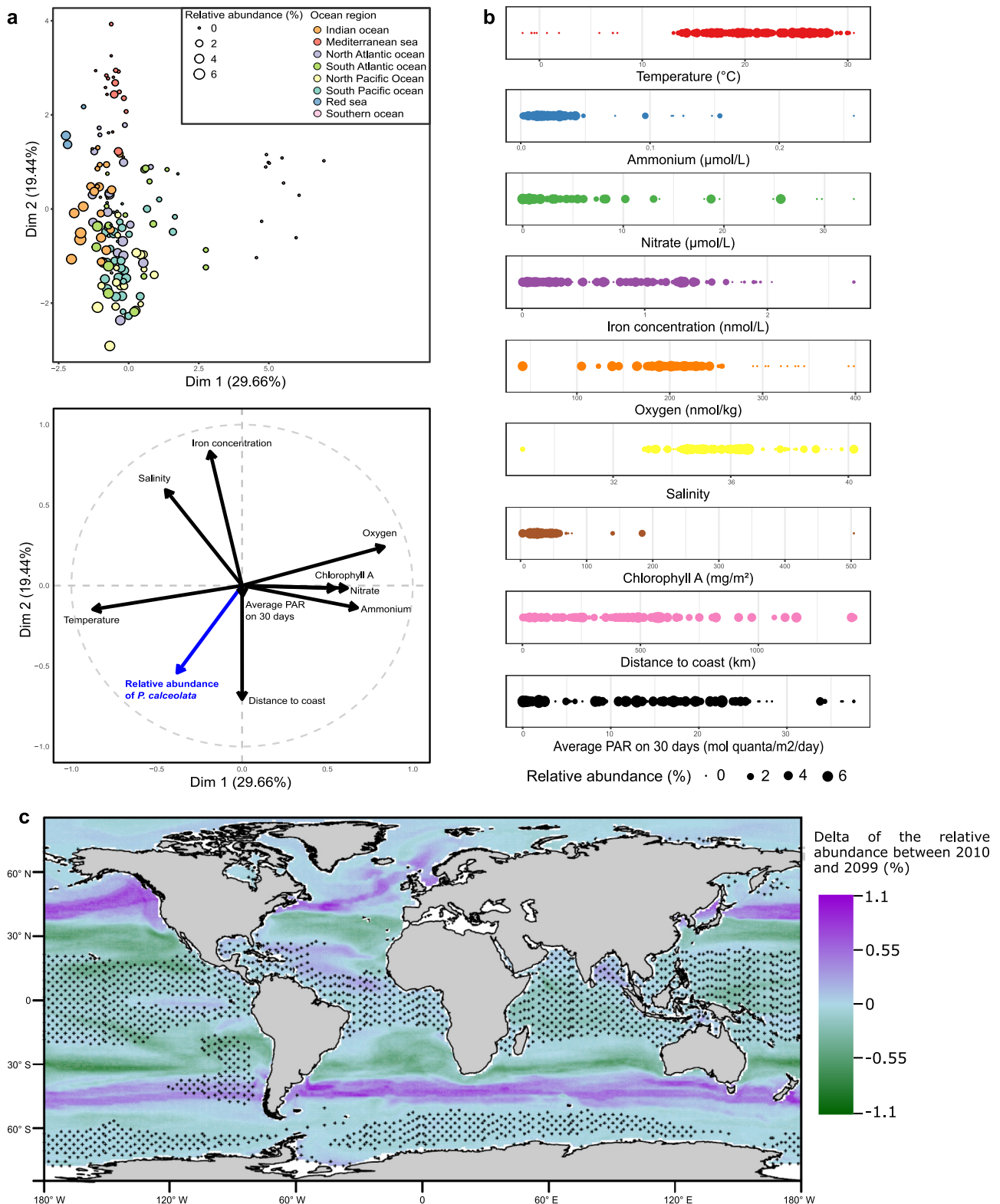
**Genes related to iron uptake, storage and usage in *P. calceolata*.** Iron is a critical metal for all photosynthetic organisms, required for the photosynthesis, the nitrogen cycle and the protection against reactive oxygen species. Since *P. calceolata* seems to thrive in iron-poor waters, we identified *P. calceolata* genes coding for iron uptake and storage, then compared their expression levels in low (<0.2 nmol/l) versus high (>0.2 nmol/l) iron conditions using Tara Oceans metatranscriptomes. *P. calceolata* has five genes encoding the phytoferritin *ISIP2A* involved in  $\text{Fe}^{3+}$  uptake via endosomal vesicles and 2 putative iron-storage protein *ISIP3*. In iron-poor environments, three *ISIP2A* and one *ISIP3* are over-expressed (Fig. 4a and Supplementary Data 9). This result indicates



**Fig. 2** Relative abundance and distribution of *P. calceolata* in the oceans. **a** World map of the relative abundance of *P. calceolata* metagenomic reads. The color code indicates the percentage of sequenced reads aligned on the genome. The DCM samples of size fractions 0.8–5  $\mu\text{m}$  (circles) or 0.8–2000  $\mu\text{m}$  (triangles) are shown. *P. calceolata* is considered to be absent when the horizontal coverage is below 25% of the genome (gray dots). **b** Boxplot of the relative abundance in each oceanic region in surface and DCM samples. Red stars indicate a significant difference between SUR and DCM samples (Wilcoxon test,  $P$  value < 0.01).

that similarly to diatoms, ISIP are upregulated following iron starvation in *P. calceolata* potentially improving cell growth in low-iron environments. Three genes encode the iron transporter ferroportin but are not differentially expressed according to the environment (Supplementary Fig. S8a). These proteins are iron exporters in multicellular organisms but their function in microalgae remains to be studied<sup>36</sup>. Finally, we identified eight Zinc/iron permeases potentially involved in iron uptake from the

environment in the *P. calceolata* genome. Among them, two are overexpressed in high-iron and one in low-iron environments. Interestingly, we note the absence of the iron permease *FTR1*, the iron-storage ferritin and the starvation induced protein *ISIP1* (involved in endocytosis of siderophores in diatoms). In comparison to *P. calceolata*, the coastal Pelagophyceae *A. anophagefferens* do not have *ISIP3* gene and a lower number of Zinc-Iron permeases.



**Fig. 3 Ecological niche of *P. calceolata*.** **a** Principal component analysis of the metagenomic-based relative abundance of *P. calceolata* in the 0.8–5 μm size fraction. Percentages of variance explained by each axis are indicated on axis titles. Top panel: each dot represents a sample with a size proportional to the relative abundance of *P. calceolata* and the colors indicate the oceanic basins. Bottom panel: nine environmental parameters are represented as vectors alongside the relative abundance of *P. calceolata* (blue vector). **b** Bubble plot of the relative abundance of *P. calceolata* for the 0.8–5 μm size fraction according to the nine environmental parameters. **c** Delta of the modeled relative abundance of *P. calceolata* between 2010 and 2099. Green areas correspond to a decrease while purple areas correspond to an increase of *P. calceolata* relative abundance. Small stars indicate locations where at least one of the predictor drivers is out of range compared to the training dataset values.

**Table 2 Environmental parameters explaining *P. calceolata* relative abundance for the 0.8–5 µm (a) and the 0.8–2000 µm (b) size fractions.**

(a)	GAM model			GAM verification		Pearson correlations	
	edf	F value	P value	k-index	k P value	r	P value
0.8–5 µm							
s(Temperature)	1	22.16	6.84e <sup>-6</sup>	1.11	0.87	0.23	0.001
s(Iron concentration)	1.257	13.12	2.26e <sup>-4</sup>	0.93	0.12	-0.25	0.001
s(PAR 30 days)	1.859	15.94	4.56e <sup>-7</sup>	1.01	0.44	-0.32	0.001
Adjusted R <sup>2</sup>	0.3						
Deviance explained	32.30%						
(b)	GAM model			GAM verification		Pearson correlations	
0.8–2000 µm	edf	F value	P value	k-index	k P value	r	P value
s(Temperature)	3.628	9.442	1.71e <sup>-6</sup>	0.96	0.31	0.57	0.0001
s(Iron concentration)	1.54	1.225	2.42e <sup>-1</sup>	0.91	0.14	-0.47	0.0001
s(PAR 30 days)	2.454	6.962	0.00027	1.14	0.93	-0.051	0.6
Adjusted R <sup>2</sup>	0.53						
Deviance explained	56.80%						

Several important ferrous proteins can be substituted by non-ferrous equivalents in iron-poor environments<sup>14</sup>. In the *P. calceolata* genome we identified 11 flavodoxin genes involved in electron transfer during photosynthesis, potentially replacing ferredoxins (17 genes; Supplementary Data 9). This number of genes is important compared to other algae, including *A. anophagefferens* (5 flavodoxins and 9 ferredoxins). Expression levels of these genes across the oceans revealed overexpression of flavodoxin genes in low-iron environments, replaced by ferredoxins in high-iron conditions (Fig. 4b). The Fructose-Bisphosphate Aldolase (FBA), necessary for gluconeogenesis and the Calvin cycle, is encoded by six genes in *P. calceolata*. Two genes are dependent on a divalent cation (FBA type II), and the four others are Zinc/Iron-independent (FBA type I). FBA type I is overexpressed in high-iron conditions and quasi-absent in low-iron conditions (Fig. 4c). Finally, all types of Superoxide dismutases (SOD) are found in the *P. calceolata* genome. Non-ferrous SODs (Cu/Zn and Ni) encoded by three genes and Mn/Fe-SOD encoded by two genes are not differentially expressed according to iron concentrations (Supplementary Fig. S8b). The gene content and the transcriptomic flexibility suggest important capacities for *P. calceolata* growth in low-iron environments.

### Genes involved in nitrogen uptake, storage, and recycling.

Because *P. calceolata* could be an important player in the nitrogen (N) cycle in oceanic ecosystems<sup>25</sup>, we explored its gene content and analyzed the expression levels of genes involved in nitrogen metabolism in the environment using *Tara* Oceans metatranscriptomes (Fig. 5 and Supplementary Data 9). The uptake of nitrogen-containing inorganic compounds is supported by 11 genes in *P. calceolata*. Three genes encode nitrate/nitrite transporters, three genes encode formate/nitrite transporters and five genes encode ammonium transporters. One formate/nitrite and two nitrate/nitrite transporter genes are significantly overexpressed in low-nitrate conditions (Supplementary Fig. S9). One ammonium transporter is overexpressed in low-ammonium environments (Supplementary Fig. S9). We identified one nitrate reductase and two nitrite reductases in the *P. calceolata* genome (Supplementary Data 9). The expression level is significantly higher in high-nitrate environments for the nitrate reductase and one nitrite reductase (Supplementary Fig. S10). But the second nitrite reductase is surprisingly less expressed in high-nitrate conditions. The number of enzymes incorporating ammonium into organic compounds (GS/GOGAT pathway) is higher in *P. calceolata* than in other species: five glutamine

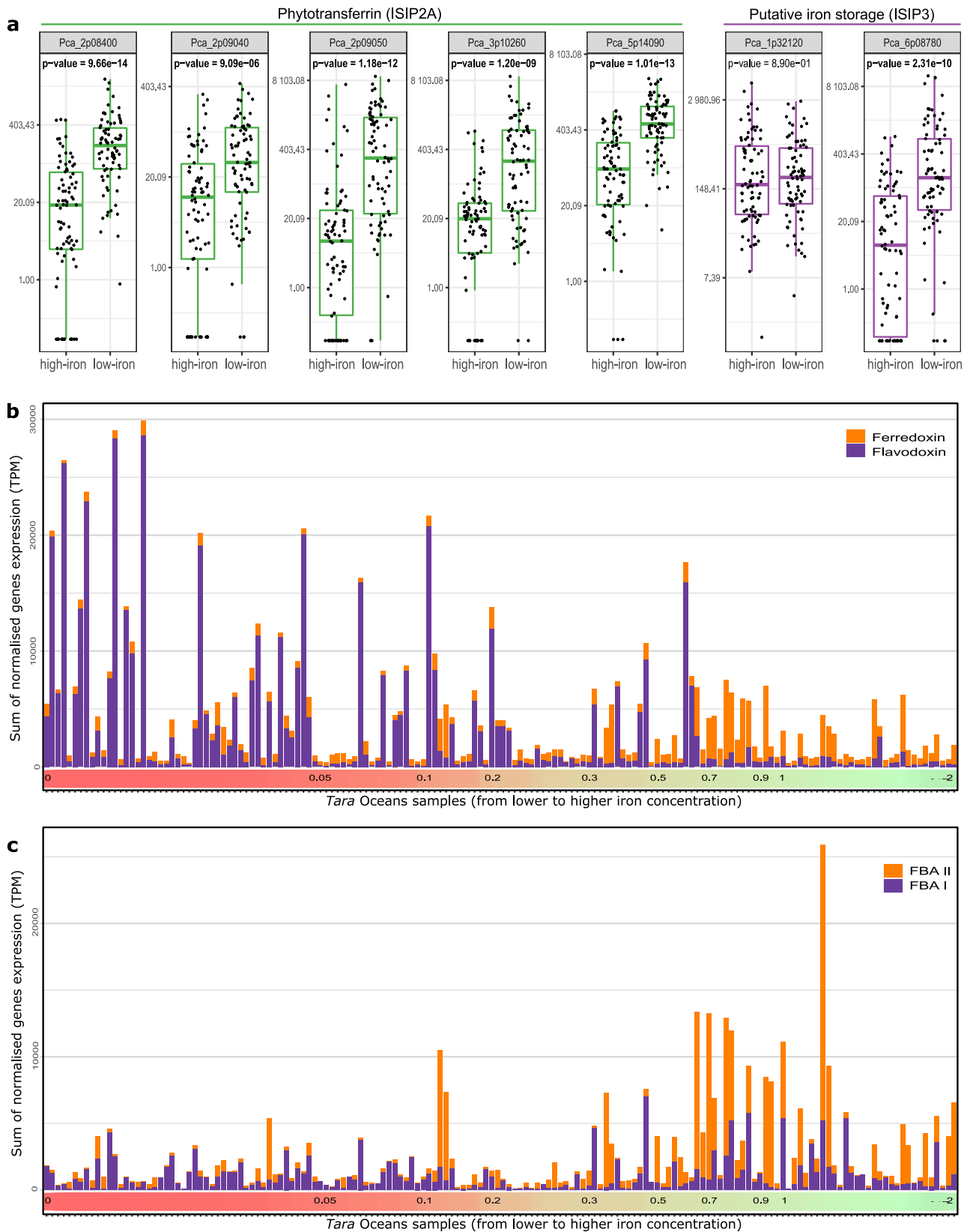
synthetase (GS) and four glutamate synthase (GOGAT) genes are present in the *P. calceolata* genome. Two GS genes are more expressed in high-nitrate samples and 1 GOGAT gene is more expressed in low-nitrate samples (Supplementary Fig. S10).

We identified three genes carrying the nitrate and nitrite sensing (NIT) domain (IPR013587) in the *P. calceolata* genome. Using NCBI non-redundant proteins and marine genomic databases (see “Methods”), 60 homologous proteins of the NIT-sensing domains of *P. calceolata* were identified. These homologs are restricted to the Pelagophyceae class (16 transcriptomes), the Dictyochophyceae class (6 transcriptomes) and one putative cryptophyte transcriptome. The phylogenetic tree of this protein family shows three subfamilies diverging before the Dictyochophyceae/Pelagophyceae separation (Supplementary Fig. S11). One *P. calceolata* protein carries a NIT-sensing domain surrounded by two transmembrane domains suggesting a capacity for external nitrate/nitrite sensing while the two other NIT genes carry a protein-kinase domain (IPR000719) suggesting phosphorylation-based signal transduction dependent on intracellular nitrate or nitrite concentration (Fig. 5b). To investigate this possibility, we studied the expression levels of NIT-sensing genes. The potential external NIT-sensing gene is indeed significantly overexpressed in low-nitrate environments. In contrast, only one of the intracellular NIT-sensing genes is upregulated in nitrate-rich environments (Fig. 5c). This result suggests an important role of the NIT-sensing genes in the acclimation to environmental nitrate concentrations.

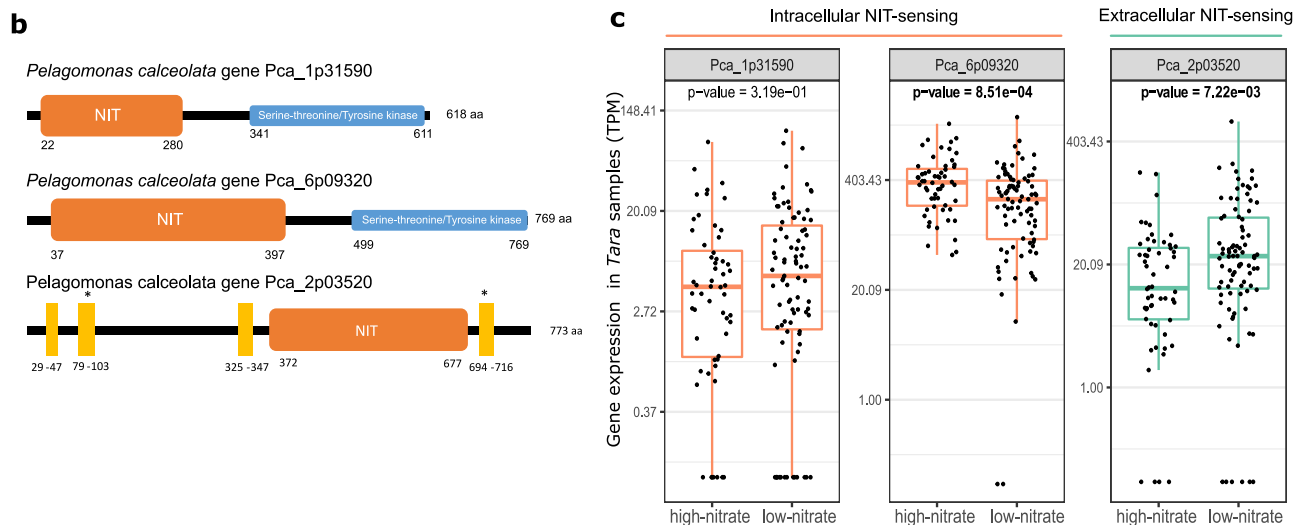
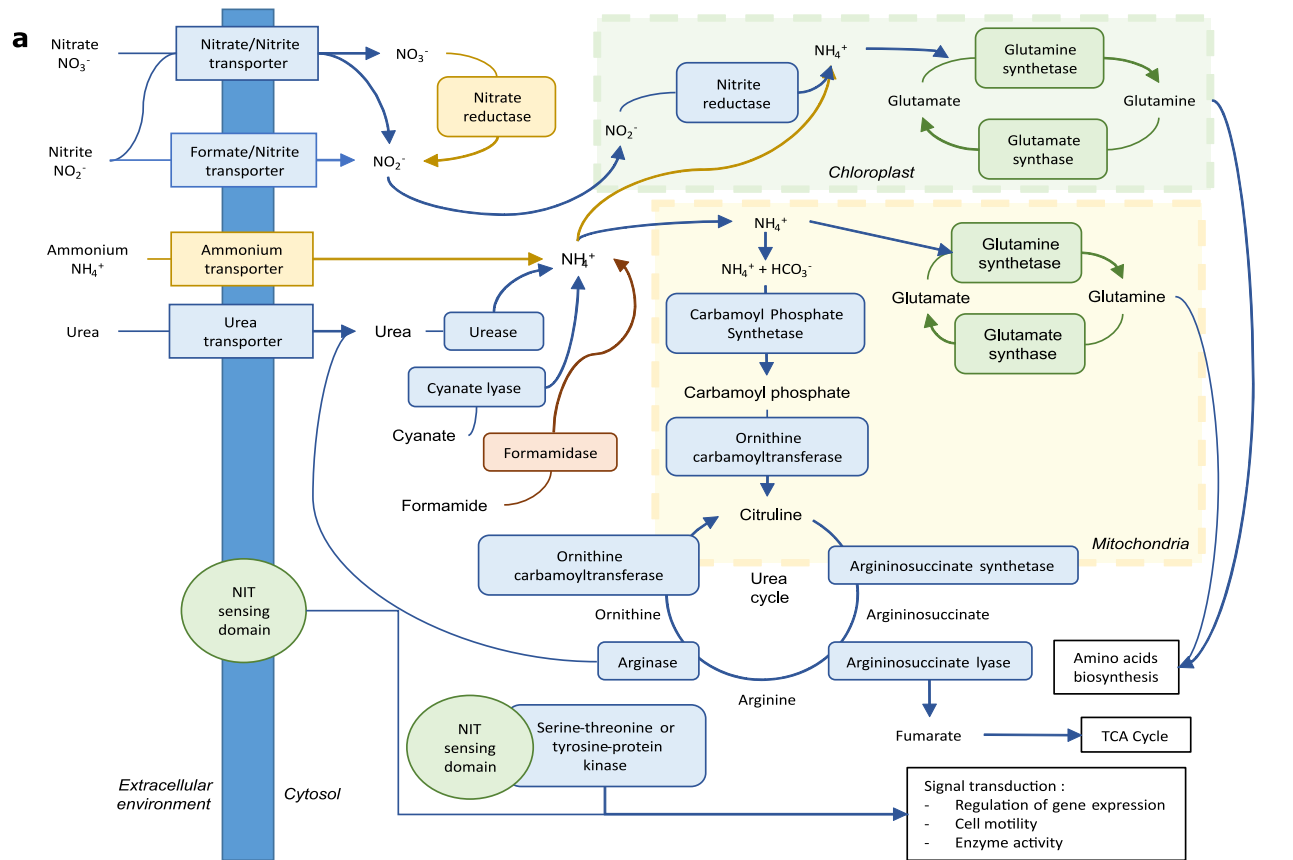
Finally, we identified genes involved in nitrogen recycling from organic compounds which are important in several species in case of inorganic nitrogen deprivation. One arginase and one cyanate lyase were detected in the *P. calceolata* genome but no gene encoding formamidase. In addition, the number of gene copies for enzymes involved in the urea cycle (carbamoyl-phosphate synthetase, ornithine carbamoyltransferase, argininosuccinate synthase and argininosuccinate lyase) is equal or slightly lower than in other algae (Fig. 5a and Supplementary Data 9). Among these genes, only the cyanate lyase is overexpressed in low-nitrate conditions suggesting cyanate is an important alternative source of organic nitrogen for *P. calceolata* (Supplementary Fig. S10).

### Discussion

The essential roles of phytoplankton in oceanic ecosystems have been illustrated many times, however, numerous lineages are still poorly explored and model organisms are restricted to a few taxa (mainly diatoms, prasinophytes, and haptophytes) limiting the global understanding of phytoplankton activity. The *P. calceolata*



**Fig. 4** Expression of iron-related genes in *P. calceolata*. **a** Relative gene expression levels normalized in transcript per million (TPM) of five phytotransferrins (*ISIP2A*) and two putative iron storage (*ISIP3*) in low-iron (<0.2 nM) and high-iron (>0.2 nM) oceanic stations. *P* values of Wilcoxon statistical tests between low- and high-iron conditions are indicated for each gene. Significant *P* values (<0.01) are in bold. **b** Relative expression levels (TPM) of genes coding for ferredoxin (orange) and its non-ferrous equivalent flavodoxin (purple) in each Tara Oceans sample. Samples are sorted from low-iron (left) to high-iron (right) conditions. Iron concentrations are indicated in nM on the colored horizontal bar. **c** Same representation for genes coding for fructose-bisphosphate aldolase II (orange) and its non-ferrous equivalent fructose-bisphosphate aldolase I (purple).



**Fig. 5 Nitrogen sensing and metabolism in *P. calceolata*.** **a** Schematic representation of N transport and assimilation in *P. calceolata* based on the gene content. The color code indicates if the number of gene copies for a specific function is overrepresented (green), equally represented (blue), underrepresented (orange) or absent (red) in *P. calceolata* genome compared to the mean of eight pico-nano photosynthetic eukaryotes. Gene copy number for each function is indicated in Supplementary Data 9. **b** Domain organization of NIT-sensing proteins in *P. calceolata*. Orange boxes are NIT-sensing domains (IPR13587), blue boxes are serine-threonine/tyrosine-kinase domain (IPR20635), and yellow rectangles are transmembrane domains. **c** Relative expression levels (TPM) of three NIT-sensing genes in low-nitrate (<2 μM) and high-nitrate (>2 μM) environments. *P* values of Mann-Whitney-Wilcoxon tests between low- and high-nitrate samples are indicated for each gene. Significant *P* values (<0.01) are in bold.

genome assembled and annotated in this study reveals a previously underestimated high abundance of *P. calceolata* in the oceans and brings new insights into specific genomic features of this algae class related to its adaptation to specific environments.

We have shown in this study that *P. calceolata* is cosmopolitan in oceanic samples above 10 °C with a relative abundance

generally >1% of all sequenced reads. In contrast to the coastal Pelagophyceae *A. anophagefferens* that can present high peaks of abundance<sup>37</sup>, no *P. calceolata* blooms were reported, but *P. calceolata* is well-adapted to an extensive range of environmental conditions as suggested by previous studies<sup>21,23</sup>. Although the abundance of an organism calculated from metabarcoding or

metagenomic data provides only an indirect and relative quantification of organism abundances, both methods suggest that *P. calceolata* is one of the most abundant pico-nano eukaryote in offshore data. The high relative abundance of *P. calceolata* measured with a metabarcoding approach has recently been confirmed with a qPCR method (average of  $5\,882 \pm 2\,855$  rRNA gene copies  $\text{mL}^{-1}$  on the surface of the eastern North Pacific)<sup>21</sup>. In addition, we have shown that the metabarcoding approach probably underestimates the relative abundance of *P. calceolata* compared to the metagenomic analysis owing to the low copy number of rRNAs in this organism. However, we cannot exclude that the large genome size of *P. calceolata* compared to bacterial genomes present in the 0.8–5  $\mu\text{m}$  size fraction overestimates its relative abundance in metagenomic datasets. Further studies may combine microscopic and flow sorting approaches with genomic data to assess the number of cells and the biomass of this organism in the oceans. Our model analysis has revealed a probable increase of *P. calceolata* relative abundance at the end of the century in high latitudes where the seawater temperature is currently too low for this species. This result is coherent with previous studies suggesting a global increase of phytoplankton in subpolar regions<sup>38,39</sup>.

Iron is essential for growth, photosynthesis, primary production, nitrogen fixation and reduction for PPEs<sup>40</sup>. Our results show that *P. calceolata* thrives in iron-poor waters and thus occupies a large ecological niche for a PPE. Two main strategies exist against iron deprivation: the optimization of iron uptake and the modulation of iron needs. Genes coding for iron chelators and ferritin are absent from the *P. calceolata* genome, and genes coding for passive iron transporters are under- or equally represented compared to other PPEs. In contrast, phytoferroportins (*ISIP2A*), putative iron-storage proteins (*ISIP3*) and ferroportins are over-represented in the *P. calceolata* genome. The expression levels of *ISIP* genes are anti-correlated with iron concentration in *P. calceolata* showing an acclimation to low-iron conditions. Because phytoferroportins are dependent on carbonate ions, ocean acidification may reduce iron uptake efficiency in many species including *P. calceolata*<sup>41,42</sup>. Compared to *P. calceolata*, the low abundance of *A. anophagefferens* in open oceans where iron is limited is consistent with the absence of genes involved in iron uptake and storage, including iron permeases and *ISIP3* genes. The presence of three ferroportin genes in *P. calceolata* is interesting since these transmembrane iron export proteins play a significant role in iron homeostasis in multicellular organisms<sup>36</sup>. Ferroportin function in microalgae is unknown but could act to export iron from endosomes to the cytoplasm<sup>43</sup>. In the green alga *Chlamydomonas reinhardtii*, a ferroportin gene is overexpressed under low Fe conditions<sup>44</sup>. Even though we did not find significant changes in the expression of the 3 ferroportin genes of *P. calceolata* according to iron concentration, the function of these genes could be investigated to understand their role in variable iron concentrations.

Modulation of iron needs seems to be a major acclimation strategy for *P. calceolata* in low-iron environments. All known molecular switches between ferrous and non-ferrous proteins are present in the *P. calceolata* genome and the transcriptomic regulation of these genes suggests a central role of these proteins for its growth in low-iron environments.

Expressing more than 90% of all nitrate transporter transcripts, pelagophytes may dominate nitrate uptake and assimilation in the North Pacific Ocean<sup>25</sup>. Indeed, *P. calceolata* contains a large collection of genes for inorganic nitrogen transporters. The main difference with the coastal *A. anophagefferens* is the reduced number of ammonium transporters in *P. calceolata*.

Organic nitrogen compounds could also be a major nitrogen source of for *P. calceolata*. We have shown that the cyanate lyase

and urease genes are expressed in many environments but only the cyanate lyase is overexpressed in low-nitrate conditions. These two genes, largely present among phytoplankton lineages, could be significant components of acclimation to low-nitrate environments<sup>45</sup>. In addition, *A. anophagefferens* grow faster on the organic nitrogen substrates (urea and glutamic acid) than on nitrate or ammonium, suggesting that the dominance of Pelagophyceae in low-nitrate environments could be due to an optimized usage of these organic molecules<sup>46</sup>.

One remarkable feature of the *P. calceolata* genome is the presence of three genes carrying NIT domains (PF08376). This NIT domain was first described in bacterial nitrite and nitrate sensor proteins<sup>47</sup>. This sensor is an alpha-helical protein playing a signal transduction role in regulating gene expression, cell motility and enzyme activity in *Klebsiella oxytoca*<sup>48</sup>. In pico-nano algae, NIT-sensing domains can be associated with a serine-threonine/tyrosine-kinase domain, suggesting signal transduction according to intracellular nitrate/nitrite concentration, or surrounded by two transmembrane domains suggesting extracellular sensing. Even though NIT-sensing domains can be found across various phyla, homologs of *P. calceolata* NIT proteins are restricted to pelagophytes and dictyochophytes. The NIT genes in *P. calceolata* are highly expressed in subtropical Pacific N-depleted waters, suggesting that these proteins have a role in transcription regulation according to nitrate availability<sup>25</sup>. Our results suggest that NIT-sensing proteins respond differently to environmental nitrate depletion. We can hypothesize that the NIT-sensing protein overexpressed in nitrate-rich environments plays a role in the intracellular regulation of stored nitrogen, activating pathways when nitrate or nitrite stocks are sufficient. In contrast, putative extracellular NIT-sensing could be an environmental nitrate or nitrite sensor activated to regulate the expression of genes involved in the acclimation to low-nitrate conditions.

In summary, due to its widespread distribution and its high abundance in the open oceans, *Pelagomonas calceolata* can serve as an ecologically-relevant model to study marine photosynthetic protists. We used the chromosome-scale genome sequence, mostly telomere-to-telomere, generated in this study to estimate its abundance in environmental datasets. We have shown that the *P. calceolata* genome has specific genomic features potentially explaining its ecological success in open oceans. Compared to the coastal Pelagophyceae *A. anophagefferens*, the *P. calceolata* genome contains specific genes involved in the acclimation to low-iron conditions. The large repertoire of genes involved in nutrient acquisition from the environment is coherent with its widespread pattern of relative abundance distribution across different environments. The ecological niche of *P. calceolata* suggests that this alga will benefit from the global climate change with the extension of oligotrophic regions and global ocean warming. Future studies could use the *P. calceolata* genome to explore adaptation and acclimation processes controlling the distribution and abundance of this alga.

## Methods

**Pelagomonas culture.** *Pelagomonas calceolata* RCC100 culture was grown in 12:12-h light:dark photoperiod in K medium with natural seawater base at 20 °C. At the Roscoff Culture Collection, cells were kept at a light intensity of  $\sim 80 \mu\text{mol photon m}^{-2} \text{s}^{-1}$  and the volume of culture was ramped up to 1 litre in mid-exponential growth phase before harvesting. RCC100 culture was not axenic and grown in the presence of undefined bacterial microbiota.

**DNA extraction, library preparation, and sequencing.** We pelleted cells from 500 ml of culture by two successive centrifugations at  $10,000 \times g$  for 15 min at 4 °C. Genomic DNA was extracted using the NucleoSpin Plant II Mini kit according to the manufacturer's instructions (Macherey-Nagel, Hoerd, France) with the following exception for the lysis step: 400  $\mu\text{L}$  of lysis buffer PL1 and 25  $\mu\text{L}$  of proteinase K 25 mg/mL were added to strain pellets, and lysates were incubated at 55 °C for 1 h at 900 rpm. DNA quantity and integrity were respectively evaluated

on a Qubit 2.0 spectrofluorometer (Invitrogen, Carlsbad, CA, USA) and a Nanodrop1000 spectrophotometer (Thermo Fisher Scientific, MA, USA).

For Illumina sequencing, DNA (1.5 µg) was sonicated using a Covaris E220 sonicator (Covaris, Woburn, MA, USA). Fragments were end-repaired, 3'-adenylated and Illumina adapters (Bioo Scientific, Austin, TX, USA) added using the Kapa Hyper Prep Kit (KapaBiosystems, Wilmington, MA, USA). Ligation products were purified with AMPure XP beads (Beckmann Coulter Genomics, Danvers, MA, USA). The library was then quantified by qPCR using the KAPA Library Quantification Kit for Illumina Libraries (KapaBiosystems), and the library profile was assessed using a High Sensitivity DNA kit on an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The library was sequenced on an Illumina NovaSeq instrument (Illumina, San Diego, CA, USA) using 150 base-length read chemistry in paired-end mode.

For ONT sequencing, the library was prepared using the 1D Native barcoding genomic DNA (with EXP-NBD104 and SQK-LSK109). Genomic DNA fragments (1 µg) were repaired and 3'-adenylated with the NEBNext FFPE DNA Repair Mix, and the NEBNext® Ultra™ II End Repair/da-Tailing Module (New England Biolabs, Ipswich, MA, USA). Adapters with barcodes provided by ONT were then ligated using the NEB Blunt/TA Ligase Master Mix (NEB). After purification with AMPure XP beads (Beckmann Coulter, Brea, CA, USA), the sequencing adapters (ONT) were added using the NEBNext Quick T4 DNA ligase (NEB). The library was purified with AMPure XP beads (Beckmann Coulter), then mixed with the Sequencing Buffer (ONT) and the Loading Bead (ONT), and loaded on a MinION R9.4.1 flow cell. Reads were basecalled using Guppy 3.1.5.

For the Hi-C sequencing, the library was prepared using the Dovetail Omni-C kit (Dovetail Genomics, Scotts Valley, CA, USA). A *P. calceolata* RCC100 culture (60 mL corresponding approximately to  $6 \times 10^7$  cells) was first centrifuged at  $5000 \times g$  for 10 min. The pellet was processed as mammalian cells, following the Mammalian Cell Protocol for Sample Preparation (version 1.4) without using DSG cross-linking reagent. Briefly, the chromatin was fixed with formaldehyde, randomly digested with DNase I and then extracted. Chromatin ends were repaired and ligated to a biotinylated bridge adapter, followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed and DNA was purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments, and a sequencing library was generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of the library. The Dovetail Hi-C library quality was checked as described above and sequenced on an Illumina MiSeq instrument (Illumina, San Diego, CA, USA) in paired mode (2\*150 bp), producing 2,832,092 reads. The Hi-C raw reads were aligned against the assembly (-s none option) using Juicer (Juicer version 1.5.6 - Juicer Tools Version 1.9.9)<sup>49</sup>. The contact map representation was generated with R version 4.1.1 using the merged nudups file.

**RNA extraction, library preparation, and sequencing.** When the cell concentration reached 10 million cells/mL in the mid-exponential growth phase, 160 mL of culture were collected by three successive filtrations on 1.2-µm polycarbonate filters of 47 mm to avoid prokaryotic contamination. To preserve cell and RNA integrity, we kept filtration time and pressure below 10 min and 20 mmHg, respectively. Then filters were stored in 15-mL Falcon tubes with 3 mL of Trizol (Invitrogen, Carlsbad, CA, USA), mixed and flash-frozen in liquid nitrogen for further processing. RNA was extracted by incubation at 65 °C for 15 min, followed by chloroform extraction. The aqueous phase was purified using a Purelink RNA Isolation kit (Ambion Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. DNA contamination was removed by digestion using the TURBO DNA-free™ Kit (Ambion Invitrogen) according to the manufacturer's DNase treatment protocol. After two rounds of 30-min incubation at 37 °C, the efficiency of DNase treatment was assessed by PCR. Quantity and quality of extracted RNA were analyzed with RNA-specific fluorimetric quantitation on a Qubit 2.0 Fluorometer using Qubit RNA HS Assay (Invitrogen). The qualities of total RNA were checked by capillary electrophoresis on an Agilent Bioanalyzer, using the RNA 6000 Nano LabChip kit (Agilent Technologies, Santa Clara, CA).

RNA-Seq library preparation was carried out from 1 µg total RNA using the TruSeq Stranded mRNA kit (Illumina, San Diego, CA, USA), allowing mRNA strand orientation. Briefly, poly(A) + RNAs were selected with oligo(dT) beads, chemically fragmented and converted into single-stranded cDNA using random hexamer priming. After second strand synthesis, double-stranded cDNA was 3'-adenylated and ligated to Illumina adapters. Ligation products were PCR-amplified following the manufacturer's recommendations. Finally, the ready-to-sequence Illumina library was quantified by qPCR using the KAPA Library Quantification Kit for Illumina libraries (KapaBiosystems, Wilmington, MA, USA), and evaluated with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The library was sequenced using 101 bp paired-end read chemistry on a HiSeq2000 Illumina sequencer. Low-quality nucleotides (Q < 20) from both ends of the reads were discarded. Illumina sequencing adapters and primer sequences were removed and reads shorter than 30 nucleotides after trimming were discarded. These trimming and cleaning steps were achieved using in-house-designed software based on the FastX package (<https://www.genoscope.cns.fr/externe/fastxtend/>). The last step identifies and discards read pairs mapped to the phage phiX genome, using the SOAP aligner<sup>50</sup> and the Enterobacteria phage PhiX174 reference sequence

(GenBank: NC\_001422.1). This processing, described in ref. <sup>51</sup>, resulted in high-quality data. Moreover, ribosomal RNA-like reads were excluded using SortMeRNA<sup>52</sup> 2.1 and SILVA databases.

**Long-read-based genome assembly.** Raw nanopore reads were used for genome assembly. The taxonomic assignment was performed using Centrifuge<sup>53</sup> version 1.0.3 to detect potential contamination. Genome size and heterozygosity rate were estimated using Genomescope<sup>54</sup> and Illumina short reads. For the genome assembly, we generated three sets of ONT reads: all the reads, 30× genome coverage with the longest reads and 30× genome coverage of the highest quality reads estimated by the Filtlong tool (<https://github.com/rwrick/Filtlong>). We applied Filtlong filtering with default parameters using *Pelagomonas* Illumina short reads as a reference (ONT reads covered by Illumina reads have higher scores). We then applied four different assemblers with default settings, Smartdenovo, Redbean, Flye and Ra on these three sets of reads (Supplementary Data 1)<sup>55–58</sup>. After the assembly phase, we selected the best assembly (Flye with all reads) based on the cumulative size and fragmentation. Indeed, the Wtdbg2 and Smartdenovo assembler generated fragmented assemblies with lower N90. Raven and Flye were very close, but only the Flye assembly with all ONT reads contained both the mitochondrial and chloroplastic circular contigs. To display connections that are not present in the contigs file, we used Bandage tool<sup>59</sup>. The selected Flye assembly was polished three times using Racon<sup>60</sup> with ONT reads, and two times with Haplo-G<sup>61</sup> and Illumina reads. Gene completeness of the assembly was estimated using the single-copy orthologous gene analysis from BUSCO v5 with the stramenopile dataset version 10 containing 100 genes<sup>28</sup>.

**Repeat masking and GC analyses.** Repetitive regions on the genome were masked using Tandem Repeat Finder tool<sup>62</sup>. Dust tool to detect low-complexity regions<sup>63</sup> and RepeatMasker<sup>64</sup> to identify interspersed repeats based on homology search within the Stramenopile clade and other low-complexity sequences. The positions of detected repeats were merged and hard-masked on the genome, amounting to 8% of its length. Ab initio identification of repeat family sequences was performed using RepeatScout<sup>65</sup>. The algorithm first calculates the frequency of all k-mers in the genome, then removes low-complexity regions and tandem repeats. In >80% of the cases, repeat families identified using ab initio approaches do not overlap with repetitive regions identified by homology search. GC content along the genome was calculated with Bedtools nuc version 2.29.2<sup>66</sup> and the coverage over a non-overlapping window of 2 Kb with Mosdepth version 0.2.8<sup>67</sup>.

**Transcriptome assembly.** RNA sequencing reads from *P. calceolata* RCC100 were assembled using Velvet 1.2.07 and Oases 0.2.08 with a k-mer size of 63 bp<sup>68,69</sup>. Reads were mapped back to the contigs with BWA-mem<sup>70</sup> and only consistent paired-end reads were kept. Uncovered regions were detected and used to identify chimeric contigs. In addition, open reading frames (ORF) and domains were searched using respectively TransDecoder (<http://transdecoder.sourceforge.net>) and CDDsearch<sup>71</sup>. Contig extremities without predicted ORFs or functional domains were removed. Lastly, we used the read strand information to orient RNA contigs. We completed the RNA contigs dataset with the two transcriptome assemblies of the RCC100 strain of *P. calceolata* from the Marine Eukaryotes Transcriptomes database (METdb) (<http://metdb.sb-roscoff.fr/metdb/>)<sup>72</sup>.

**Gene prediction.** Nuclear gene prediction was performed using 23,696 Pelagomonadales proteins (mainly *A. anophagefferens*) downloaded from the NCBI website. Proteins were aligned on the genome in a two-step strategy. First, BLAT<sup>73</sup> (version 36 with default parameters) was used to rapidly localize corresponding putative regions of these proteins on the genome. The best match and the matches with a score greater than or equal to 90% of the best match score were retained. Then, the regions with BLAT alignments were masked and we aligned the same set of proteins using BLAST<sup>74</sup>, which can identify more divergent matches. Second, alignments were refined using Genewise<sup>75</sup> (version 2.2.0 default parameters, except the -splice model option to detect non-canonical splicing sites), which is more accurate for detecting intron/exon boundaries. Alignments were kept if more than 50% of the length of the protein is aligned on the genome. Additionally, the transcriptome assemblies of *P. calceolata* RCC969, RCC2362, RCC706 and RCC981 included in the METdb were translated into proteins and aligned to the genome using BLAT, a BLAT score > 50 % filter, and alignments refined with Genewise as previously described.

We selected alignments from the newly generated transcriptome assembly and the two assemblies available in METdb belonging to the *P. calceolata* RCC100 strain to build a training set for the AUGUSTUS ab initio gene predictor<sup>76</sup>. Only gene models with complete coding DNA sequences were retained for training and 1000 genes were set aside for testing AUGUSTUS accuracy. Initial training produced exon and intron parameters for *P. calceolata* species. Parameters were optimized using successive steps of training and testing. We calculated gene prediction accuracy by running AUGUSTUS on the test set. At the exon level, AUGUSTUS performed well in terms of sensitivity (0.619) and specificity (0.669). We thus run AUGUSTUS on the masked genome based on trained parameters.

The ab initio prediction and all the transcriptomic and protein alignments were combined using Gmove, an easy-to-use predictor with no need for a pre-calibration

step<sup>77</sup>. Briefly, putative exons and introns, extracted from predictions and alignments, were used to build a graph, where nodes and edges represent exons and introns respectively. From this graph, Gmove extracts all paths and searches open reading frames (ORFs) consistent with the protein evidence. We trimmed untranslated transcribed regions that overlapped the coding part of a neighbor gene and renamed the genes following the standard nomenclature. Mono-exonic genes models encoding proteins of less than 200 amino acids without significant protein match (1006 genes) were excluded. Chloroplast and mitochondrial genes (contig 7 and 8) were predicted using previously published annotations for *P. calceolata*<sup>23,78</sup>. Following this pipeline, we predicted 16,667 genes with 0.45 intron per gene on average.

**Functional annotation.** Predicted gene models of *P. calceolata* nuclear genome (contig 1 to contig 6) were annotated for protein function using InterProScan v5.41-78.0<sup>79</sup>. A protein alignment against the NR database (01-12-2021 version) was performed with diamond v0.9.24<sup>80</sup>. The best protein match with a functional annotation and an *e*-value < 10<sup>-5</sup> was retained. KEGG Orthologues (KO) were identified with the HMM search tool KofamKoala v1.3.0 and KO annotations with an *e*-value < 10<sup>-5</sup> and a score above the HMM threshold were retained<sup>81</sup>. Finally, Gene Ontology (GO) terms and Enzyme commission (EC) numbers were recovered from the Interproscan and KO analysis respectively. Previously published chloroplast and mitochondrial gene names and functions were reported on the corresponding genes<sup>23,78</sup>. All functional gene annotations of *P. calceolata* are available in Supplementary Data 2.

In order to compare the functional annotation of *P. calceolata* with other small free-living photosynthetic eukaryotes, we applied the same analysis to the predicted proteins available for the following species: *Aureococcus anophagefferens*, *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, *Nannochloropsis oceanica*, *Bathycoccus prasinos*, *Micromonas pusilla*, *Ostreococcus lucimarinus*, and *Emiliania huxleyi* (references are indicated in Table 1).

We defined a list of 23 meiosis-specific genes using three previous studies<sup>31-33</sup>. KO annotations and Interproscan domains were used to recover these genes in the *P. calceolata* genome. Transcriptomic reads of *P. calceolata* were mapped back to the predicted genes with BWA-MEM v 2.2.1. Reads aligned over more than 80% of their length were retained to estimate the relative expression levels of meiosis genes.

Genes containing the NIT-sensing domain were identified based on Interproscan annotations. Eukaryotic homologs of the three *P. calceolata* NIT-sensing genes were retrieved with a BLASTP (*e*-value < 10<sup>-5</sup>, coverage > 100 aa) against 27.7 million proteins from NR, the METdb<sup>72</sup> transcriptome database, eukaryotic algal proteomes from the JGI database, *Tara* Oceans single-cell amplified genomes and metagenome assembled genomes (SMAGs)<sup>82</sup>. The 60 retrieved proteins and the 3 *P. calceolata* NIT-domain-containing proteins were then aligned with Mafft v3.0 (<https://mafft.cbrc.jp/alignment/server/>) and a Maximum Likelihood phylogenetic tree (Jones-Taylor-Thornton substitution model and 100 bootstraps) was made with MEGAX software. Transmembrane regions in NIT-sensing domain-containing proteins were identified with TMHMM v 2.055.

**Estimation of *P. calceolata* relative abundance in environmental metagenomic reads.** We used metagenomic datasets of *Tara* Oceans and *Tara Polar Circle* expeditions to detect and estimate the relative abundance of the *P. calceolata* in the oceans. Datasets from water samples collected on the photic zone: surface (SUR) and deep-chlorophyll maximum (DCM) were analyzed. Size-fractionated water samples containing pico-nano algae (organisms < 5 µm in cell diameter) were selected: 0.2–3 µm (100 samples), 0.8–2000 µm (119 samples) and 0.8–5 µm (148 samples)<sup>51</sup>. Metagenomic reads were aligned on the *P. calceolata* genome assembled in this study with BWA-mem version 0.7.15 with default parameters<sup>83</sup>. Alignments with at least 90% of identity over 80% of read length were retained for further analysis. In the case of several possible best matches, a random one was picked. In order to remove putative PCR duplicates, multiple read pairs aligned at the same position on the *P. calceolata* genome were removed with samtools rmdup version 1.10.2<sup>70</sup>. For the metagenomic abundance, we divided the total number of reads aligned on the *P. calceolata* genome assembled in this study by the total number of sequenced reads for each sample. For the metabarcoding abundance, we used the 18S V9 rRNA OTU table published in 2019 and available here <https://zenodo.org/record/3768510#YEX2S9zjjaQ34>. Bacterial and archaea OTUs were removed for this analysis.

***P. calceolata* relative abundance models.** Two *P. calceolata* relative abundance models were performed based on the in situ environmental conditions measured at the time of sampling or using the WOA18 datasets. The environmental parameters measured during the expedition are available in the Pangaea database (<https://www.pangaea.de/>) and are described in ref. <sup>84</sup>. Iron concentrations are annual means derived from PISCES2 model<sup>85</sup> and described in ref. <sup>13</sup>. Ammonium concentrations at the date and location of sampling are derived from the MITgcm Darwin model and available in the Pangaea database<sup>86</sup>. Environmental parameters for each sample are available in Supplementary Data 10. We consider oceanic samples as “low-iron” if they contain less than 0.2 nM of iron, “low-nitrate” if they

contain less than 2 µM of nitrate and “low-ammonium” if they contain less than 25 nM of ammonium. These thresholds were defined with the distribution of nutrient concentrations in the dataset and previous studies<sup>13,14</sup>. Pearson’s correlations between the relative abundance of *P. calceolata* and all environmental parameters were calculated with the *cor* function the R package FactoMineR version 2.4 and the GGally package version 2.1.0. Principal component analysis (PCA) was performed with 9 parameters presenting significant Pearson’s correlations. We used a Generalized Additive Model (GAM) for its ability to fit non-linear and non-monotonic functions and for its low sensitivity to extreme values to model the relative abundance of *P. calceolata* as a function of iron concentration, temperature and PAR light<sup>87</sup>. This function is implemented in the mgcv R package version 1.8–33.

The mean of several climatologies of the Earth System Models under the RCP8.5 scenario (ESM8.5) was used to define the environmental conditions at the end of the century following the method of ref. <sup>35</sup>. *P. calceolata* relative abundance models based on the WOA18 at locations, depths, and months of *Tara* Oceans samples or ESM8.5 were obtained using four machine-learning techniques as described in ref. <sup>35</sup>, with some differences. The four tools were trained in regression mode, we used the neuralnet R package<sup>88</sup> with the following parameters: hidden = 1, decay = 0.1, mxit = 1e<sup>6</sup> and size = 5, 6 splines were selected for the GAM. We performed 100-fold random cross-validation for each model and evaluated their performance using the Pearson correlation coefficient and rmse. We used the ensemble model approaches<sup>35</sup> for final global-scale models of the relative abundance of *P. calceolata* (i.e., the mean projections of the validated machine-learning techniques). Figures 2a, 3c, and Supplementary Fig. S7 were generated using the world map of R package “maps”, “mapdata” and “ggmap”. World maps data are imported from the public domain Natural Earth project.

**Environmental metatranscriptomic reads mapping and filtering.** Metatranscriptomic reads from *Tara* Oceans datasets were aligned on predicted coding sequences of the *P. calceolata* genome with BWA-mem 2.2.1 using default parameters. We selected reads aligned with more than 95% of identity over 80% of the read length. We kept nuclear genes covered by at least ten reads in a minimum of ten samples and removed those detected in more than 90% of samples which probably aggregate metagenomic reads from other organisms (cross-mapping). To obtain robust relative expression values, we removed samples where less than 25% of *P. calceolata* genes were detected. We finally got the expression of 15,679 genes of *P. calceolata* across 167 samples. In all gene expression analyses, we normalized the gene expression levels in transcripts per kilobase million (TPM). Gene expression levels of *P. calceolata* in all *Tara* Oceans samples are available here: <https://doi.org/10.5281/zenodo.6983365>.

**Statistical analysis and reproducibility.** Sample sizes are indicated in the methods, in the figure captions or in the main text. Nonparametric Wilcoxon signed-rank tests were applied with the two-sided alternative hypothesis and not paired. Two sides Fisher statistical tests were applied for gene enrichments analysis. All statistical tests in the manuscript were generated with R version 4.0.3 and *P* values < 0.01 are considered significant.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

*Pelagomonas calceolata* genomic and transcriptomic reads, genome assembly, and gene prediction are available at the ENA (EMBL-EBI) website under the accession number PRJEB47931. *P. calceolata* transcriptomes are available under the accession number PRJEB34158, runs ERR3497221 and ERR3497222. *Tara* Oceans and *Tara* Polar Circle metagenomic sequences are archived at the ENA under the following accession numbers: PRJEB9740, PRJEB9691, PRJEB4352, and PRJEB1787. All other data are available from the corresponding author upon reasonable request.

Received: 13 December 2021; Accepted: 2 September 2022;

Published online: 16 September 2022

## References

- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Boyce, D. G., Lewis, M. R. & Worm, B. Global phytoplankton decline over the past century. *Nature* **466**, 591–596 (2010).
- Henson, S. A., Cael, B. B., Allen, S. R. & Dutkiewicz, S. Future phytoplankton diversity in a changing climate. *Nat. Commun.* **12**, 5372 (2021).

4. Vaulot, D., Eikrem, W., Viprey, M. & Moreau, H. The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol. Rev.* **32**, 795–820 (2008).
5. Agawin, N. S. R., Duarte, C. M. & Agustí, S. Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnol. Oceanogr.* **45**, 591–600 (2000).
6. Morán, X. A. G., López-Urrutia, Á., Calvo-Díaz, A. & Li, W. K. W. Increasing importance of small phytoplankton in a warmer ocean. *Glob. Change Biol.* **16**, 1137–1144 (2010).
7. Li, W. K. W., McLaughlin, F. A., Lovejoy, C. & Carmack, E. C. Smallest algae thrive as the arctic ocean freshens. *Science* **326** <https://doi.org/10.1126/science.1179798> (2009).
8. Benner, I., Irwin, A. J. & Finkel, Z. V. Capacity of the common Arctic picoeukaryote *Micromonas* to adapt to a warming ocean. *Limnol. Oceanography Lett.* **5**, 221–227 (2020).
9. Sunda, W. G. & Huntsman, S. A. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.* **50**, 189–206 (1995).
10. Raven, J. A. The twelfth Tansley Lecture. Small is beautiful: the picophytoplankton. *Funct. Ecol.* **12**, 503–513 (1998).
11. Morel, F. M. M. & Price, N. M. The biogeochemical cycles of trace metals in the oceans. *Science* **300**, 944–947 (2003).
12. Gao, X., Bowler, C. & Kazamia, E. Iron metabolism strategies in diatoms. *J. Exp. Bot.* **72**, 2165–2180 (2021).
13. Caputi, L. et al. Community-level responses to iron availability in open ocean plankton ecosystems. *Glob. Biogeochemical Cycles* **33**, 391–419 (2019).
14. Carradec, Q. et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **9**, 373 (2018).
15. Morrissey, J. et al. A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr. Biol.* **25**, 364–371 (2015).
16. Moore, C. M. et al. Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* **6**, 701–710 (2013).
17. Kumar, A. & Bera, S. Revisiting nitrogen utilization in algae: a review on the process of regulation and assimilation. *Bioresour. Technol. Rep.* **12**, 100584 (2020).
18. Smith, S. R. et al. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **10**, 4552 (2019).
19. Berg, G. M., Glibert, P. M., Lomas, M. W. & Burford, M. A. Organic nitrogen uptake and growth by the chrysophyte *Aureococcus anophagefferens* during a brown tide event. *Mar. Biol.* **129**, 377–387 (1997).
20. Andersen, R. A., Saunders, G. W., Paskind, M. P. & Sexton, J. P. Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the pelagophyceae classis nov. *J. Phycol.* **29**, 701–715 (1993).
21. Choi, C. J. et al. Seasonal and geographical transitions in eukaryotic phytoplankton community structure in the Atlantic and Pacific Oceans. *Front. Microbiol.* **11**, 542372 (2020).
22. Duerschlag, J. et al. Niche partitioning by photosynthetic plankton as a driver of CO<sub>2</sub>-fixation across the oligotrophic South Pacific Subtropical Ocean. *ISME J* 1–12 <https://doi.org/10.1038/s41396-021-01072-z> (2021).
23. Worden, A. Z. et al. Global distribution of a wild alga revealed by targeted metagenomics. *Curr. Biol.* **22**, R675–R677 (2012).
24. Dimier, C. é, Brunet, C., Geider, R. & Raven, J. Growth and photoregulation dynamics of the picoeukaryote *Pelagomonas calceolata* in fluctuating light. *Limnol. Oceanogr.* **54**, 823–836 (2009).
25. Dupont, C. L. et al. Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J* **9**, 1076–1092 (2015).
26. Kang, Y. et al. Transcriptomic responses of four pelagophytes to nutrient (N, P) and light stress. *Front. Mar. Sci.* **8**, 636699 (2021).
27. Huff, J. T., Zilberman, D. & Roy, S. W. Mechanism for DNA transposons to generate introns on genomic scales. *Nature* **538**, 533–536 (2016).
28. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
29. Nambiar, M. & Smith, G. R. Repression of harmful meiotic recombination in centromeric regions. *Semin Cell Dev. Biol.* **54**, 188–197 (2016).
30. Pessia, E. et al. Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biol. Evol.* **4**, 675–682 (2012).
31. Chi, J., Mahé, F., Loidl, J., Logsdon, J. & Dunthorn, M. Meiosis gene inventory of four ciliates reveals the prevalence of a synaptonemal complex-independent crossover pathway. *Mol. Biol. Evol.* **31**, 660–672 (2014).
32. Ramesh, M. A., Malik, S.-B. & Logsdon, J. M. A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr. Biol.* **15**, 185–191 (2005).
33. Schurko, A. M. & Logsdon, J. M. Using a meiosis detection toolkit to investigate ancient asexual ‘scandals’ and the evolution of sex. *Bioessays* **30**, 579–589 (2008).
34. Ibarbalz, F. M. et al. Global trends in marine plankton diversity across kingdoms of life. *Cell* **179**, 1084–1097.e21 (2019).
35. Frémont, P. et al. Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat. Clim. Chang.* **12**, 393–401 (2022).
36. Ward, D. M. & Kaplan, J. Ferroportin-mediated iron transport: expression and regulation. *Biochim Biophys. Acta* **1823**, 1426–1433 (2012).
37. Gobler, C. J., Lonsdale, D. J. & Boyer, G. L. A review of the causes, effects, and potential management of harmful brown tide blooms caused by *Aureococcus anophagefferens* (Hargraves et sieburth). *Estuaries* **28**, 726–749 (2005).
38. Agustí, S., Lubián, L. M., Moreno-Ostos, E., Estrada, M. & Duarte, C. M. Projected changes in photosynthetic picoplankton in a warmer subtropical ocean. *Front. Mar. Sci.* **5**, 506 (2019).
39. Anderson, S. I., Barton, A. D., Clayton, S., Dutkiewicz, S. & Rynearson, T. A. Marine phytoplankton functional types exhibit diverse responses to thermal change. *Nat. Commun.* **12**, 6413 (2021).
40. Martin, J. H. et al. Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**, 123–129 (1994).
41. Shi, D., Xu, Y., Hopkinson, B. M. & Morel, F. M. M. Effect of ocean acidification on iron availability to marine phytoplankton. *Science* **327**, 676–679 (2010).
42. McQuaid, J. B. et al. Carbonate-sensitive phytoferritin controls high-affinity iron uptake in diatoms. *Nature* **555**, 534–537 (2018).
43. Turnšek, J. et al. Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway. *eLife* **10**, e52770 (2021).
44. Urzica, E. I. et al. Systems and trans-system level analysis identifies conserved iron deficiency responses in the plant lineage[W][OA]. *Plant Cell* **24**, 3921–3948 (2012).
45. Mao, X. et al. Diversity, prevalence, and expression of cyanase genes (cynS) in planktonic marine microorganisms. *ISME J.* **16**, 602–605 (2022).
46. Ou, L., Cai, Y., Jin, W., Wang, Z. & Lu, S. Understanding the nitrogen uptake and assimilation of the Chinese strain of *Aureococcus anophagefferens* (Pelagophyceae). *Algal Res.* **34**, 182–190 (2018).
47. Shu, C. J., Ulrich, L. E. & Zhulin, I. B. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. *Trends Biochem. Sci.* **28**, 121–124 (2003).
48. Wu, S. Q., Chai, W., Lin, J. T. & Stewart, V. General nitrogen regulation of nitrate assimilation regulatory gene nasR expression in *Klebsiella oxytoca* M5al. *J. Bacteriol.* **181**, 7274–7284 (1999).
49. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
50. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
51. Alberti, A. et al. Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci. Data* **4**, 170093 (2017).
52. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
53. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* <https://doi.org/10.1101/gr.210641.116> (2016).
54. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
55. Vaser, R. & Šikić, M. Yet another de novo genome assembler. *BioRxiv.* <https://doi.org/10.1101/656306> (2019).
56. Liu, H. et al. SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte* **2021**, 1–9 (2021).
57. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
58. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
59. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
60. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**, 737–746 (2017).
61. Aury, J.-M. & Istace, B. Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads. *NAR Genomics Bioinform.* **3**, lqab034 (2021).
62. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
63. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
64. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker. <http://repeatmasker.org/> (2013).
65. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).

66. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
67. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
68. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
69. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
70. Li, H. et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
71. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).
72. Niang, G. et al. METdb: A genomic reference database for marine species. *F1000Research*, <https://doi.org/10.7490/f1000research.1118000.1> (2020).
73. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
74. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
75. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
76. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
77. Dubarry, M. et al. Gmove a tool for eukaryotic gene predictions using various evidences. *F1000Research*, <https://doi.org/10.7490/f1000research.1111735.1> (2016).
78. Sibbald, S. J., Lawton, M. & Archibald, J. M. Mitochondrial genome evolution in pelagophyte algae. *Genome Biol. Evol.* **13**, evab018 (2021).
79. Quevillon, E. et al. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).
80. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
81. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
82. Delmont, T. O. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics* **2**, 100123 (2022).
83. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
84. Pesant, S. et al. Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2**, 150023 (2015).
85. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Dev.* **8**, 2465–2513 (2015).
86. Clayton, S. et al. Biogeochemical versus ecological consequences of modeled ocean physics. *Biogeosciences* **14**, 2877–2889 (2017).
87. Ravindra, K., Rattan, P., Mor, S. & Aggarwal, A. N. Generalized additive models: building evidence of air pollution, climate change and human health. *Environ. Int.* **132**, 104987 (2019).
88. Günther, F. & Fritsch, S. neuralnet: training of neural networks. *R. J.* **2**, 30–38 (2010).
89. Gobler, C. J. et al. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc. Natl Acad. Sci. USA* **108**, 4352–4357 (2011).
90. Guo, L. et al. Genome assembly of *Nannochloropsis oceanica* provides evidence of host nucleus overthrow by the symbiont nucleus during speciation. *Commun. Biol.* **2**, 1–12 (2019).
91. Bowler, C. et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).
92. Armbrust, E. V. et al. The genome of the diatom thalassiosira pseudonana: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
93. Worden, A. Z. et al. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes micromonas. *Science* **324**, 268–272 (2009).
94. Palenik, B. et al. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *PNAS* **104**, 7705–7710 (2007).
95. Moreau, H. et al. Gene functionalities and genome structure in Bathycoccus prasinus reflect cellular specializations at the base of the green lineage. *Genome Biol.* **13**, R74 (2012).
96. Read, B. A. et al. Pan genome of the phytoplankton *Emiliania underpins* its global distribution. *Nature* **499**, 209–213 (2013).

## Acknowledgements

We thank the commitment of the following people who made this work possible: the Genoscope/CEA, the CNRS (in particular the Federation de Recherche R2022/Tara Oceans GO-SEE), Marie-José Garet-Delmas from the Roscoff Culture Collection for growing the RCC100 strain, Claude Scarpelli for support in high-performance computing and Linda Sperling for language editing. Computations were performed using the cobalt HPC machine. We acknowledge the financial support of FRANCE GENOMIQUE (ANR-10-INBS-09-08) and Oceanomics (ANR-11-BTBR-0008). We also thank the Tara Expedition Foundation and their partners for the organization of marine scientific expeditions (<http://oceans.taraexpeditions.org>). This article is contribution number 139 of Tara Oceans.

## Author contributions

S.R., C.B., and M.G. performed *P. calceolata* cultures and DNA/RNA extractions. A.A., E.P., and C.C. coordinated DNA/RNA sequencing. B.I, B.N., and J.M.A. did the assembly and annotation of the genome. M.C., S.M., Q.C., and J.M.A. carried out genomic analysis. N.G., E.F., and Q.C. analyzed environmental datasets. P.F. and O.J. worked on *P. calceolata* models. Q.C. and N.G. wrote the paper with a strong support of J.M.A. and P.W. C.B, L.B., and K.L. did the Hi-C library preparation, sequencing and analysis, respectively. All authors contributed to the manuscript preparation and approved the final version of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03939-z>.

**Correspondence** and requests for materials should be addressed to Quentin Carradec.

**Peer review information** *Communications Biology* thanks Senjie Lin, Ana Cabello and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Linn Hoffmann and Luke R. Grinham. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

# Chapitre 2 - Nitrogen metabolism in the picoalgae *Pelagomonas calceolata*: disentangling cyanate lyase function under nutrient stress

---

## Préambule au Chapitre 2

Dans ce deuxième chapitre, notre objectif est de comprendre les mécanismes d'acclimatation au manque d'azote de *P. calceolata*. En effet, nous avons vu dans notre première étude que cette espèce possédait un ensemble de gènes pouvant être impliqués dans la réponse transcriptomique au manque de nitrate, et qu'un certain nombre de ces gènes candidats présentaient une expression significativement différente selon la concentration en nitrate dans les échantillons.

Nous avons d'abord recherché les gènes différentiellement exprimés dans les échantillons environnementaux pauvres en nitrate par rapport aux échantillons environnementaux riches en nitrate. Nous avons pu identifier un petit nombre de gènes jouant un rôle dans l'acclimatation environnementale au manque d'azote.

Pour aller plus loin dans notre compréhension des mécanismes d'acclimatation aux déplétions en azote, nous avons cultivé *P. calceolata* dans différentes concentrations de nitrate et avec différentes sources d'azote (nitrate, ammonium, urée, cyanate), afin d'étudier les gènes différentiellement exprimés en réponse à ces conditions. La comparaison de ces résultats de transcriptomique avec ceux obtenus précédemment nous a permis d'obtenir une meilleure compréhension des mécanismes d'acclimatation aux déplétions en azote et particulièrement le rôle de la cyanate lyase. Cette enzyme, encore peu étudiée, permet à un certain nombre de procaryotes d'assimiler le cyanate comme source d'azote organique alternative. Nous montrons ici que son rôle chez *P. calceolata*, et potentiellement chez les PPE, pourrait être en réalité assez différent.

Dans cette partie, j'ai réalisé les analyses de métatranscriptomiques environnementales. J'ai également mené les expériences de mises en culture dans différentes sources et quantité d'azote, et j'ai réalisé l'analyse des transcriptomes obtenus.



## Nitrogen metabolism in the picoalgae *Pelagomonas calceolata*: disentangling cyanate lyase function under nutrient stress.

\*Nina Guérin, Choé Seyman, Adrien Thurotte, Céline Orvain, Laurie Bertrand, Priscillia Gourvil, Ian Probert, Benoit Vacherie, Élodie Brun, Ghislaine Magdelenat, Karine Labadie, Patrick Wincker, Quentin Carradec

\*To be completed/reordered

### 1 - Abstract

Among nitrogen sources compounds in the environment, cyanate is a potential important actor given the activity and prevalence of cyanate lyase genes in microalgae. However, the conditions in which this gene is activated and the actual capacities of microalgae to assimilate cyanate remain underexplored. Here we studied the nitrogen metabolism of the abundant and cosmopolite Ochrophyte *P. calceolata* (Stramenopiles) with environmental metatranscriptomes and culture experiments under different nitrogen sources and concentrations (nitrate, ammonium, urea and cyanate). We observed that in nitrate-poor oceanic regions, the cyanate lyase gene is one of the most differentially expressed gene, suggesting that cyanate is an important alternative for *P. calceolata* persistence in oligotrophic environments. In the lab, we confirmed that this gene is overexpressed in low-nitrate medium together with several genes involved in nitrate recycling from endogenous molecules (purines and amino acids). *P. calceolata* is able to grow on various nitrogen sources: nitrate, urea and cyanate but not ammonium. RNA sequencing of these cultures revealed that the cyanate lyase gene is surprisingly underexpressed in cyanate conditions indicating that this gene is not involved in extracellular cyanate catabolism to ammonia. Taken together, environmental datasets and lab experiments reveal that if the cyanate lyase is essential in nitrite-poor environments it's probably to reduce the toxicity of cyanate as a consequence of endogenous nitrogenous compound recycling rather than the use extracellular cyanate to produce nitrate.

### 2 - Introduction

Nitrogen is essential to many biological processes, such as photosynthesis, amino acid and nucleic acid biosynthesis, thus its bioavailability in the oceans impact the growth of primary producers (Moore et al. 2013). After fixation of  $N_2$  gas by diazotrophs prokaryotes, inorganic ammonium ( $NH_4^+$ ), nitrite ( $NO_2^-$ ) and nitrate ( $NO_3^-$ ) are the primary source of fixed nitrogen for phytoplankton (David A. Hutchins and Capone 2022). Nitrate is the most abundant nitrogen compound, whereas nitrite and ammonium are typically less abundant (Collos and Berges 2002). Nevertheless, ammonium is considered a preferred nitrogen source due to its lower energy cost during assimilation, since it requires less reductase steps, and thus less energy, trace metals and water molecules (John A. Raven, Wollenweber, and Handley 1992). While

the growth of many species can be stimulated by ammonium, the choice of the most suitable nitrogen source varies depending on the species (Raven, Wollenweber, and Handley 1993). In general, photosynthetic eukaryotes mostly uptake nitrate, and cyanobacteria mostly uptake ammonium (Berthelot et al. 2019; Isaji et al. 2022). Most oceanic surface waters are depleted in inorganic nitrogen compounds because of photosynthesis organisms' quick removal of nitrogen in the photic zone. Nutrients are mainly supplied to surface waters by upward vertical flux, but they remain limited except in polar waters, upwelling zones and near the shore where there are terrestrial or anthropogenic inputs of nutrients (Levitus et al. 1993). A consequence of ocean warming is an enhanced stratification reducing the supply of nutrients to the euphotic zone (Fu, Randerson, and Moore 2016). At the end of the twenty-first century, a decrease of nitrate concentration of  $1.06 \pm 0.45 \text{ mmol.m}^{-3}$  on average of surface water is projected under the IPCC high-emission scenario SSP5-8.5 (Kwiatkowski et al. 2020).

Since inorganic nitrogen sources concentration can vary, phytoplankton developed a large set of strategies to face temporal and spatial changes in nitrate, nitrite and ammonium availability. It has been shown that the optimisation of inorganic nitrogen uptake can be achieved by regulating the expression of transporters in several phytoplankton groups, especially Prasinophytes and diatoms (Lampe et al. 2019). Storage and recycling of nitrogen-rich proteins and amino acids is also an important strategy, especially in diatoms (Scarsini et al. 2022; McCarthy et al. 2017; Alipanah et al. 2015).

Various phytoplankton can metabolise dissolved organic N (DON) such as urea, purines or amino acids (Kumar and Bera 2020). In the ocean, DON concentration is often higher than DINs and contributes to autotrophic production and growth of primary producers, especially in coastal and estuarine environments, which contain a large set of nitrogenous compounds such as urea, amino acids and nucleic acids (Sipler and Bronk 2015). All these compounds, especially urea and amino acids have been shown to support phytoplanktonic growth in low N-conditions. For example, urea can be uptake as a nitrogen source by many phytoplankton species, such as diatoms *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, *Chaetoceros didymus*, *Skeletonema costatum* and dinoflagellates *Karenia mikimotoi*, *Dinophysis acuminata* and *D. acuta* (Price and Harrison 1988; Cerón García et al. 2005; García-Portela et al. 2020; Meyer, Rydzyk, and Pohnert 2022). Growth of the harmful-bloom forming algae *Aureococcus anophagefferens* can be supported by urea (Berg et al. 1997). Diatoms such as *P. tricornutum* and *T. pseudonana* present urea transporters that are overexpressed in N-limited conditions (Smith et al. 2019; S. Bender et al. 2014). *Emiliania huxleyi* can use a broader range of organic nitrogen source, including urea, hydroxyurea, hydroxanthine, purines and small amides such as acetamides and formamides (Palenik and Henson 1997).

Cyanate ion ( $\text{OCN}^-$ ) is the smallest nitrogenous organic compound recently described as a potential important source of N for microorganism. In the oceans, cyanate comes

from different sources: terrestrial inputs, spontaneous decomposition of carbamoyl-phosphate and urea released by zooplankton or senescent diatom cells, and photoproduction from organic matter (Kamennaya, Chernihovsky, and Post 2008; Widner, Mulholland, and Mopper 2016). Cyanate is measured up to 45 nM in ocean subsurface (Widner, Mordy, and Mulholland 2018). However, organisms actually capable of cyanate uptake and metabolism in the environment as well as the underlying molecular mechanisms remain unclear. Growth on cyanate as sole N source was first described in *E. coli* then in cyanobacteria from genus *Synechococcus* and *Prochlorococcus*, in the ammonia-oxidizing Thaumarchaeota *Nitrososphaera gargensis*, in some ascomycetes yeasts, and in the dinophyceae *Prorocentrum* (Anderson, Sung, and Fuchs 1990; Kamennaya and Post 2011; Hu et al. 2012; 2014; Palatinszky et al. 2015; Linder 2018). In bacteria, the cyanate lyase enzyme (cynS gene) catalyses the bicarbonate dependant reaction of cyanate to ammonia and carbon dioxide (Johnson and Anderson 1987). Cyanate lyase homologs are present in the genome of many organisms including the dominant eukaryotic phytoplankton lineages (Mao et al. 2022). Cyanate lyase could have a role in the detoxification of cyanate resulting from intracellular urea and/or carbamoyl phosphate decomposition or in the production of CO<sub>2</sub> to support Calvin cycle (Guilloton and Karst 1987; Miller and Espie 1994). In the environment, the expression of cyanate lyase in most prokaryotic and eukaryotic phytoplankton is increased in N-limited environments, suggesting that these organisms might use cyanate as alternative N source (Wurch et al. 2011; Dong et al. 2014; Smith et al. 2019; Mao et al. 2022). <sup>15</sup>N stable isotope probing revealed that cyanate uptake could account for up to 10% of total nitrogen uptake in natural communities from offshore oligotrophic Atlantic, particularly in surface waters (Widner, Mulholland, and Mopper 2016). Since cyanate lyase expression is equally important as urease in the environment, it can be presumed to have a very crucial ecological role (Mao et al. 2022). However, uptake of cyanate has not been demonstrated in eukaryotic phytoplankton yet. A recent study have shown that cyanate enrichment in natural phytoplankton populations induces *Synechococcus* growth but not eukaryotic phytoplankton (Sato, Hashihama, and Takeda 2023). This leads us to question the role of cyanate lyase in organic nitrogen assimilation in photosynthetic eukaryotes.

In order to disentangle the endogenous role of cyanate lyase in microeukaryotes and more globally have a better understanding of eukaryotic phytoplankton acclimation to changing N availability in the environment, we used the cosmopolite photosynthetic pico-eukaryote *Pelagomonas calceolata* (Pelagophyceae) as a model for pelagic phytoplankton (Worden et al. 2012; Guérin et al. 2022). *P. calceolata* RCC100 strain genome contains a large set of genes involved in nitrogen metabolism (Guérin et al. 2022). Among them, genes coding for arginase, urease and cyanate lyase may suggest the capacity of organic nitrogen compounds metabolism. In the environment, several nitrogen ions transporters, nitrate and nitrite reductases, glutamine synthetase, nitrate/nitrite sensing proteins and the cyanate lyase are significantly expressed at higher levels in low-N environments (Dupont et al. 2015; Guérin et al. 2022).

Here we, cultivated *P. calceolata* under high or limited nitrate conditions and tested several inorganic (nitrate, ammonia) and organic (urea, cyanate) sources of nitrogen on *P. calceolata* growth and physiology. For all these conditions, we performed RNA sequencing and identified differentially expressed genes. We compared our results with a transcriptome-wide differential analysis of *P. calceolata* gene expression levels in environmental metatranscriptomes of *Tara* Oceans expedition according to *in situ* measurements of nitrate concentrations.

### 3 - Results

#### ***In situ* gene expression levels of *P. calceolata* according to nitrate concentration**

To study the transcriptomic regulation of the nitrogen metabolism of *P. calceolata in situ*, we used all metatranscriptomes of *Tara* Oceans datasets (Alberti et al. 2017). We aligned metatranscriptomic reads on the predicted mRNAs of *P. calceolata* RCC100 genome and selected 124 samples where at least 75% of the genes are. For 112 of these samples with available *in situ* measurements of nitrate concentration, 69 have nitrate concentrations below 2  $\mu\text{M}$  and are considered "low-nitrate", and 43 have nitrate concentrations above 2  $\mu\text{M}$  and are considered "high-nitrate" (Figure 1A).

Differential expression analysis between "high-nitrate" and "low-nitrate" environments reveal 375 genes significantly overexpressed in low-nitrate samples ( $p\text{-value} < 0.01$ ) and 6 of them with a log2 fold change higher than 2 (Figure 1B, Supplementary Table 1). Three of these genes are involved in nitrogen compound transport across cell membranes: a formate or nitrite transporter (carrying PF01226 protein domain) and an ammonium or urea transporter (PF00909 domain) as previously described (Guérin et al. 2022). In addition, an uric acid or xanthine permease (K23887 domain) and a formyltetrahydrofolate deformylase (10-FDF; EC 3.5.1.10), enzyme potentially related to ammonium recycling from glycine and an unknown gene carrying a RNA-binding domain of the Pumilio family (K17943) are also overexpressed in low-nitrate environments. Finally, the cyanate lyase (PF02560 domain) is one of the most overexpressed genes in nitrate-low environments.

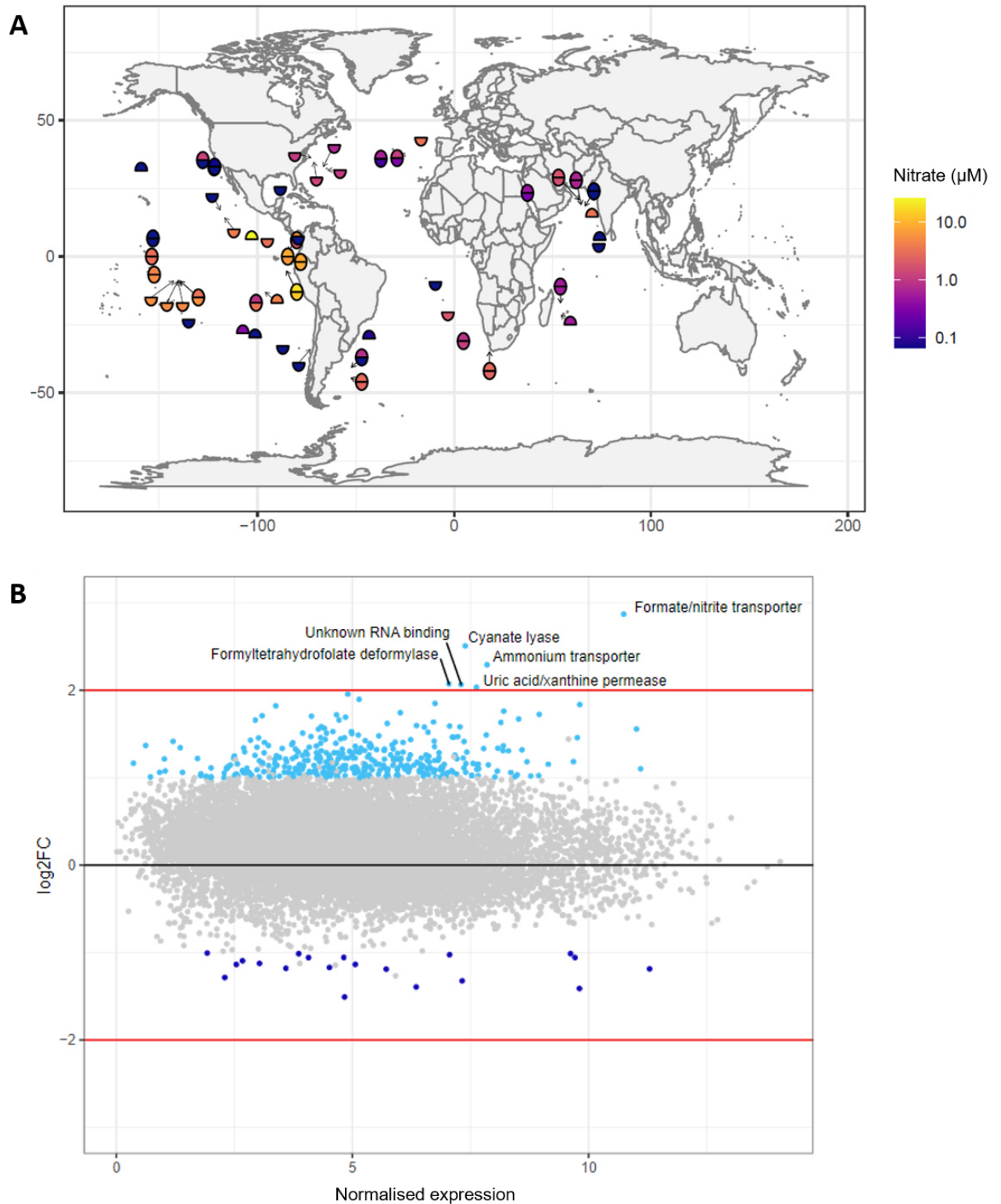


Figure 1. *Pelagomonas calceolata* response to environmental nitrate concentrations. A) Nitrate concentration across Tara Oceans samples that were incorporated in this analysis. B) Differentially expressed genes between low-nitrate environmental samples ( $\text{NO}_3^- < 2 \mu\text{M}$ ,  $n=69$ ) and high-nitrate environmental samples ( $\text{NO}_3^- > 2 \mu\text{M}$ ,  $n=43$ ). Log<sub>2</sub> fold change between low- and high-nitrate samples of each gene according to their mean normalised expression level. Genes with  $p$ -value  $< 0.01$  and  $\log_2\text{FC} > 1$  or  $< -1$  are coloured in blue. Data are available in Supplementary Table 1.

## Growth and gene expression variations under low-N conditions

*P. calceolata* was incubated with high-nitrate (882  $\mu\text{M}$ ), intermediate-nitrate (441  $\mu\text{M}$ ) or low-nitrate (220  $\mu\text{M}$ ) concentrations for a minimum of 3 weeks with weekly transfer on fresh media. Reduction of  $\text{NO}_3^-$  concentration affects *P. calceolata* maximal cell density (Figure 2A). Cells were collected by filtration in exponential growth phase for high- and intermediate- $\text{NO}_3^-$  concentration and at the beginning of the plateau for 220  $\mu\text{M}$ . After polyA+ RNA extraction and sequencing, the 9 samples had between 49 million and 68 million paired-end reads that were used for the alignment on *P. calceolata* genome (RCC100 strain). 16 659 out of 16 667 *P. calceolata* genes were detected (at least one read in one sample). For high, intermediate and low-nitrate conditions, correlations of gene expression within triplicates are higher (Pearson's  $r$  from 0.90 to 0.98) than correlations between conditions Pearson's  $r$  from 0.59 to 0.81) (Supplementary Figure 1).

Differential expression analysis was performed between the high-nitrate condition and each tested condition (Supplementary Table 2, Supplementary Table 3). In the intermediate nitrate condition, we identified 1 282 differentially expressed genes compared to the standard condition. There are 482 genes overexpressed when the nitrate concentration is divided by two compared to the standard. These genes mostly involved in amino acid catabolism and diverse types of transporters. The most differentially expressed genes with known domains are coding for sodium/solute transport (PF00474), taurine dioxygenase (PF02668), S-adenosyl-L-methionine-dependent (SAM) methyltransferase (IPR029063), tryptophan/tyrosine permease (PF03222), and carbon-nitrogen hydrolase domain (PF00795) (Figure 2B). The 800 genes underexpressed in intermediate nitrate concentrations are mostly involved in initiation, elongation and translation factor and in ribosomal proteins. The most differentially expressed genes are aminotransferase (PF00155), carbamoyl-phosphate synthetase, chaperonin DnaJ (K09522), polyadenylate-binding protein (K13126), chaperonin GroEL (K04077), chlorophyll A/B binding protein (PF00504), uridine kinase (K00876-EC:2.7.1.48), cytochrome b5 (PF00173) and heat shock protein 70kD (PF00012) (Figure 2B).

In the low- nitrate condition, 799 genes are differentially expressed compared to the standard condition. 230 genes are overexpressed and 569 are underexpressed in the low-nitrate condition. Genes overexpressed in low-nitrate are enriched in amino acid catabolism and diverse types of transporters. The most expressed genes are sodium/solute symporter (PF00474), taurine dioxygenase (PF02668), tryptophan/tyrosine permease (PF03222), SAM methyltransferase (IPR029063), xanthine permease (K23887) (Figure 2C). Genes underexpressed in low-nitrate are coding for translation, initiation and elongation factor and ribosomal proteins. The most differentially expressed genes are coding for aminotransferase (PF00155), uridine kinase (K00876-EC:2.7.1.48), ATPase family (K17681) and LL-diaminopimelate aminotransferase (K10206-EC:2.6.1.83 ) (Figure 2C).

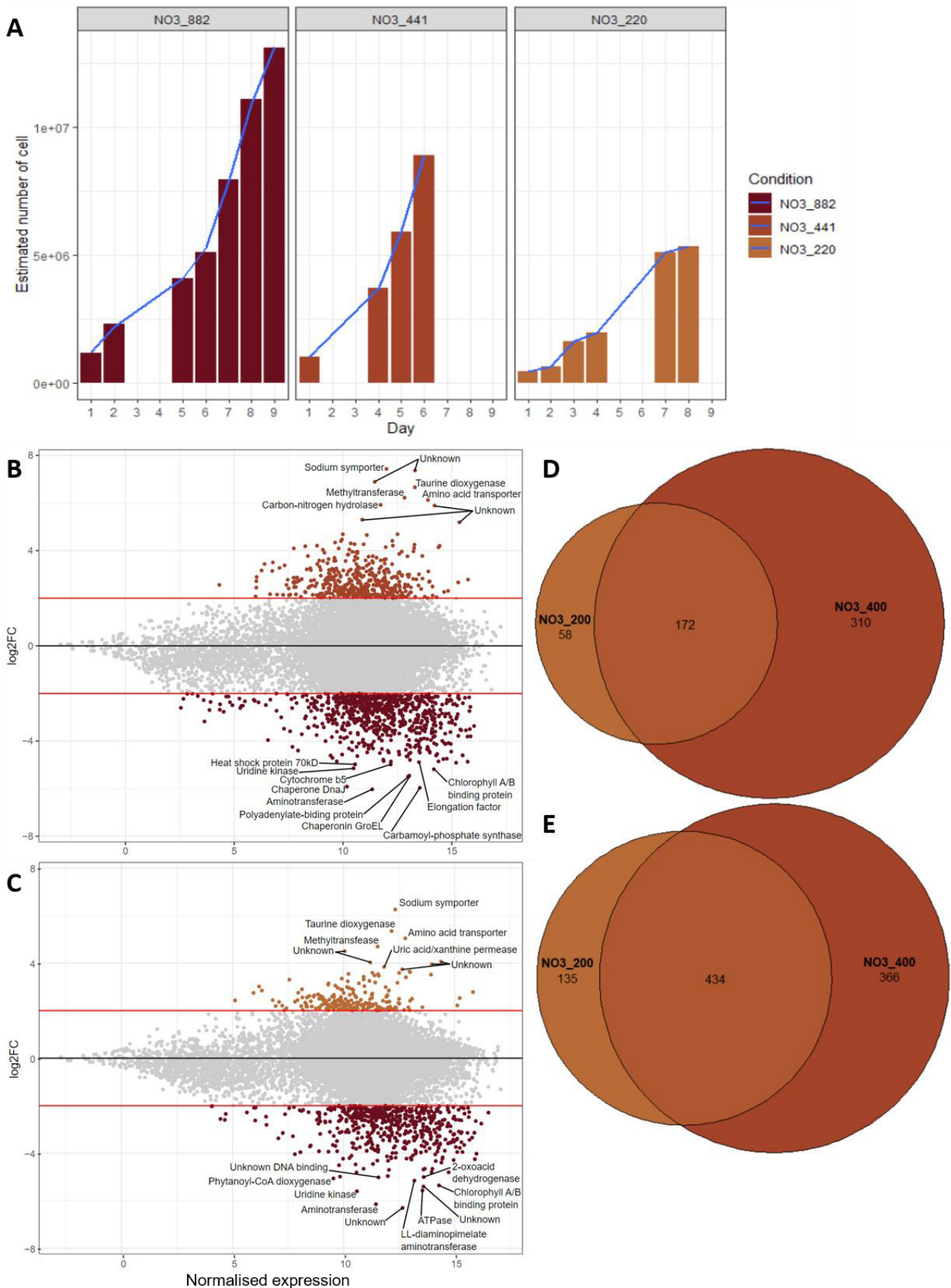


Figure 2. *Pelagomonas calceolata* growth and transcriptomic response to low-nitrate culture conditions. A) Growth of *P. calceolata* under different nitrate concentrations was measured by fluorescence and then converted into an estimated number of cells. B, C) Differentially expressed genes in 441  $\mu$ M (B) and 220  $\mu$ M (C) of nitrate compared to 882  $\mu$ M of nitrate. Genes with p-value < 0.01 and log2fold change > 2 are coloured. Data are available in Supplementary Tables 2 and 3. D, E) Euler diagram of genes overexpressed (D) or underexpressed (E) in 441  $\mu$ M, 220  $\mu$ M of nitrate or both.

Intermediate and low-nitrate experiments have in common 604 DEG: 172 overexpressed and 434 underexpressed in both 441  $\mu$ M and 220  $\mu$ M compared to the high-nitrate condition (Figure 2D, 2E).

### ***P. calceolata* gene expression levels cultivated with different nitrogen compounds.**

*P. calceolata* was incubated with nitrate (882  $\mu$ M), ammonium (882  $\mu$ M), urea (441  $\mu$ M) or cyanate (882  $\mu$ M) for a minimum of 3 weeks with weekly transfer on fresh media. We observed similar cell density for nitrate, cyanate and urea in exponential growth phase (Figure 3A). This results shows that *P. calceolata* is able to grow with cyanate or urea as sole source of nitrogen. The maximum density reached by *P. calceolata* is strongly limited on ammonium (Figure 3A). After polyA<sup>+</sup> RNA extraction and sequencing, the 12 samples had between 45 million and 84 million paired-end reads that were used for the alignment on *P. calceolata* genome (RCC100 strain). 16 659 out of 16 667 *P. calceolata* genes were detected (at least one read in one sample). For those four conditions, triplicates are highly correlated with Pearson's *r* correlations above 0.98. *P. calceolata* transcriptome under nitrate condition is more correlated with the cyanate and the ammonium conditions (Pearson's *r* from 0.96 to 0.95) than with the urea conditions (Pearson's *r* from 0.73 to 0.82) (Supplementary Figure 1).

In the ammonium condition, despite a growth curve strongly impacted, only 71 genes are underexpressed and 45 are overexpressed compared to the nitrate condition (Supplementary Table 4). The most underexpressed genes are genes coding for genes involved in nitrate assimilation such as nitrite reductase (K00362-EC:1.7.1.15), formate/nitrite transporter (PF01226) and nitrate/nitrite transporter (K02575), some are protein kinase (PF00069). The most overexpressed genes have domain of ammonium transporter (K03320), glycoside hydrolase (PF00722), heat shock type protein with DNA binding (PF00447) and some specialised motifs such as uncharacterised protein (K07126) and EF hand motif (PF13499) (Figure 3 B).

On the contrary, *P. calceolata* presents a good acclimation to cyanate and urea and many differentially expressed genes. In the urea condition, 921 genes are differentially expressed compared to the nitrate condition. 281 genes are underexpressed and 640 are overexpressed when the nitrate is replaced by urea (Supplementary Table 5). The most differentially underexpressed genes are involved in nitrate assimilation such as nitrite reductase (K00362-EC:1.7.1.15), and nitrate/nitrite transporter (K02575). There is also genes coding for chlorophyll A/B binding protein (PF00504) and for protein kinase (PF00069). The most differentially overexpressed genes are carrying specialised repeat sequences and motifs such as ankyrin repeats domain (PF12796), Macro and CHAT domain (PF01661, PF12770), IQ calmodulin-binding motif (PF00612) and EF hand domain (PF13202) (Figure 3 C).

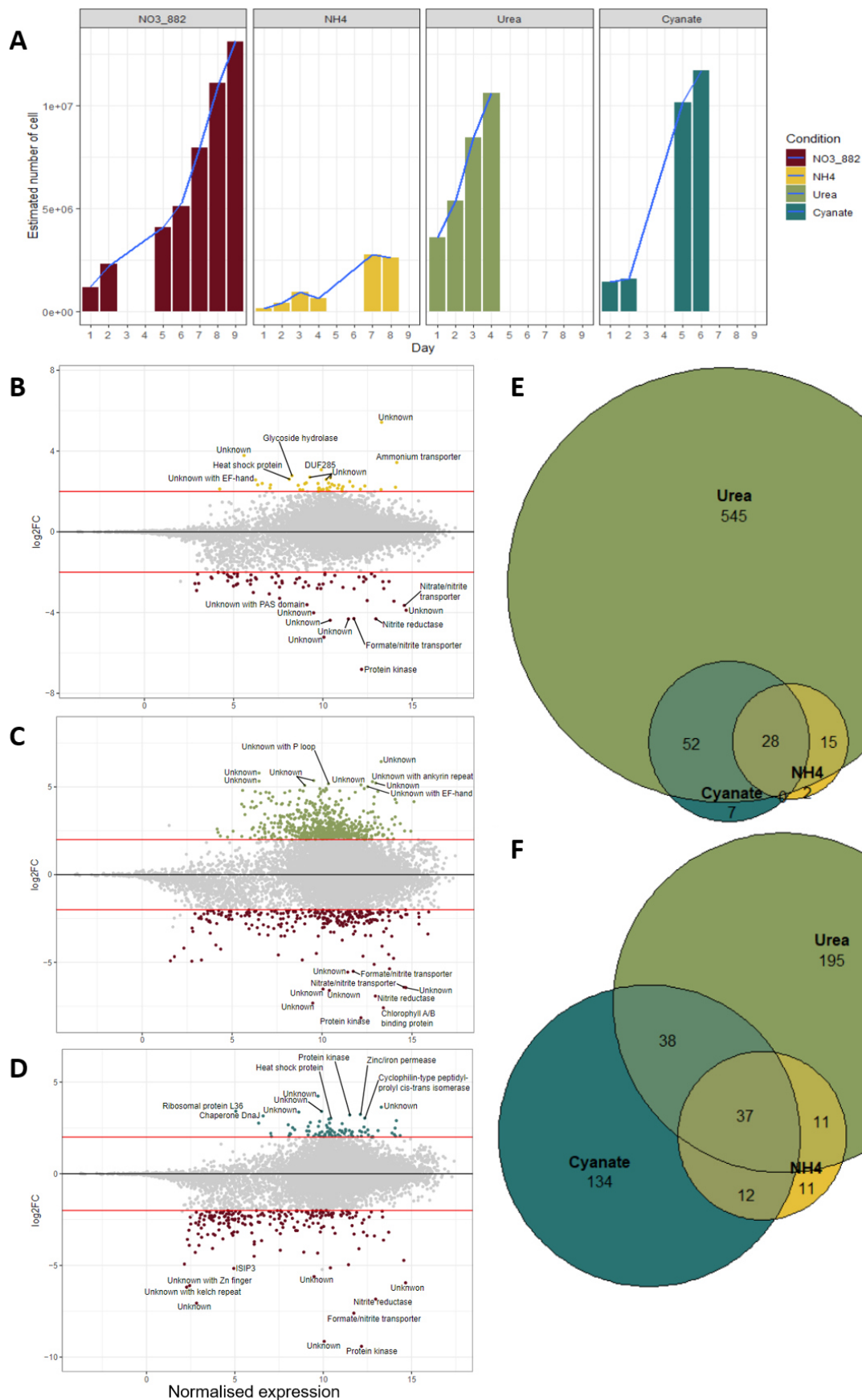


Figure 3. Growth and transcriptomic response of *P. calceolata* cultivated with different nitrogen compounds. A) Growth of *P. calceolata* under various nitrogen sources (882  $\mu\text{M}$   $\text{NO}_3^-$ , 882  $\mu\text{M}$   $\text{NH}_4^+$ , 441  $\mu\text{M}$  urea, 882  $\mu\text{M}$  cyanate) measured by fluorescence converted into estimated number of cell. B, C, D) Differentially expressed genes in ammonium (B), urea (C) and cyanate (D) compared to nitrate. Genes with p-value < 0.01 and log<sub>2</sub>fold change > 2 are coloured. Data are available in Supplementary Tables 4, 5 and 6. E, F) Euler diagram of genes overexpressed (E) or underexpressed (F) in at least one of the alternative nitrogen sources.

In the cyanate condition, 221 genes are underexpressed and 87 are overexpressed when the nitrate is replaced by cyanate (Supplementary Table 6). The most differentially underexpressed genes contain domains for nitrate assimilation such as formate/nitrite transporter (PF01226) and nitrite reductase (K00362-EC:1.7.1.15), there is also protein kinase (PF00069), and specialised motifs and repeats Kelch motif (PF01344) and tetratricopeptide repeat (PF13374). The most differentially overexpressed genes are involved in protein biosynthesis and folding, such as ribosomal protein L36 (PF00444), a cyclophilin type peptidyl-prolyl cis-trans isomerase (PF00160), a DnaJ chaperon (PF00226) and a heat shock protein 20 (IPR008978), along with a zinc/iron permease (PF02535), a protein kinase (PF00069) and the uncharacterised protein K07126 (Figure 3D).

Alternative nitrogen source experiments have 65 DEG in common compared to the nitrate condition, 28 are overexpressed and 37 are underexpressed in alternative nitrogen source (Figure 3E, 3F).

## **Nitrogenous compound transport**

In each family of nitrogen compound transporters, we observed genes with important variations of expression when the nitrogen source or concentration is modified (Figure 4, Table 1). Among three high-affinity nitrate transporters (NRT2, KO2575), two are downregulated when nitrate is replaced by another N source, but they are not regulated with the decrease of nitrate concentration. Two out of 3 formate/nitrite transporter (FNT, IPR000292) are underexpressed in all alternative N source and one of them is overexpressed in low-nitrate conditions *in situ* and in lab culture. Among the five ammonium transporters (Amt), one gene is overexpressed in low-nitrate conditions *in situ* and slightly overexpressed in the intermediate-NO<sub>3</sub> condition. This gene is also downregulated when the nitrate is replaced by another nitrogen source. This gene might be high-affinity transporters while three other Amt genes are slightly underexpressed when nitrate is depleted, and might be low-affinity transporters. Interestingly, the last low-affinity transporter is overexpressed in urea, cyanate and ammonium conditions suggesting that this Amt is regulated according to the presence or absence of nitrate in the environment. A single urea transporter (IPR004937) is described in *P. calceolata* genome but is not differentially regulated in any of our tested conditions.

Gene	Gene name	Function	Environmental low nitrate	Intermediate	Low	Ammonium	Urea	Cyanate
Pca_2p10630	NRT2	Nitrate/nitrite transporter				-1.30	-1.57	
Pca_2p27760	NRT2	Nitrate/nitrite transporter	1.56			-3.65	-6.42	-4.73
Pca_2p31470	NRT2	Nitrate/nitrite transporter	1.17					
Pca_2p00350	FNT	Formate/nitrite transporter		1.08				
Pca_4p09810	FNT	Formate/nitrite transporter	2.87	2.78	2.78	-4.30	-5.51	-7.61
Pca_4p16880	FNT	Formate/nitrite transporter				-2.47	-2.59	-2.36
Pca_1p08820	AMT	Ammonium transporter	2.29	1.09		-2.48	-3.27	-2.28
Pca_2p16510	AMT	Ammonium transporter		-3.43				
Pca_5p16370	AMT	Ammonium transporter			-1.28	-1.21	-1.53	-1.19
Pca_6p00630	AMT	Ammonium transporter						
Pca_6p05940	AMT	Ammonium transporter		-1.56		3.43	4.10	2.90
Pca_4p23330	UT	Urea transporter						
Pca_2p02240	NirB-NirD	Nitrite reductase	1.20			-4.31	-6.92	-6.84
Pca_4p19310	NirA	Ferredoxin nitrite reductase		-1.89		-2.47	-4.65	-3.18
Pca_1p15820	NR	Nitrate reductase				-2.00	-6.00	-6.00
Pecal_2p03910	Xan_ur_permease	Xanthin/ Uric acid permease	2.03	3.77	3.95		-2.34	-2.69
Pca_2p05120	UreD-UreF	urease accessory protein						
Pca_2p05150	Urease	urease			1.40	1.31	2.12	1.40
Pca_2p05090	UreG	urease accessory protein						
Pca_4p25220	UreG	urease accessory protein						
Pca_6p12900		arginase						
Pca_3p16260	CYN	cyanate lyase	2.51		1.85	-1.81	-3.23	-2.30
Pca_1p31590	NIT	Nitrate sensing				-1.92	-1.74	-1.84
Pca_2p03520	NIT	Nitrate sensing (transmembrane domains)		2.69	1.92			
Pca_6p09320	NIT	Nitrate sensing		-2.59				
Pca_1p07940	GSIII	Glutamine synthetase		-2.56	-1.45			
Pca_1p27820	GSIII	Glutamine synthetase	1.44	2.04	1.49			
Pca_2p07880	GSIII	Glutamine synthetase	1.39			-1.07		
Pca_2p12040	GS	Glutamine synthetase					-3.50	
Pca_5p23600	GS	Glutamine synthetase					-2.08	-1.57
Pca_1p08930	NADPH-GOGAT	Glutamate synthase						
Pca_2p12240	GOGAT	Glutamate synthase		1.38			2.35	
Pca_5p07200	Fd-GOGAT	Glutamate synthase		-1.54				
Pca_5p22300	NADPH-GOGAT	Glutamate synthase		-1.43	-1.96			
Pca_2p09950	CPS	carbamoyl-phosphate synthetase	1.02	-5.97	-3.05	-1.09	-3.26	-1.47
Pca_3p22990	CPS	carbamoyl-phosphate synthetase		-2.48			-1.06	
Pca_1p14490	OCT	Ornithine carbamoyltransferase						
Pca_3p14120	OCT	Ornithine carbamoyltransferase		-1.46	-2.53			
Pca_4p01540	ASL	Argininosuccinate lyase		-2.84	-3.49			
Pca_5p06610	ASS	Argininosuccinate synthase		-4.86			-1.94	
no gene found	XDH	Xanthine dehydrogenase/oxidase						
Pca_6p01310	hpxO	FAD-dependent urate hydroxylase						
Pca_6p01330	hpxO	FAD-dependent urate hydroxylase						
Pca_2p06780	uraH, pucM, hiuH,	5-hydroxyisourate hydrolase / 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazole decarboxylase						
Pca_2p06490	allB	allantoinase						
no gene found	alc	allantoicase						
Pca_6p05430	nodU	carbamoyltransferase						
Pca_6p05480	nodU	carbamoyltransferase		1.79				1.03
Pca_6p05520	nodU	carbamoyltransferase		1.90	1.68	1.14	1.87	1.46
Pca_6p05580	nodU	carbamoyltransferase						
Pca_6p08530	nodU	carbamoyltransferase	1.40					

Table 1. Differential expression of genes involved in nitrogen metabolism based on their functional annotation. Log2 fold changes are shown only when log2FC > 1 or < -1 and p-value < 0.01.

Finally, the uric-acid/xanthine permease overexpressed in low-nitrate environments *in situ* is also significantly overexpressed in low-nitrate cultures and is underexpressed when the nitrate is replaced by cyanate or urea. Its function might be the transport of purine degradation products such as uric acid, xanthine and vitamin C. This expression pattern suggests that this transporter under nitrate starvation recycles intracellular nitrogen compounds but do not transport extracellular organic nitrogen molecules. It was suggested that the overexpression of similar the *A. anophagefferens* xanthine permease under low-N conditions was a consequence of the conversion of xanthine into urea and ammonia (Gann et al. 2022). We search for the genes involved in this pathway in *P. calceolata*. We found the genes coding for almost all steps except for xanthine dehydrogenase/oxidase and the allantoicase (Table 1). None of them were differentially expressed except the carbamoyl-phosphate transferase that was down-

regulated in both low-nitrate and alternative nitrogen sources.

## Nitrate sensing

In Guérin et al., 2022, we identified three genes carrying a nitrate/nitrite sensing domain that we suspected to be involved in the nitrogen metabolism (Guérin et al. 2022). These three genes present distinct expression patterns according to nitrate concentration or source (Figure 4, Table 1). The nitrate/nitrite sensing gene carrying transmembrane domains is overexpressed in low-nitrate samples but not differentially expressed when the nitrate is replaced by another nitrogen source. This gene has been shown to be overexpressed by *P. calceolata* in low-nitrate environments (Dupont et al. 2015; Guérin et al. 2022) but is not significant in our environmental DESeq2 analysis (log2fold change >1 or <-1 and p-value>0.01). The two other nitrate-sensing genes carry a serine–threonine/tyrosine kinase domain, and might play a role in phosphorylation-based signal transduction. These two genes present two different expression patterns: the first one is slightly overexpressed under nitrate as nitrogen source while the second is underexpressed in intermediate-nitrate concentration. This suggests that each gene has a specific role to respond intracellular or extracellular nitrate concentrations.

## Nitrate reduction and storage

The nitrate reductase, the NADH-dependant nitrite reductase (NirB-NirD) and the ferredoxin nitrate reductase (NirA) have the same expression pattern with high gene expression levels when the nitrogen source is nitrate, and a strong downregulation under urea, cyanate and ammonium (Figure 4, Table 1). Overexpression under nitrate compared to other nitrogen sources is coherent with previous results on many algae including the diatom *P. tricornutum* and the Pelagophyte *A. anophagefferens* (Dong et al. 2014; Smith et al. 2019).

Four of the five genes coding for glutamine synthases (GS) in *P. calceolata* are differentially expressed in at least one experiment (Figure 4, Table 1). We do not know their cellular localisation. Two genes are underexpressed when the nitrogen source is urea, suggesting that the urea metabolism do not require these genes. Two others are differentially expressed according to the nitrate concentration in the lab with opposite patterns. Similarly to GS, the four *P. calceolata* glutamate synthase (GOGAT) could be localised in the chloroplast or in the mitochondria. Two of them (the putative Fd-GOGAT and one of the putative NADPH-GOGAT) are slightly overexpressed in high-nitrate samples. Reversely, the third GOGAT is slightly overexpressed in low-nitrate experiments and strongly when the nitrogen source is urea. Finally, the other NADH-GOGAT gene is not differentially expressed in our experiments (Figure 4, Table 1).

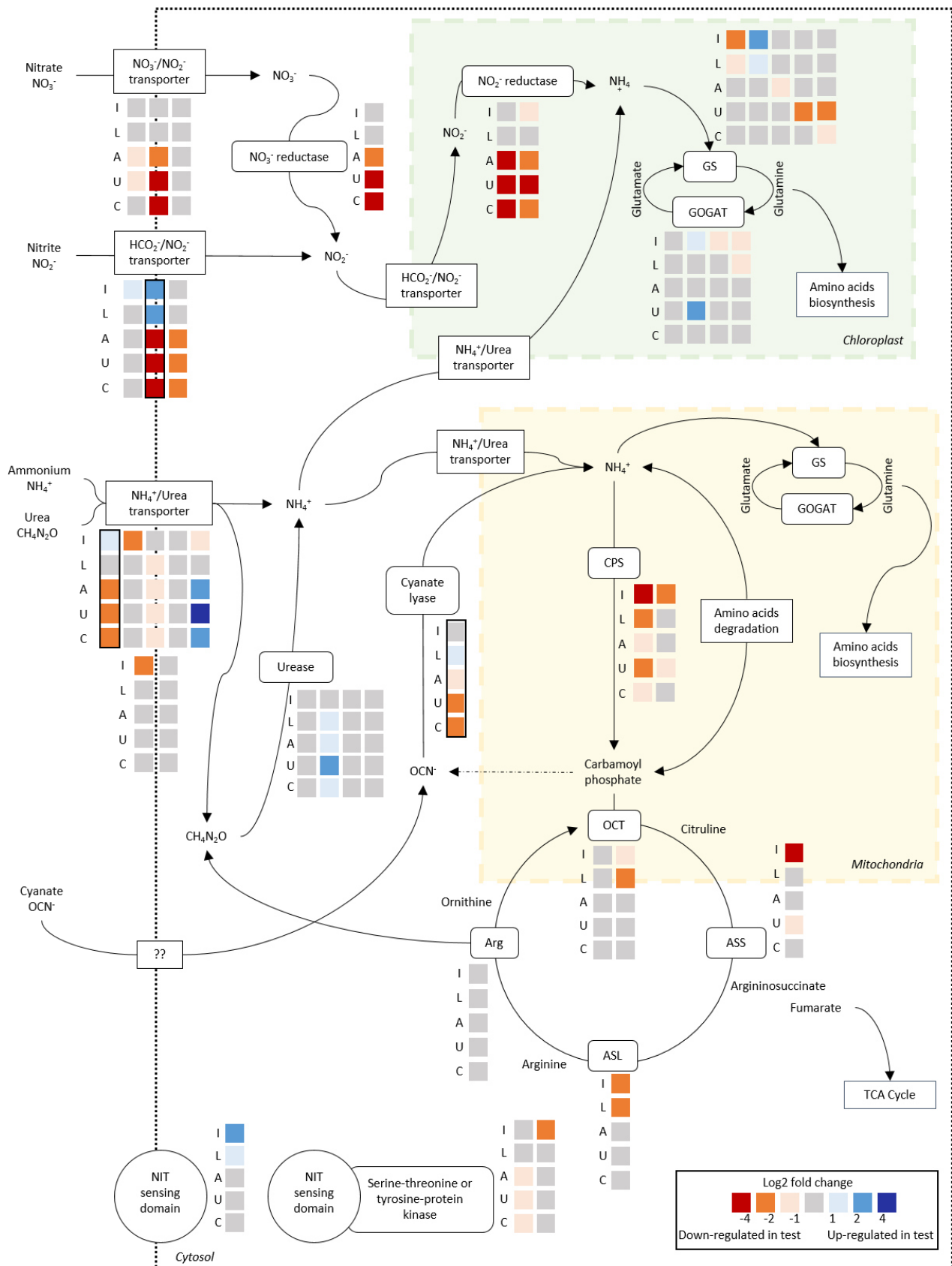


Figure 4. Impact of changing N source and quantity on N metabolism of *P. calceolata*. Each square represent the differential expression of one gene in one condition compared to the standard (882  $\mu\text{M}$   $\text{NO}_3^-$ ). Columns are the different genes, and rows are the different conditions. I: intermediate nitrate (441  $\mu\text{M}$ ); L: low nitrate (220  $\mu\text{M}$ ); A: ammonium; U: urea; C: cyanate. GS: glutamine synthetase; GOGAT: glutamate synthase; CPS: carbamoyl-phosphate synthetase; OCT: ornithine-carbamoyl transferase; ASS: arginosuccinate synthase; ASL: arginosuccinate lyase; Arg: arginine. The colour for the square indicates if the gene is underexpressed (red) or overexpressed (blue). Genes framed in black are overexpressed in environmental low-nitrate samples.

## Cyanate lyase, urease and urea cycle.

Consistently with low-nitrate environments, *P. calceolata* cyanate lyase gene is slightly overexpressed in our low-nitrate experiment (Figure 4, Table 1). Surprisingly, under alternative nitrogen source, including cyanate, the cyanate lyase is underexpressed. This result shows that the cyanate lyase is not involved in assimilation of extracellular cyanate and could instead be involved in nitrogen recycling from intracellular molecules (see discussion).

*P. calceolata* encodes one urease and three urease accessory proteins: one UreD-UreF and two UreG (Figure 4, Table 1). The urease is slightly overexpressed in low-nitrate samples and overexpressed in alternative nitrogen source, especially in urea. This result shows extracellular urea as well as intracellularly produced urea are hydrolysed into ammonia by the urease.

The urea cycle in *P. calceolata* is composed of two carbamoyl phosphate synthetases (CPS), 2 ornithine carbamoyltransferase (OCT), 1 arginosuccinate synthase (ASS), 1 arginosuccinate lyase (ASL) and 1 arginase (Figure 4, Table 1). Except for the arginase and one OCT, all the genes involved in the urea cycle are overexpressed in high-nitrate conditions. This pattern is expected to remove excess of ammonia from the cell through the urea cycle.

## 4 - Discussion

### *P. calceolata* growth under various nitrogen sources

Our results show that *P. calceolata* has a similar growth under nitrate, urea and cyanate as sole N source. Based on previous metabarcoding and metatranscriptomic studies, it was already suggested that main inorganic N source for *P. calceolata* was nitrate, based on how its abundance was positively impacted by mixing events, and how it dominates nitrate assimilation in oligotrophic areas (Dupont et al. 2015; Choi et al. 2020). As many other phototrophs, *P. calceolata* can grow on organic nitrogen sources. For example, *A. anophagefferens* has a greater growth when cultivated under urea compared to many others N sources such as nitrate, glutamic acid and ammonium (Berg et al. 1997). Similarly to *A. anophagefferens*, ammonium as N source strongly affected *P. calceolata* growth. To date, *P. calceolata* and *Prorocentrum donghaiense* are the only eukaryotic algae shown to be able to grow under cyanate as only N source (Hu et al. 2012).

### Global response to low nitrate conditions

When facing low-nitrate, *P. calceolata* presents a common response in the lab and *in situ*, but the use of biological replicates in the lab gives access to a larger number of genes to study the metabolism. There is an overexpression of genes involved in degradation of metabolites, such as peptidases, nucleases, hydrolases, dioxygenases, transferases and transporters for amino acids, nucleotides. There is also a

downregulation of gene involved in urea and pyruvate cycle (CPS, methylglyoxal synthase), porphyrin metabolism (uroporphyrinogen decarboxylase), DNA transcription (helicase, initiation and elongation factors), protein biosynthesis (ribosomal protein, tRNA synthetase), and proteins involved in protein folding (GroES chaperonine, heatshock protein 70). It seems that under low-nitrate, *P. calceolata* has a double strategy to save on nitrogen needs by downregulated protein biosynthesis and to recycle on intracellular N by catabolising metabolites. This was already observed in other PPEs such as diatoms and Pelagophyte *A. anophagefferens*. In diatoms, nitrogen limitations induce common responses among different species: a decrease in carbon fixation, protein biosynthesis, carbohydrates, fatty acid metabolism and an increase in nitrogen recycling (S. Bender et al. 2014). In *A. anophagefferens* nitrate-depletion induced up-regulation of amino acid and nucleotide transporters, and amides degradation enzymes and nitrate assimilation pathway (Dong et al. 2014). We hypothesize that under low-nitrate, *P. calceolata* recycle intracellular nitrogen by degrading metabolites such as amino acids, nucleotides, allocate gathered N to essential functions and simultaneously decrease its nitrogen needs by reducing protein metabolism and synthesis.

## **Global response to alternative nitrogen source**

It seems that the response of *P. calceolata* to changing nitrogen source is consistent regardless of the new source, whether it is inorganic such as ammonium or organic like urea or cyanate. Overexpressed function seems to be related to breakdown of metabolites, sign that there is maybe a similar form of intracellular N recycling than the one observed in N-depleted conditions. Underexpressed functions are clearly plastidic nitrogen assimilation pathways, with downregulation of genes such as formate/nitrite transporters, nitrate/nitrite transporters and nitrite reductase. Photosynthesis seems to be also negatively impacted, with downregulation of gene coding for chlorophyll A/B binding protein and photosystems II. Decrease of nitrate assimilation under changing N source was observed in other Pelagophytes. Urea is the most studied alternative nitrogen source, and its presence negatively affects plastidic N assimilation in Pelagophyceae CCMP 2097 (Terrado et al. 2015). In *A. anophagefferens* it was observed that urea as N-sources triggered upregulation of genes involved in protein, amino acid, spermine and sterol synthesis (Dong et al. 2014), but these functions were not observed in *P. calceolata*. Since *A. anophagefferens* presented a stronger growth under urea as N source, expressions of these biosynthesis genes might be a sign of enhanced growth, while growth of *P. calceolata* under urea and nitrate are equivalent.

## **Ammonium usage in *P. calceolata***

In our experiments, *P. calceolata* growth under ammonium was slow compared to nitrate and urea, which is surprising considering that ammonium is the preferred source of nitrogen for most of microalgae (John A. Raven, Wollenweber, and Handley 1992). Despite this low affinity for ammonium, the five ammonium transporters genes in *P.*

*calceolata* genome are highly expressed and differentially regulated according to different N conditions. Ammonium transporters are generally up regulated in low-nitrate conditions in diatoms (S. Bender et al. 2014; Remmers et al. 2018; Smith et al. 2019; Scarsini et al. 2022) and *A. anophagefferens* (Gry Mine Berg et al. 2008). In *A. anophagefferens*, ammonium transporters are overexpressed when nitrate is the nitrogen source (Dong et al. 2014). In *P. calceolata*, only single ammonium transporter gene is significantly overexpressed in low-nitrate samples. The cellular localization remain to be studied for these transporters but we cannot exclude that several of them are present on mitochondrial or plastid membranes rather than on the plasma membrane.

### **Urea cycle in *P. calceolata***

Similarly, to diatoms, *P. calceolata* presents a metazoan-like ornithine-urea cycle that seems to facilitate rapid recovery from prolonged N limitation by redistributing nitrogen-containing metabolites (Allen et al. 2011). Except for the arginase and one of the 2 OCT, all the genes involved in the urea cycle are overexpressed in high-nitrate conditions suggesting that the excess of nitrate in the media requires an elimination through the urea cycle. Furthermore, the underexpression of ASS and the two CPS under urea is opposite to the pattern observed in *A. anophagefferens* (Dong et al. 2014). It seems that the regulation of the urea cycle in *P. calceolata* is more similar to diatoms such as *P. tricornutum*, *T. pseudonana* or *F. cylindrus* (S. Bender et al. 2014; Smith et al. 2019).

### **Uric acid/xanthine permease**

Among the genes that are overexpressed both in environmental and experimental low-nitrate conditions, one gene carries the domain K23887 coding for a uric-acid/xanthine permease. Its function is the transport of purine degradation products such as uric acid, xanthine and vitamin C.

Overexpression of uric-acid/xanthine transporters under low-nitrate condition as already been observed in the arctic Pelagophyceae CCMP2097 (Terrado et al. 2015) and in *A. anophagefferens*. It was suggested to be a sign of purine uptake as N source under inorganic N deprivation, which could be a key process to bloom despite low-nitrate availability (Gry Mine Berg et al. 2008; Wurch, Gobler, and Dyhrman 2014; Dong et al. 2014). We can hypothesise that expression of this permease is linked to catabolism of purines as a recycling mechanism in *P. calceolata*, in contrary to *A. anophagefferens* where no correlation were found between the expression of this gene and DIN concentration in the environment (Wurch, Gobler, and Dyhrman 2014).

The conversion of xanthine into urea and ammonia under low-N conditions has been suggested in *A. anophagefferens* (Gann et al. 2022). Since two enzymes of this pathway were missing in *P. calceolata* and most genes were not regulated by changing nitrate quantity (Table 1), we cannot conclude that *P. calceolata* converts xanthine into

ammonium under low-nitrate condition as a way to recycle intracellular N as *A. anophagefferens* does.

Another hypothesis might be that the overexpression of this permease is a sign of utilisation of N storage in the form of uric acid, xanthine and/or guanine crystals. Indeed, in the symbiotic dinoflagellate *Symbiodinium*, uric acid crystals are produced and might serve as a strategy to store N under oligotrophic condition (Clode et al. 2009). In the case of the symbiosis between *Pocillopora damicornis* and its symbiont, <sup>15</sup>N labelling shows that nitrogen from assimilated ammonium, nitrate and aspartic acid is quickly incorporated in the algae as uric acid crystals. When retransferred in low-N conditions, the nitrogen is reallocated from the algae to the coral symbiont, supporting the hypothesis of uric acid crystals as N storage to face low-N conditions (Kopp et al. 2013). Until recently, this mechanism was believed to be specific to symbiotic dinoflagellate, but a recent study shows that purine inclusions are widespread among major eukaryotic super groups (Pilátová et al. 2022). They identified hypoxanthine-guanine phosphoribosyl transferase as key enzyme in purine usage after crystal degradation. A gene coding for this enzyme is present in *P. calceolata* genome but is not overexpressed in any of our tested conditions. Among the identified transporters, the uric-acid/xanthine permease belongs to the nucleobase-ascorbate transporter family. If recent research highlights the possibility that organisms such as *P. calceolata* might be able to store nitrogen in purine crystals, it is hasty to pretend that overexpression of the purine permease under low-nitrate conditions is a sign of crystals' nitrogen reallocation. Even so, it remains an interesting lead to investigate in further study, since such mechanisms seem to be underestimated in photosynthetic pico-eukaryotes.

### **Role of the cyanate lyase in *P. calceolata***

In our study, we have shown that *P. calceolata* can grow under cyanate as sole nitrogen source. Cyanate is rarely tested as nitrogen source in eukaryote algae, therefore this capacity has only been shown in *Prorocentrum donghaiense* (Dinophyceae), at the expense of a strongly reduced growth rate (Hu et al. 2012). Considering the large diversity of algae capable of growth under various organic nitrogen source, it is then possible that cyanate would supports growth of many taxa.

It has been suggested that all phytoplankton overexpressing the cyanate lyase in nitrate-low environments are therefore capable of cyanate uptake and metabolism (Dong et al. 2014; Smith et al. 2019; Mao et al. 2022). Although, *P. calceolata* overexpressed its cyanate lyase in low-nitrate environments (*in situ* and in the lab), we have shown that this gene is underexpressed when *P. calceolata* is cultivated with cyanate. In consequence, we believe that the cyanate lyase is not responsible for external cyanate metabolism and should not be used as a marker of cyanate uptake. Firstly, cyanate transports in eukaryotic cells remain undescribed, and the bacterial transporter of cyanate CynABD is not conserved in eukaryotes (Widner and Mulholland

2017). Secondly, and in agreement with our results, Linder studied several ascomycetes yeast species and their capacity to grow on cyanate as sole N source (Linder 2018). They observed that growth was not supported by some species despite having gene coding for cyanate lyase. On the contrary, cyanate supported the growth of several species despite the absence of cyanate lyase coding genes. Thus, it is possible that a gene yet undescribed allowed assimilation of cyanate as N source for eukaryotic organisms.

This result raises two questions. Firstly, why is cyanate lyase overexpressed in low-nitrate conditions if it is not to uptake cyanate as an alternative nitrogen source? Secondly, how can cyanate support growth of *P. calceolata* if the gene coding cyanate lyase is downregulated?

An alternative function of cyanate lyase is the detoxification of intracellularly produced cyanate. Cyanate can be generated through spontaneous fast abiotic decomposition of carbamoyl-phosphate (Anderson, Sung, and Fuchs 1990; Ter-Ovanesian et al. 2021) or the slow abiotic decomposition of urea (Dirnhuber and Schütz 1948). As described earlier, *P. calceolata* recycles its amino acids and proteins in order to retrieve intracellular nitrogen resources. In this process, many enzymes are involved in the breaking down of metabolites, including large protein family carbamoyltransferase (K00612), which can produce carbamoyl-phosphate (CP). Among *P. calceolata* genes that are overexpressed in low-nitrate environments, two of the five genes coding for carbamoyltransferase are slightly overexpressed in intermediate-nitrate condition, and one in low-nitrate condition. We hypothesise that these enzymes increase intracellular concentration of CP, which is in principle taken up by the urea cycle. However, under low-nitrate conditions the urea cycle is downregulated, which could lead to an increase of intracellular cyanate, requiring the cyanate lyase activity to detoxify the cell. This hypothesis is summarised on Figure 5.

Our results tends to show that the cyanate lyase of *P. calceolata* is involved in detoxification of intracellular cyanate, which is a by-product of carbamoyl-phosphate generated by amino-acid degradation and not taken up by the urea cycle under low-nitrate conditions. In complement to the work of Sato, Hashihama, and Takeda 2023 and the phylogeny of Mao et al. 2022, we hypothesise that the detoxification of cyanate lyase is the main role of this enzyme in eukaryotic microalgae. This function could explain its overexpression in low nitrate and underexpression in cyanate, instead of the external cyanate metabolism as alternative N source, which was proven in several bacterial and fungal species, but not in PPEs. Further experiments on cyanate uptake in *P. calceolata* and other PPEs are required in order to confirm our hypothesis that PPE's cyanate lyases are not involved in external cyanate assimilation.

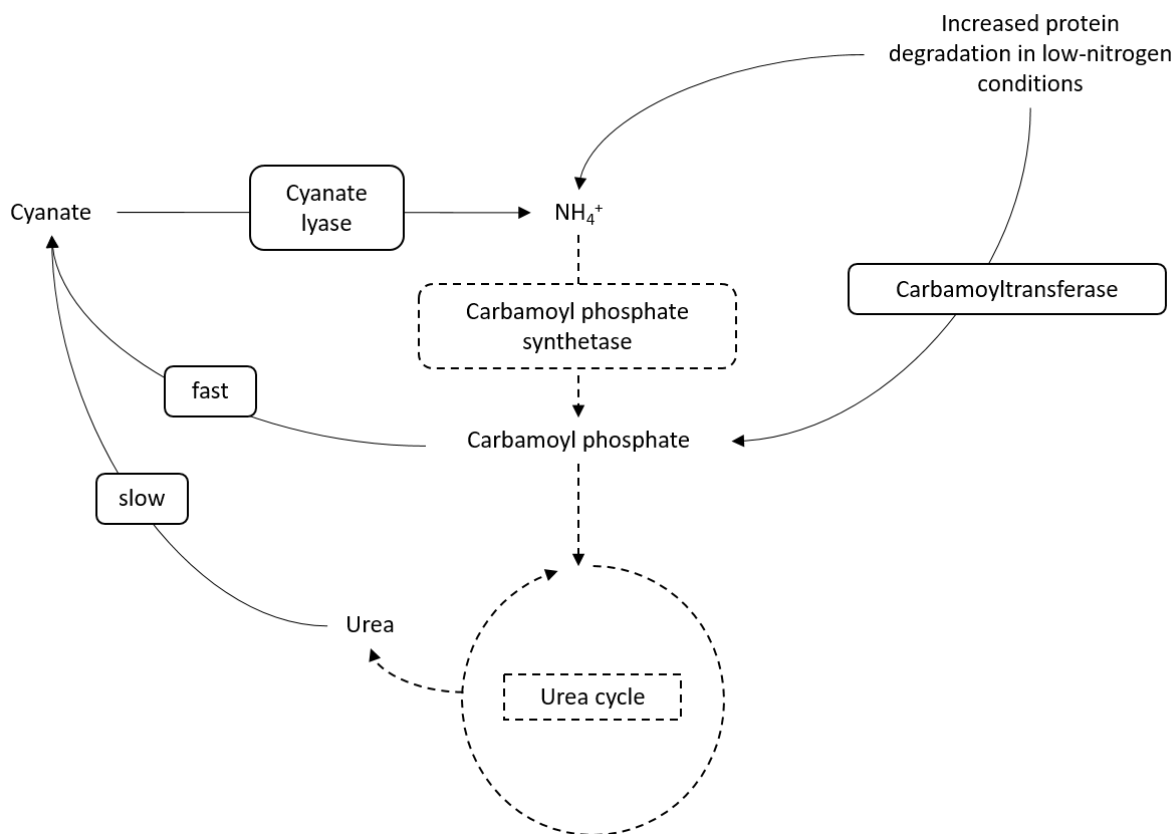


Figure 5. Putative role of the cyanate lyase as a detoxifier under low-nitrate conditions. Instead of assimilation of environmental cyanate into ammonium, we hypothesized based on our results that cyanate lyase plays a role as a detoxifying system for intracellular cyanate. Under low-nitrate conditions, the increased of protein degradation by (among others) carbamoyltransferases leads to an abnormal accumulation of carbamoyl phosphate. Instead of being incorporated into the downregulated urea cycle, it is spontaneously degraded into cyanate, which requires up-regulation of cyanate to avoid intoxication of the cell.

## 5 - Materials & Methods

### Differential analysis of gene expression levels in environmental datasets.

In order to detect gene expression of *P. calceolata* in the oceans, we used metatranscriptomic datasets from *Tara Oceans* and *Tara Polar Circle* expeditions (Alberti et al. 2017). Metatranscriptomic reads were aligned on the 16 667 predicted mRNA sequences of *P. calceolata* genome (ENA PRJEB34158) (Guérin et al. 2022) with bwa-mem2 version 2.2.1 with default's parameters (Li et al. 2009). We selected reads aligned with more than 95% of identity over 80% of the read length. We kept nuclear genes covered with a minimum of 10 reads in at least 10 samples. To eliminate putative cross-mapped genes (i.e. highly conserved genes which potentially aggregate reads from other organisms) we removed genes detected in more than 90% of samples including those where *P. calceolata* is not present. Finally, 124 samples with more than 75% of detected *P. calceolata* genes were kept for the next steps. We finally got the expression of 15 617 genes across 124 environmental samples. The environmental

parameters measured during the expedition are available and described in the Pangaea database (<https://www.pangaea.de/>) (Pesant et al. 2015). We selected all available datasets from seawater samples of the photic zone (73 surface and 51 deep-chlorophyll maximum samples), and from 2 size-fractions: 80 from the 0.8 - 5  $\mu\text{m}$  and 44 from the 0.8-2000  $\mu\text{m}$ . Nitrate concentrations are calculated from *in situ* sensor (SATLANTIC) data, calibrated using water samples. Samples are considered "low-nitrate" if they contain less than 2  $\mu\text{M}$  of nitrate. Differential expression analyses were conducted with DESeq2 package version 1.32.0 (Love, Huber, and Anders 2014) under R version 4.1.1 across the 15,617 genes and the 112 environmental samples with available *in situ* nitrate concentration measurements. Pairwise comparisons were performed across 43 "high-nitrate" samples (nitrate concentration > 2  $\mu\text{M}$ ) and 69 "low-nitrate" samples (nitrate concentration < 2  $\mu\text{M}$ ) and with the DESeq function with default parameters and log2fold change values were calculated with lfcShrink. We considered genes with a p-value < 0.01 and a log2 fold change > 2 or < -2 as differentially expressed. Data for differentially expressed genes are available in Supplementary Tables 1 to 6.

### **Cultures of *P. calceolata* RCC100 strain**

Non-axenic RCC100 strain of *P. calceolata* was cultivated in artificial seawater (ASW) supplemented by L1 medium (as described in Guillard and Hargraves 1993). Briefly, ASW is prepared by dissolving 24.55g of sodium chloride (NaCl), 0.75g of potassium chloride (KCl), 4.07g of magnesium chloride hexahydrate ( $\text{MgCl}_2 \cdot 6 \text{H}_2\text{O}$ ), 1.11g of calcium chloride ( $\text{CaCl}_2$ ), 2.95g of magnesium sulphate ( $\text{MgSO}_4$ ) and 0.21g of sodium bicarbonate ( $\text{NaHCO}_3$ ) in 1L of sterile water. 1 ml of trace metals, vitamins and nutrients from the L1 medium Bigelow Kit (MKL150L) were added to reach following concentrations : 882  $\mu\text{M}$  of sodium nitrate ( $\text{NaNO}_3$ ), 36.2  $\mu\text{M}$  of monosodium phosphate ( $\text{NaH}_2\text{PO}_4^-$ ) and 106  $\mu\text{M}$  of sodium silicate ( $\text{Na}_2\text{SiO}_3$ ). Cultures were maintained at 20°C, under a 12:12h light-dark photoperiod and a blue light at the intensity of 20  $\mu\text{mol. m}^{-2} \cdot \text{s}^{-1}$  of photosynthetic photons. For the nitrate depletion experiment, the nitrate concentration was divided by 2 or 4 for a final concentration of 441  $\mu\text{M}$  or 220  $\mu\text{M}$  respectively. For the cyanate and the ammonium experiment, the nitrate was replaced by potassium cyanate (KOCN) or ammonium chloride ( $\text{NH}_4\text{Cl}$ ) at the same concentration (882  $\mu\text{M}$ ). For the urea experiment, the nitrate was replaced by urea ( $\text{CH}_4\text{N}_2\text{O}$ ) concentrated at 441  $\mu\text{M}$  in order to have the same concentration of nitrogen atoms in all conditions.

RCC100 was grown in each tested condition during an acclimation phase of minimum 8 days then diluted in fresh medium in triplicates during 5 to 9 days until they reached a minimum concentration of 10 million cells per ml. Growth and fluorescence was daily followed with a Qubit instrument (Invitrogen™ Qubit™ 3 Fluorometer Q33216, blue excitation at 470 nm, far red emission 665-720 nm). Cells were counted on a Thoma cell under microscope at the day of the harvesting (Marienfeld Thoma counting chamber, depth 0.1 mm, 0640710). *P. calceolata* cells were harvested by filtration on a 1.2  $\mu\text{m}$  mixed-cellulose esters membrane filter (MF-Millipore, rawp04700) with a

peristaltic pump (SFP-100), then transferred in 15ml tubes, flash-frozen in liquid nitrogen, and stored at -80°C until RNA extraction.

## **RNA extraction and sequencing**

Flash-frozen filters with *P. calceolata* cells were vortexed in QIAzol, then RNAs were extracted following RNeasy Plus Universal Mini Kit procedure (Qiagen, Ref 73404). All extracted RNAs were treated 6U of TURBO™ DNase (2 U/μl) (Thermo Fisher Scientific, Ref. AM2238) then purified with RNA Clean and Concentrator-5 kit (Zymo Research, Ref. ZR1016), keeping only large RNAs (> 200 nt) for the RNAseq library preparation. 100 ng of treated RNAs were used to create Illumina library (Illumina Stranded mRNA Prep, Ligation). Briefly, poly(A) + RNAs were selected with oligo(dT) beads, chemically fragmented by divalent cations under high temperature, converted into single-stranded cDNA using random hexamer priming, followed by second strand synthesis and 3'-adenylated. A pre-index anchor was ligated and a PCR amplification step with 15 cycles was done to add a 10bp unique dual index adapter sequences (IDT® for Illumina® RNA UD Indexes, Ligation). All libraries were quantified by Qubit dsDNA HS Assay measurement. A size profile analysis was performed in an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Libraries were sequenced in 2x150 bp on Illumina NovaSeq 6000 sequencer (Illumina, San Diego, CA, USA) in order to obtain 50 millions paired-end reads. After Illumina sequencing, a quality control process was applied to the reads that passed the Illumina quality filters (Alberti et al. 2017). Briefly, Illumina sequencing adaptors and primer sequences were removed, then low-quality nucleotides (Q < 20) were discarded from both ends of the reads. Sequences between the second unknown nucleotide (N) and the end of the read were also trimmed. We discarded reads shorter than 30 nucleotides after trimming with an adaptation of fastx\_clean tool (<https://www.genoscope.cns.fr/fastxtend/>). In the last step, we discarded reads that were mapped to the Enterobacteria phage PhiX174 genome (GenBank: NC\_001422.1), using bowtie2 v2.2.9 (-L 31-mp 4-rdg 6,6-local-no-unal) (Langmead and Salzberg 2012). Remaining rRNA reads were removed using SortMeRNA v2.1 and SILVA databases (Kopylova, Noé, and Touzet 2012; Quast et al. 2013).

## **Analysis of *P. calceolata* gene expression levels in different nitrogen conditions**

Similarly, to the analysis of environmental metatranscriptomes, RNAseq reads were aligned with bwa-mem2 version 2.2.1 on the *P. calceolata* genome. Reads with a minimal size of 50 bp and aligned with more than 95% of identity over 80% of the read length were selected. 16 659 out of 16 667 genes were detected in at least one sample. We only kept nuclear genes, then normalised the gene expression levels in transcripts per kilobase per million mapped reads (TPM).

Pearson's correlation matrices were computed with R using cor function between

triplicates and across conditions, based on the gene expression normalised in TPM. Hierarchical clustering of euclidian distances between samples were performed with `dist` and `hclust` functions of the `stats` package on R version 4.1.1

As described above for the metatranscriptomes, differential expression analysis on transcriptomic samples was carried out with `DESeq2`. Identification of differentially expressed genes was performed by pairwise comparisons between the standard condition (882  $\mu\text{M}$  of nitrate) and each other condition (220  $\mu\text{M}$  of nitrate, 441  $\mu\text{M}$  of nitrate, 882  $\mu\text{M}$  of ammonium, 882  $\mu\text{M}$  of cyanate and 441  $\mu\text{M}$  of urea). We considered as differentially expressed genes with a p-value  $< 0.01$  and a  $\log_2$  fold change  $> 2$  or  $< -2$ .

All figures were generated with `ggplot2` version 3.4.1 under R version 4.0.3 . The Euler diagrams was done with `eulerr` version 6.1.1 and `graphics` version 4.0.3.

## Chapitre 3 – L'impact de la profondeur sur le métabolisme du carbone de *Pelagomonas calceolata*

---

### Préambule au chapitre 3

Dans ce troisième chapitre, notre objectif est de comprendre comment *P. calceolata* s'acclimate aux différentes profondeurs. En effet, la profondeur d'échantillonnage est un des paramètres impactant le plus l'abondance relative et l'expression des gènes de *P. calceolata* dans l'environnement.

En analysant la liste des gènes différentiellement exprimés en surface par rapport à la DCM, nous avons constaté que le cycle de Calvin et la photorespiration sont surexprimés dans les eaux de surface. On sait encore peu de choses sur les paramètres environnementaux impactant ces cycles centraux du métabolisme du carbone.

Notre objectif ici est donc de se questionner sur l'impact de la profondeur sur l'expression des gènes de *P. calceolata*, comprendre pourquoi ces cycles en particuliers sont surexprimés en surface par rapport à la DCM, et ce que cela implique d'un point de vue métabolique.

J'ai réalisé l'ensemble des analyses de ce chapitre.

Dans le cadre de ce chapitre, j'ai pu travailler avec M. Gilles Peltier, directeur de recherche CEA à l'Institut de Biosciences et Biotechnologies Aix-Marseille (BIAM), CEA Cadarache/ Cité des Énergies. Je le remercie pour nos échanges et pour sa contribution.



## 1 - Introduction

L'océan est un environnement hautement stratifié. Les variations de paramètres physiques des différentes masses d'eau (température, densité, salinité) génèrent une partition verticale, créant autant d'environnements de composition physico-chimiques différentes. Bien que pouvant se mélanger selon les saisons ou les courants, cette stratification verticale impacte fortement la distribution des micro-organismes dans la colonne d'eau. Les organismes photosynthétiques nécessitant de la lumière ne sont que dans la zone photique, qui s'étend de la surface jusqu'à environ 200 m de profondeur. La taille de la zone photique varie avec la turbidité de l'eau et la latitude. Elle peut elle-même être de nouveau divisée en deux zones : la surface et la profondeur de chlorophylle maximum (ou DCM). La première est en général plus illuminée, plus chaude, mais oligotrophe. La seconde présente une quantité de lumière plus faible, mais une disponibilité en nutriments plus importante, permettant ainsi une plus grande abondance d'organismes photosynthétiques, ce qui se traduit par une plus grande concentration de pigments à cette profondeur. Les différences de luminosité, concentration de nutriments, ou température peuvent fortement impacter le métabolisme central du phytoplancton.

La DCM marque en général la profondeur où la concentration de nutriments et la quantité de lumière sont équilibrées pour permettre une croissance optimale du phytoplancton. Plus haut, il n'y aurait plus suffisamment de nutriments, et plus bas plus assez de lumière. Les organismes photosynthétiques présentent donc une abondance et une activité cellulaire (mesurée par un ratio ADN:ARN) plus importantes à la DCM qu'en surface (Giner et al. 2020). Les Pelagophyceae, comme beaucoup d'organismes photosynthétiques, ont tendance à être présents en plus grande abondance relative à la DCM (Cabello et al. 2016; D. H. Choi et al. 2016). Dans le premier chapitre, nous avons montré, par analyse de la distribution métagénomique de *P. calceolata*, qu'elle était significativement plus abondante à la DCM qu'à la surface dans la mer Méditerranée, la mer Rouge et l'Océan indien, relativement aux autres eucaryotes de la fraction de taille 0.8-5  $\mu\text{m}$  (Chapitre 1, Figure 2). Cette tendance est d'ailleurs également observée dans les autres bassins océaniques, et dans la fraction de taille 0.8-2000  $\mu\text{m}$  (Annexes du chapitre 1, Figure Supplémentaire 6). Ainsi, nous avons confirmé que *P. calceolata* était relativement plus abondante à la DCM qu'à la surface. Cependant, une étude de métatranscriptomique de 2015 montre que *P. calceolata* domine la communauté picoeucaryote dans les eaux de surface du Pacifique, et que son rôle écologique y est potentiellement sous-estimé (Dupont et al. 2015). Du fait de sa petite taille, *P. calceolata* devrait être plus apte à se développer dans les eaux oligotrophes en surface, mais également être plus sensible aux dégâts causés par une forte luminosité (Raven 1998).

De nombreuses études s'intéressent à la façon dont la composition de la communauté planctonique varie avec la profondeur. Mais il y a assez peu d'exemples d'études s'intéressant plus spécifiquement à la manière dont la profondeur affecte la

physiologie. Pourtant, on s'attend à ce que les pressions biotiques et abiotiques subies par les cellules soient très différentes entre ces 2 environnements. Dans ce troisième chapitre, je vais décrire les effets de la profondeur sur l'expression des gènes de *P. calceolata*, quelles fonctions et métabolismes sont associés à chaque strate, et comment cette algue s'acclimate aux contraintes spécifiques de ces différents niveaux.

## 2 - Méthodes

Pour réaliser ces analyses, je me suis basée sur l'analyse de l'expression des gènes de *P. calceolata* dans l'environnement marin, en utilisant les données métatranscriptomiques issues de l'expédition *Tara Oceans*. Les lectures métatranscriptomiques de *Tara Oceans* ont été alignées sur le transcriptome de référence de *P. calceolata*, puis filtrées, et la table d'expression a été normalisée en nombre de transcrits par millions de lectures alignés (TPM) comme décrit dans la partie Material & Methods du chapitre 2.

L'analyse des résultats a été faite sous R version 4.1.1 et R-Studio version 2022.02.0, sauf la comparaison avec les OCT générés expérimentalement qui a été faite sous R version 4.0.3 et R-studio version 1.3.1093.

L'analyse des corrélations canoniques a été réalisée avec la fonction `cca` du package `vegan` version 2.6-4, et la figure a été faite avec `ggplot2` version 3.4.1. Les paramètres suivants ont été intégrés à la CCA : la température mesurée *in situ* (°C), la concentration en ammonium ( $\mu\text{M}$ ) modélisée par le modèle Darwin (<http://darwinproject.mit.edu/>), la concentration de nitrate mesurée *in situ* ( $\mu\text{M}$ ), la concentration en nitrite mesurée *in situ* ( $\mu\text{M}$ ), la concentration en phosphate *in situ* ( $\mu\text{M}$ ) la concentration en fer (nM) modélisée avec PISCES 2 (Aumont et al. 2015), la quantité totale de carbone *in situ* ( $\mu\text{mol/kg}$ ), le rayonnement photosynthétique actif (PAR) moyen sur 30 jours ( $\text{quanta/m}^2/\text{jour}$ ), et la concentration en chlorophylle A *in situ* corrigée ( $\text{mg/m}^3$ ).

Les paramètres environnementaux de la surface et de la DCM ont été comparés grâce à des tests de Wilcoxon-Mann-Whitney pairés, réalisés avec la fonction `pairwise.wilcox.test`, issue du package `stats` version 4.1.1. Nous avons appliqué une correction de la p-value de Benjamini & Hochberg et fixé le seuil de significativité des p-values à 0.01.

L'analyse des gènes différentiellement exprimés entre la surface et la DCM a été réalisée avec le package `DESeq2` version 1.32.0 avec les fonctions `DESeqDataSetFromMatrix`, `DESeq` et `lfcShrink`. L'expression des gènes de *P. calceolata* a été comparée entre 73 échantillons de surface et 51 échantillons de DCM. Les gènes présentant une valeur de  $\log_2$  fold-change supérieur à 2 ou inférieur à -2 et ayant une p-value ajustée inférieure à 0.01 sont considérés comme différentiellement exprimés. Le mantel plot a été réalisé avec `ggplot2`. Les analyses d'enrichissement ont été effectuées en comparant le nombre d'occurrence d'un domaine dans la liste des gènes différentiellement exprimés par rapport à son occurrence dans le génome, avec un test

exact de Fisher.

Les expériences de culture dans différentes quantités et sources d'azote, l'extraction de l'ARN, le séquençage, l'alignement sur le transcriptome de référence et l'analyse de DEG en fonction de la quantité et de la source d'azote sont décrits en détails dans le chapitre 2. La comparaison des DEG en fonction de la profondeur avec ceux générés expérimentalement et les figures correspondantes ont été réalisées avec les packages `eulerr` version 6.1.1 et `graphics` version 4.0.3.

Les gènes d'intérêt impliqués dans le cycle de Calvin, la photorespiration et les mécanismes de concentration du carbone ont été identifiés sur base de l'annotation fonctionnelle du génome, réalisée dans le cadre du chapitre 1. Les gènes présentant une valeur de  $\log_2$  fold-change supérieure à 1 ou inférieure à -1 et ayant une p-value ajustée inférieure à 0.01 sont considérés comme différentiellement exprimés. Les boîtes à moustaches permettant de visualiser l'expression de ces gènes ont été réalisées avec `ggplot2`.

La carte thermique (heatmap) de leurs expressions a été réalisée à partir d'une version normalisée entre 1 et 10 de la table d'expression en TPM de ces gènes uniquement, en utilisant la fonction `Heatmap` du package `ComplexHeatmap` 2.10.0. Le partitionnement des données d'expression (clustering) a été réalisé en calculant une matrice de distance euclidienne à partir de la table d'expression en TPM normalisée entre 1 et 10, puis en partitionnant avec la méthode complète. Le nombre de clusters optimal a été calculé avec la fonction `fviz_nbclust` du package `factoextra` 1.0.7, en utilisant la méthode Elbow (Figure Supplémentaire 1). Les couleurs de la palette Spectral du package `RColorBrewer` version 1.1-3 ont été utilisées pour réaliser la heatmap.

### 3 - Résultats et Discussion

#### **L'expression des gènes de *P. calceolata* est fortement impactée par la profondeur**

Le paramètre environnemental impactant le plus fortement l'expression des gènes de *P. calceolata* est la profondeur à laquelle l'organisme a été échantillonné. En effet, le profil d'expression des gènes de *P. calceolata* dans l'environnement est très fortement variable selon la profondeur d'échantillonnage. L'analyse canonique des corrélations (Figure 1) permet de visualiser la séparation des profils d'expression des gènes de *P. calceolata* entre les échantillons de surface (bleu) et à la DCM (vert).

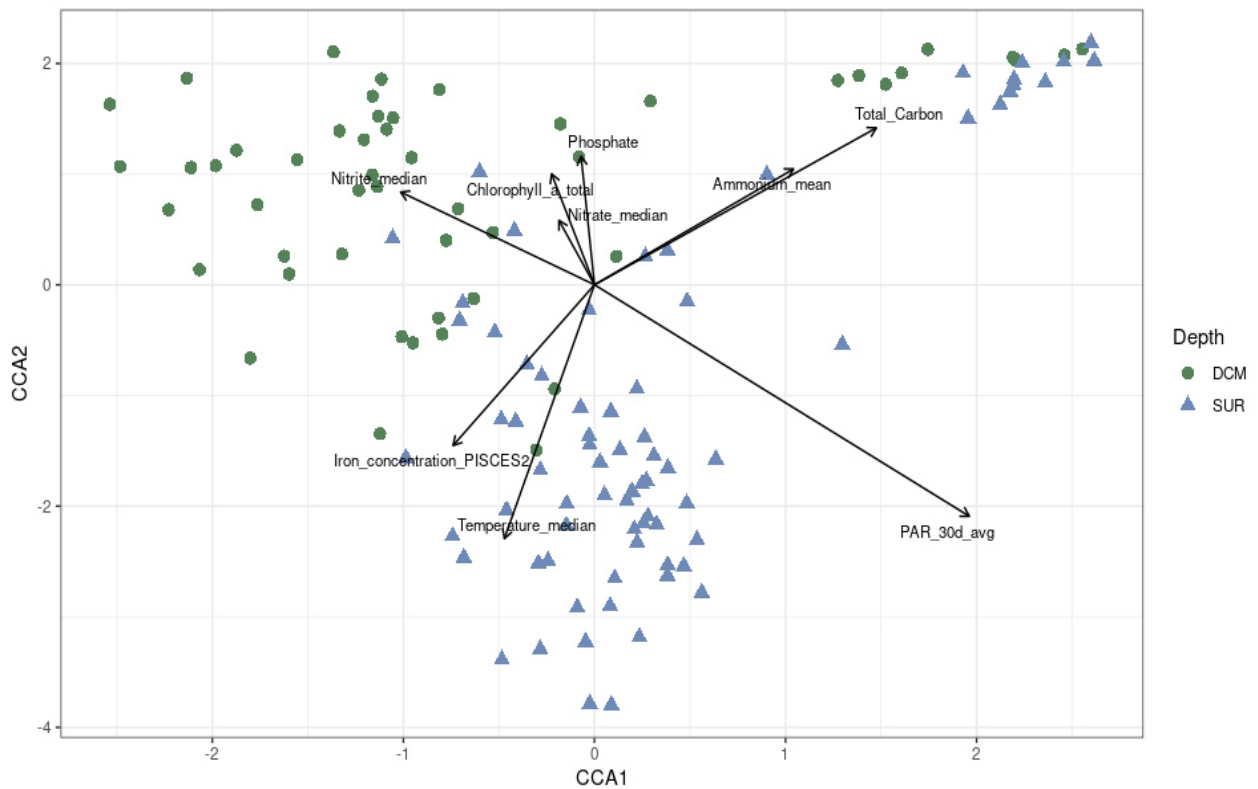


Figure 1. Analyse des corrélations canoniques de l'expression des gènes de *P. calceolata* normalisée (TPM) dans les échantillons *Tara* Oceans en fonction de paramètres environnementaux : la température (°C), la concentration en ammonium ( $\mu\text{M}$ ), en nitrate ( $\mu\text{M}$ ), en nitrite ( $\mu\text{M}$ ), en phosphate ( $\mu\text{M}$ ), en fer (nM), la quantité totale de carbone ( $\mu\text{mol/kg}$ ), le rayonnement photosynthétique actif (PAR) moyen sur 30 jours (quanta/m<sup>2</sup>/jour), et la concentration en chlorophylle A (mg/m<sup>3</sup>). L'axe CCA1 explique 10 % de la variance totale, l'axe CCA2 6 %.

Comme évoqué dans l'introduction, la surface et la DCM sont des environnements différents. La première est oligotrophe mais riche en lumière, et la seconde est moins illuminée mais plus riche en nutriments. Ces différences physico-chimiques sont observables dans les données *Tara* Oceans. En comparant les valeurs de chaque paramètre physico-chimique des échantillons de surface avec ceux de la DCM, au moyen d'un test de Wilcoxon-Mann-Whitney pairé, on constate que la concentration en pigments et de certains nutriments (phosphate, nitrite et silicate) est plus importante à la DCM, alors que la température et la quantité de lumière sont plus importantes à la surface (Figure 2, Table Supplémentaire 1). Au regard de ces différences environnementales, il n'est pas surprenant que la profondeur d'échantillonnage impacte autant l'expression des gènes de *P. calceolata*.

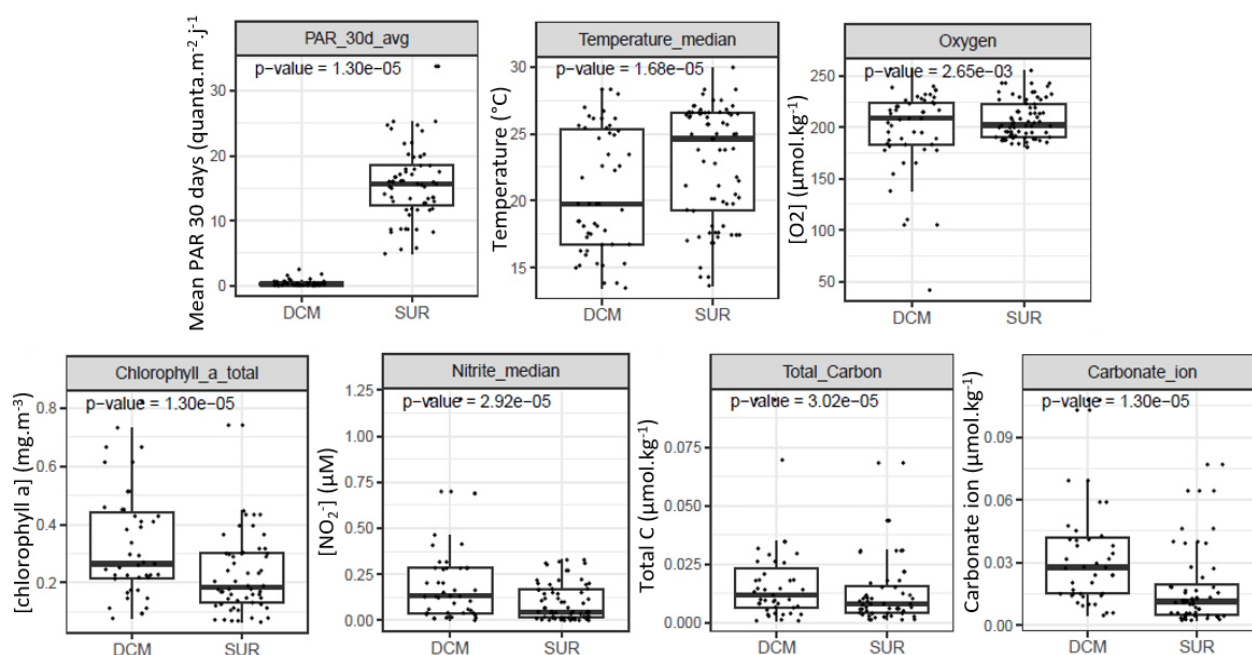


Figure 2. Distribution des valeurs de plusieurs paramètres environnementaux qui varient significativement entre la DCM et la surface (SUR). La p-value est celle obtenue avec le test de Wilcoxon-Mann-Whitney païré. Sur la ligne du haut : le rayonnement photosynthétique actif moyen sur 30 jours ( $\text{quanta.m}^{-2}.\text{jour}^{-1}$ ), la température ( $^{\circ}\text{C}$ ) et la concentration en oxygène dissous ( $\mu\text{mol/kg}$ ) sont significativement plus élevés en surface. Sur la ligne du bas : la concentration en chlorophylle A ( $\text{mg/m}^3$ ), la concentration en nitrite ( $\mu\text{M}$ ), la quantité totale de carbone ( $\mu\text{mol/kg}$ ) et la concentration en ions carbonate ( $\mu\text{mol/kg}$ ) sont significativement plus élevés à la DCM. D'autres paramètres présentent des distributions significativement différentes entre la surface et la DCM, les résultats des tests sont visibles dans la Table Supplémentaire 1.

Dans une démarche similaire à celle appliquée dans le chapitre 2, nous avons utilisé les données méatranscriptomiques *Tara Oceans* afin d'étudier l'impact de la profondeur sur la réponse transcriptomique de *P. calceolata*, ce qui permet d'étudier son métabolisme directement dans des conditions naturelles. Les niveaux d'expression des gènes entre les échantillons de surface et les échantillons de DCM ont été comparés grâce à une analyse d'expression différentielle, révélant deux physiologies très distinctes de *P. calceolata* (Table Supplémentaire 2). Ainsi, avec un  $\log_2$  fold-change supérieur à 2 et une  $p\text{-value} < 10^{-3}$ , 416 gènes sont significativement surexprimés à la DCM et 201 gènes sont significativement surexprimés à la surface (Figure 3).

À la DCM, beaucoup des gènes surexprimés n'ont pas de domaines fonctionnels connus. Ainsi, sur les 10 gènes les plus fortement surexprimés à la DCM, un seul est porteur de domaines InterProScan connus : les domaines IPR032710, IPR001202 et IPR036020, codant respectivement pour un facteur de transport nucléaire NFT-2, et des domaines conservés à tryptophanes (WW domain). En recherchant les domaines fonctionnels enrichis à la DCM, on constate un important enrichissement de gènes codant pour des dynéines. Au total, 16 domaines IPR enrichis à la DCM sur 21 sont porteurs de domaines fonctionnels de dynéines (Table Supplémentaire 3). Ces protéines sont des marqueurs de l'activité du cytosquelette et de l'axonème. Cette observation suggère qu'à la DCM *Pelagomonas* se divise plus rapidement et/ou est

plus motile. La première hypothèse indiquerait que la DCM représente l'environnement de croissance optimale de *P. calceolata*. La seconde mettrait en avant une forte capacité de déplacement de cette microalgue, par exemple pour rechercher des nutriments ou échapper aux prédateurs. En 2015, une analyse métatranscriptomique sur 8 stations du Pacifique nord montrait également une surexpression de transcrits codant pour des dynéines chez *P. calceolata* dans les échantillons de DCM. Les auteurs supposent une motilité accrue de l'algue dans ces échantillons profonds et plus riches en nutriments (Dupont et al. 2015). Longtemps ignorée chez le phytoplancton, la migration verticale aurait été sous-estimée dans les études environnementales (Wirtz et al. 2022). La motilité du phytoplancton dépendrait de facteurs tels que la disponibilité en nutriments et les besoins en luminosité, et jouerait pourtant un rôle essentiel dans la stratification (Cullen 2015). Elle a surtout été étudiée chez les dinoflagellés et les diatomées. Les flagelles, l'accumulation de lipides, la formation d'inclusions minérales et la formation de colonies sont autant de mécanismes permettant aux microorganismes de se déplacer dans la colonne d'eau (Dunne 2022). La méconnaissance de ces mécanismes pourrait impacter les prédictions faites sur l'impact du changement climatique sur le phytoplancton.

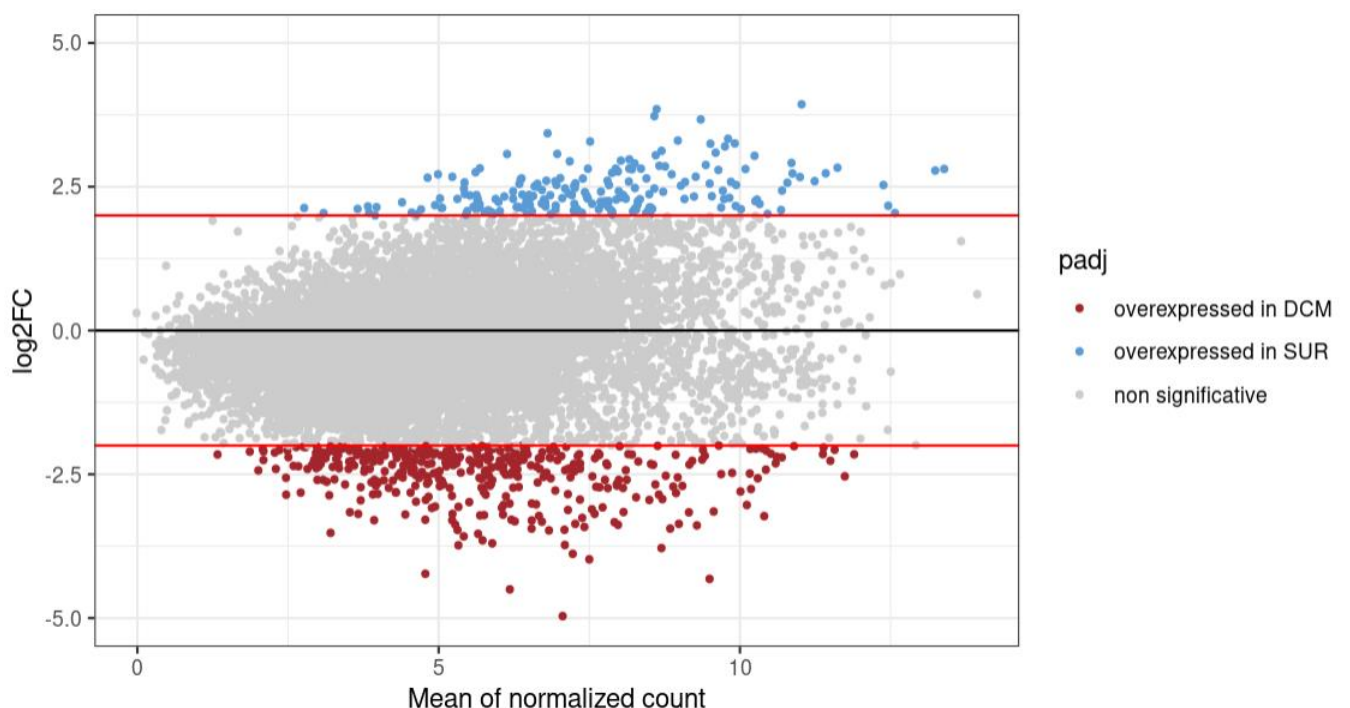


Figure 3. Mantel plot des gènes différentiellement exprimés chez *P. calceolata* en fonction de la profondeur, représenté comme le  $\log_2$  du rapport du niveau moyen d'expression (fold-change) en surface par rapport à celui de la DCM, en fonction du  $\log_{10}$  de l'expression moyenne normalisée. En bleu, les gènes significativement plus exprimés à la surface ( $\log_2$  fold-change  $>2$  et p-value ajustée  $<0.01$ ). En rouge, les gènes significativement plus exprimés à la DCM ( $\log_2$  fold-change  $< -2$  et p-value ajustée  $<0.01$ ). Les résultats complets sont disponibles en Table Supplémentaire 2.

À la surface, les fonctions enrichies sont moins nombreuses, avec 9 termes de la ressource Gene Ontology (GO) et 2 domaines InterProScan. Celles-ci sont assignées à

des activités d'oxydo-réduction, de catalyse, de la néoglucogenèse, de biosynthèse des acides gras et de la chlorophylle, et de traduction de l'ADN (Table Supplémentaire 3). En s'intéressant aux gènes spécifiquement surexprimés à la surface, on constate une abondance de fonctions liées à la biosynthèse des acides gras, des acides aminés ou des protéines, au stress oxydatif, au cycle de Calvin et à la photorespiration (Table Supplémentaire 2). Bien que la surface soit un environnement assez stressant de par la plus forte quantité de lumière et la plus faible quantité de nutriments, *P. calceolata* présente majoritairement une physiologie de biosynthèse de métabolites et de fixation du carbone. Cependant, la surexpression de la photorespiration, un métabolisme coûteux pour la cellule, interpelle et nécessite une étude plus approfondie de ce mécanisme et des gènes impliqués.

## **L'impact de la profondeur sur l'assimilation de l'azote**

Avant de m'intéresser aux cycles métaboliques dont l'expression est modulée par la profondeur, j'ai voulu comparer les résultats de cette analyse de gènes différentiellement exprimés avec ceux générés expérimentalement (voir chapitre 2). J'ai voulu observer si les gènes liés à la réponse aux changements de sources (ammonium, urée, cyanate à la place de nitrate) et de quantité d'azote (441  $\mu\text{M}$  ou 220  $\mu\text{M}$  de nitrate à la place de 882  $\mu\text{M}$ ) étaient modulés selon la profondeur, car la concentration de certains nutriments est un des paramètres qui participe à différencier ces deux environnements (Figure 2, Table Supplémentaire 1). J'ai étudié les gènes communs entre les deux analyses (Figure 4).

Parmi les 541 gènes surexprimés dans 441  $\mu\text{M}$  de nitrate, ou 220  $\mu\text{M}$  de nitrate, ou les deux, 25 sont également surexprimés à la DCM, ce qui représente 6 % de l'ensemble des gènes surexprimés à la DCM. Il s'agit principalement de gènes inconnus ou pour lesquels l'annotation fonctionnelle ne suffit pas pour décrire la fonction. En revanche, parmi les 936 gènes sous-exprimés lors d'une déplétion en nitrate, 108 sont surexprimés à la surface, ce qui représente 53 % de tous les gènes surexprimés à la surface. On retrouve un certain nombre de gènes impliqués dans le cycle de Calvin, la néoglucogenèse, la photorespiration, la réponse au stress oxydatif ou la biosynthèse de protéines (Figure 4A, Table Supplémentaire 4).

Parmi les 650 gènes surexprimés dans au moins une des trois sources alternatives d'azote (ammonium, urée, cyanate), 29 sont en commun avec ceux sous-exprimés à la surface, soit 14 % de tous les gènes surexprimés en surface. On y retrouve notamment la cyanate lyase et un transporteur de formate/nitrite. Parmi les 435 gènes sous-exprimés dans une source alternative d'azote, 83 sont en commun avec les gènes surexprimés à la DCM, soit 19 % de tous les gènes surexprimés à la DCM. Un certain nombre code pour des dynéines, ou pour des gènes inconnus. (Figure 4B, Table Supplémentaire 4)

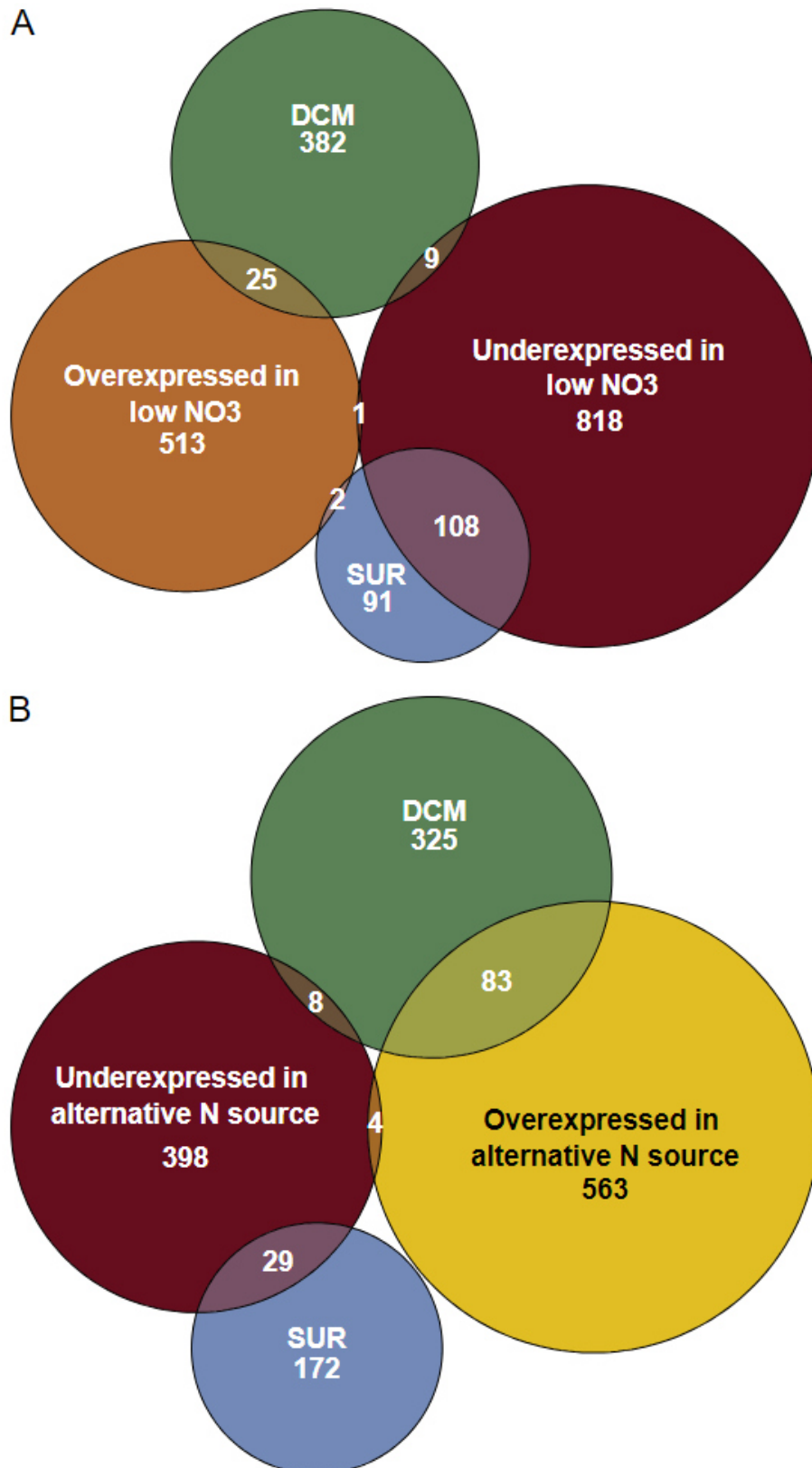


Figure 4. Diagramme d'Euler des gènes de *P. calceolata* différemment exprimés, communs ou spécifiques entre la surface, la DCM, et A) ceux différemment exprimés lorsque la quantité de nitrate diminue, passant de 882  $\mu\text{M}$  à 441  $\mu\text{M}$  ou 220  $\mu\text{M}$ , B) ceux différemment exprimés lorsque le nitrate est remplacé par une source d'azote alternative : l'ammonium, l'urée ou le cyanate.

Le métabolisme de *P. calceolata* en surface présente certaines similarités avec le métabolisme observé lorsque le nitrate est la source d'azote principale, et lorsque sa concentration est en excès. En effet, on retrouve parmi les gènes surexprimés en surface un certain nombre de gènes également surexprimés dans la condition contrôle. En revanche, le métabolisme de *P. calceolata* à la DCM présente des similarités avec celui exprimé lorsque l'on change la disponibilité ou la source d'azote. On retrouve parmi les gènes surexprimés en DCM des gènes également surexprimés lors de la déplétion en nitrate, ou lorsque que le nitrate est remplacé par de l'urée, du cyanate ou de l'ammonium. Selon ces observations, il semblerait que sur le plan de la nutrition en azote, la surface soit un environnement plutôt riche en nitrate, alors que la DCM serait un environnement plutôt pauvre en nitrate et plus riche en source d'azote alternative.

Le phosphate et le nitrite sont les seuls nutriments dont on a les mesures *in situ* présentant une concentration significativement plus abondante à la DCM qu'à la surface (Figure 2, Tableau Supplémentaire 1), donc on ne peut pas affirmer que les concentrations en azote organique (urée, cyanate) et en ammonium soient les principaux paramètres expliquant ces similarités de métabolisme. Cependant, cette observation nous amène à questionner le mode trophique de *P. calceolata* selon la profondeur. La surexpression des gènes de la condition standard en surface pourrait être un signe que *P. calceolata* assimile le nitrate dans cette zone, ce qui est cohérent avec les observations faites lors d'autres études, notamment celle de métatranscriptomique sur 8 échantillons du Pacifique. Les auteurs concluent que *P. calceolata* joue probablement un rôle important dans l'assimilation du nitrate dans les eaux de surface, où elle domine la communauté planctonique eucaryote (Dupont et al. 2015). Également, sa petite taille devrait en faire un organisme plus apte à se développer dans les milieux oligotrophes (Raven 1998). Ainsi, la similarité entre gènes exprimés en surface et en condition standard pourrait être un signe que *P. calceolata* est finalement peu affectée par le manque de nutriments en surface. La mise en place d'un métabolisme de limitation en nitrate et d'utilisation de sources alternatives d'azote à la DCM pourrait plus être un signe de compétition accrue entre les espèces pour le nitrate et d'une concentration plus importante d'azote organique issu de la décomposition de la biomasse (Kirkham et al. 2013). Ainsi, les différences d'acclimatation entre la surface et la DCM pourraient être en partie expliquées par un changement de mode trophique entre ces deux environnements, bien que ce ne soit pas suffisant pour en comprendre la totalité.

## **Le cycle de Calvin et la photorespiration : des métabolismes surexprimés en surface**

Parmi les métabolismes surexprimés en surface par rapport à la DCM, deux sont

particulièrement intéressants : le cycle de Calvin et la photorespiration. Ces deux cycles, liés par leur enzyme initiatrice, la ribulose-1,5-bisphosphate carboxylase/oxygénase (ou Rubisco), sont très étudiés en laboratoire. Les gènes qui les composent sont bien décrits, notamment chez les plantes et les algues vertes comme *Chlamydomonas reinhardtii*. Cependant, ces métabolismes sont rarement étudiés d'un point de vue environnemental.

#### Le cycle de Calvin

Le cycle de Calvin-Benson-Bassham est un cycle biochimique ayant lieu dans le chloroplaste et permettant la fixation du carbone. Grâce à l'énergie générée par la photosynthèse sous forme d'ATP, le cycle de Calvin fixe le dioxyde de carbone pour produire du glucose. Il commence par la carboxylation du ribulose-1,5-bisphosphate (RuBP) avec le CO<sub>2</sub> par la Rubisco. Cette enzyme permet de fixer le carbone dans la RuBP et de générer deux molécules de 3-phosphoglycérate (3PG). Le 3PG est ensuite phosphorylé par la phosphoglycérate kinase (PGK) en 1,3-diphosphoglycérate (DPG). Cette dernière est réduite en glycéraldéhyde-3-phosphate (G3P) par la glycéraldéhyde 3-phosphate déshydrogénase (GAPDH). La G3P est la molécule de glucide générée par le cycle de Calvin, qui va servir à la synthèse de molécules organiques essentiels au métabolisme cellulaire. Le cycle de Calvin se poursuit afin de régénérer le RuBP pour qu'il puisse de nouveau jouer son rôle d'accepteur de CO<sub>2</sub>. Grâce aux activités successives d'une série d'enzymes, la G3P qui n'est pas utilisée pour la glycolyse ou la néoglucogenèse est transformée en ribulose-5-phosphate (RuP). Ce dernier est phosphorylé par la phosphoribulokinase (PRK) en RuBP, permettant de recommencer le cycle.

Chez *P. calceolata*, les gènes codant les enzymes impliquées dans le cycle de Calvin sont pour la plupart différenciellement exprimées selon la profondeur (Figure 5, Table Supplémentaire 5). Ainsi, parmi les 3 gènes codant pour la PGK, 2 sont surexprimés en surface par rapport à la DCM. Sur les 2 gènes codant pour la GAPDH, un seul est détectable dans les données environnementales et il est également surexprimé en surface. Parmi les enzymes permettant la régénération du RuBP, 2 des 7 triosephosphate isomérases, 1 des 2 ribose 5-phosphate isomérases et 1 des 4 ribulose-phosphate 3-épimérases sont significativement plus exprimées en surface qu'à la DCM. Le seul gène codant pour la PRK est aussi surexprimé en surface. Le gène codant pour la Rubisco est exclu de l'analyse, car les gènes chloroplastiques sont en général trop conservés avec les organismes proches et on ne peut pas isoler spécifiquement la Rubisco de *P. calceolata* dans une analyses métatranscriptomique. On constate que sur les 9 enzymes impliquées dans le cycle de Calvin, 6 d'entre elles présentent au moins une copie qui est au moins 4 fois plus exprimée à la surface, et que presque toutes sont au moins 2 fois plus exprimées à la surface qu'à la DCM. Ces résultats semblent indiquer que le cycle de Calvin est surexprimé chez *P. calceolata* lorsqu'elle est à la surface.

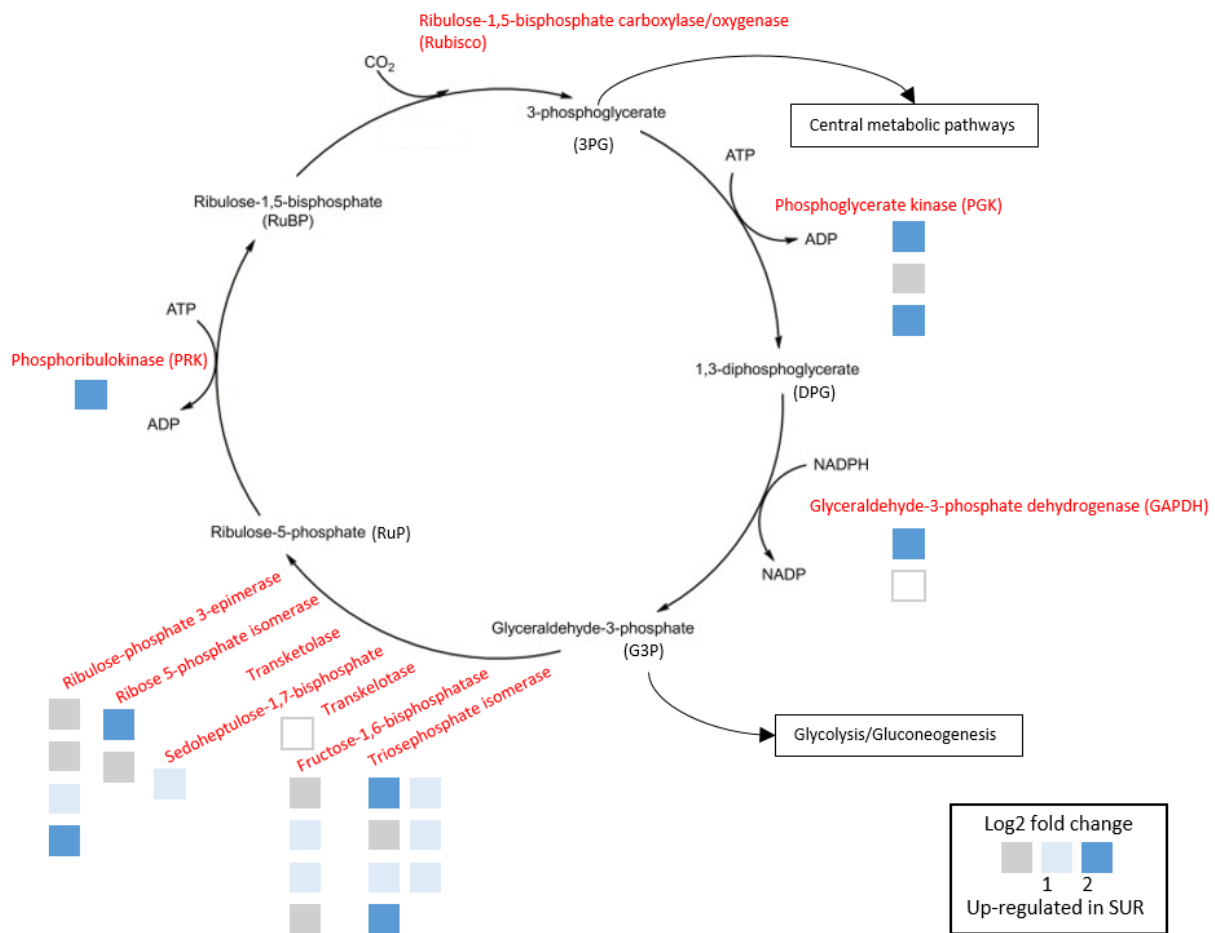


Figure 5. Le cycle de Calvin chez *P. calceolata* et l'expression différentielle des gènes impliqués selon la profondeur. Les molécules sont en noir et les enzymes en rouge. Les modules métaboliques sont encadrés. Chaque carré représente un des gènes codant pour l'enzyme indiquée. Leurs couleurs indiquent le niveau d'expression différentielle de ces gènes en fonction de la profondeur ( $\log_2$  fold-change  $> 1$  ou  $< -1$ , p-value ajustée  $< 0.01$ ). Les carrés blancs indiquent qu'il existe un gène codant pour cette enzyme dans le génome de *P. calceolata* mais que l'expression de celui-ci ne satisfait pas les filtres appliqués sur la quantité de lectures métrascriptomiques. La liste des gènes du cycle de Calvin chez *P. calceolata* est disponible dans la Table Supplémentaire 5.

Chez les plantes terrestres, la lumière joue un rôle important dans la régulation des enzymes impliquées dans ce cycle (Sun et al. 2003). Également, des stress abiotiques comme l'augmentation de la température et la sécheresse peuvent en impacter négativement l'expression (Nouri, Moumeni, and Komatsu 2015). Chez les bactéries, l'importance de la lumière serait moindre, du fait de structures protéiques différentes et moins sensibles aux états redox (Tamoi and Shigeoka 2015). La régulation du cycle de Calvin est moins décrite chez les microalgues. Chez la diatomée *P. tricornutum*, l'impact de la lumière sur la régulation du cycle de Calvin diffère selon les enzymes : la PRK et la GAPDH chloroplastique sont différentiellement régulées par la disponibilité de la lumière, contrairement à la GAPDH cytoplasmique, la fructose-1,6-bisphosphatase et la sedoheptulose-1,7-bisphosphatase (Jensen et al. 2017). Le cycle de Calvin pourrait être sous-exprimé chez les diatomées en réponse à des privations en azote (S. Bender et al. 2014; Alipanah et al. 2015).

En surface, la quantité de lumière est significativement plus importante qu'à la DCM (Figure 2, Table Supplémentaire 1). Cette augmentation de la lumière disponible en surface pourrait être responsable de cette surexpression du cycle de Calvin par *P. calceolata* en surface. Ainsi, la surexpression du cycle de Calvin pourrait être la conséquence d'une luminosité plus intense, fournissant plus d'énergie photosynthétique à la cellule et intensifiant les besoins de fixation du carbone et l'activité de la Rubisco.

#### La photorespiration

La Rubisco, enzyme centrale de la fixation du dioxyde de carbone, est également capable de fixer le dioxygène. Au lieu de scinder le RuBP pour y intégrer le CO<sub>2</sub> et générer deux molécules de 3PG, la Rubisco peut intégrer du O<sub>2</sub> dans le RuBP, générant une molécule de 3PG et une molécule de 2-phosphoglycolate (2PG). L'activité oxygénase de la Rubisco initie la photorespiration. Le 2PG est hydrolysé en glycolate par la phosphoglycolate phosphatase (PGP), avant de quitter le chloroplaste. Une fois exporté dans le peroxyosome, le glycolate est oxydé en glyoxylate par la glycolate oxydase (GOX). L'ajout de L-glutamate par la glycine transaminase permet de générer de la glycine. Cette dernière est exportée dans la mitochondrie où elle est prise en charge dans le système de clivage de la glycine (GCS). Cet ensemble de 4 protéines en interaction intervient en réponse à une augmentation de la concentration de glycine. Il utilise cette dernière et le tétrahydrofolate pour générer du 5,10-méthylènetétrahydrofolate, du CO<sub>2</sub> et de l'ammoniac NH<sub>3</sub>. Dans le cadre de la photorespiration, le GCS est couplé avec la sérine hydroxyméthyltransférase (SHMT), qui convertit le 5,10-méthylènehydrofolate généré par le GCS et la glycine générée par les étapes précédentes de la photorespiration en tétrahydrofolate et en sérine. Cette sérine est de nouveau exportée dans le peroxyosome, puis la sérine-glyoxylate transaminase (SGAT) catalyse sa transformation en hydroxypyruvate et en glycine grâce à l'apport de glyoxylate. L'hydroxypyruvate est oxydé par l'hydroxypyruvate réductase (HPR) en D-glycérate. Ce dernier rejoint le chloroplaste où a lieu la fin du cycle. Il est converti en 3PG par la glycérate kinase. Le cycle de la photorespiration permet donc de générer un deuxième 3PG, qui, avec celui produit par l'oxygénation du RuBP lors de la première réaction du cycle par la Rubisco, permet de déclencher le cycle de Calvin. La photorespiration se met en place lorsque la concentration de CO<sub>2</sub> est insuffisante pour déclencher l'activité carboxylase de la Rubisco. Nécessitant de la lumière et produisant du CO<sub>2</sub>, il s'agit d'un processus coûteux en énergie. Contrairement à la respiration, la photorespiration ne génère pas d'énergie. Et contrairement à la photosynthèse, elle n'amène pas à la production de glucides. Il est estimé que la photorespiration relargue 25 % du carbone fixé lors de la photosynthèse (Busch 2020).

Chez *P. calceolata*, nous avons trouvé des gènes codant pour la plupart des enzymes de la photorespiration (Figure 6, Table Supplémentaire 5), sauf pour la GOX, la glycine transaminase et l'HPR. Sur les 8 gènes codant pour la PGP, 3 ont un log<sub>2</sub> fold-change compris entre 1 et 2, et 2 ont un log<sub>2</sub> fold-change supérieur à 2, indiquant qu'ils sont

surexprimés à la surface par rapport à la DCM. Les enzymes du GCS sont toutes au moins 2 fois plus exprimées dans les échantillons de surface. Les deux gènes codant pour les SHMT sont 2 fois plus exprimés en surface qu'à la DCM. La SGAT et la glycérate kinase, chacune présente en une seule copie, sont toutes les deux significativement surexprimées à la surface qu'à la DCM. Dans l'ensemble, les gènes codant pour des enzymes de la photorespiration sont plus exprimés à la surface qu'à la DCM, indiquant que chez *P. calceolata*, ce métabolisme est significativement plus actif quand l'algue est en surface.

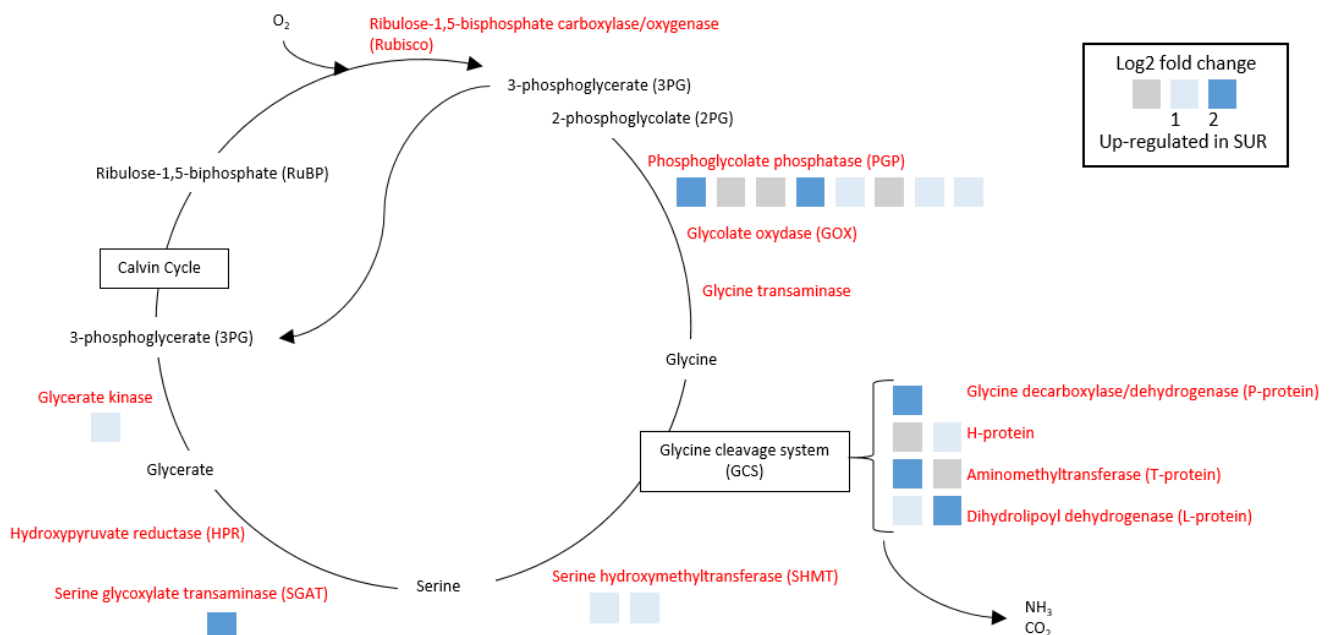


Figure 6. La photorespiration chez *P. calceolata* et l'expression différentielle des gènes impliqués selon la profondeur. Les molécules sont en noir et les enzymes en rouge. Les modules métaboliques sont encadrés. Chaque carré représente un des gènes codant pour l'enzyme indiquée. Les carrés et leurs couleurs indiquent le niveau d'expression différentielle de ces gènes en fonction de la profondeur ( $\log_2$  fold change  $> 1$  ou  $< -1$ , p-value ajustée  $< 0.01$ ). La liste des gènes de la photorespiration chez *P. calceolata* est disponible dans la Table Supplémentaire 5.

Puisqu'elle est coûteuse en énergie, on peut donc se demander pourquoi ce métabolisme est surexprimé en surface. Plusieurs conditions peuvent amener à une augmentation de la photorespiration chez les plantes et les microalgues. Une élévation de la température peut enclencher la photorespiration. En effet, à des températures plus élevées, la Rubisco est moins performante pour distinguer le CO<sub>2</sub> et l'O<sub>2</sub>. De plus, la chaleur rend le CO<sub>2</sub> moins soluble, ce qui en diminue la concentration dans le chloroplaste par rapport à l'O<sub>2</sub>. Puisque la température est significativement plus importante à la surface qu'à la DCM (Figure 2), on peut se demander si ce n'est pas cette augmentation qui induit l'intensification de la photorespiration en surface. Également, la photorespiration pourrait agir comme un mécanisme de protection contre une forte luminosité, en dissipant l'énergie excessive et en atténuant les effets nocifs des produits de réactions photochimiques qui s'accumulent. Chez les algues vertes par exemple, la photorespiration aurait un rôle de régulateur de la photosynthèse, protégeant le métabolisme contre les toxines (Segura Broncano et al.

2023). Puisque la quantité de lumière est significativement plus importante en surface qu'à la DCM (Figure 2), l'intensification de la photorespiration afin de profiter de son rôle photoprotecteur paraît être une hypothèse raisonnable. Le principal facteur amenant à l'activité oxygénase de la Rubisco est une plus faible disponibilité en  $\text{CO}_2$  par rapport à l' $\text{O}_2$ . Chez *C. reinhardtii*, la photorespiration est généralement surexprimée en réponse à une limitation du  $\text{CO}_2$ , et serait le métabolisme carboné le plus significativement modulé face à ce stress (Wang, Stessman, and Spalding 2015; Burlacot and Peltier 2023). Dans notre jeu de données, la concentration en oxygène dissous est significativement plus importante en surface, et à l'inverse, la concentration d'ions carbonate et la quantité de carbone sont plus élevées à la DCM (Figure 2). Si la surface est un milieu plus oxygéné et appauvri en carbone, on peut supposer que c'est cela qui induit cette surexpression de la photorespiration à la surface.

#### Les mécanismes de concentration du carbone

Pour favoriser le cycle de Calvin plutôt que la photorespiration, il faut que le site actif de la Rubisco présente une concentration de  $\text{CO}_2$  suffisante par rapport à celle d' $\text{O}_2$  pour éviter l'activation de son activité oxygénase. Chez les plantes et les microalgues, les mécanismes de concentration du carbone (CCM) permettent d'améliorer les performances photosynthétiques en augmentant la concentration de  $\text{CO}_2$  autour de la Rubisco, améliorant la fixation du carbone et inhibant la photorespiration. Les CCM sont une série de transporteurs actifs du carbone inorganique ( $\text{CO}_2$ ,  $\text{H}_2\text{CO}_3$ ,  $\text{HCO}_3^-$ ,  $\text{CO}_3^{2-}$ ) qui permettent de condenser le carbone autour de la Rubisco, particulièrement dans le pyrénioïde. Il s'agit d'une microstructure interne au chloroplaste, analogue au carboxysome cyanobactérien, contenant une centaine d'enzymes, y compris la Rubisco. Présent chez les algues vertes comme *C. reinhardtii* et chez les diatomées *P. tricornutum* et *T. pseudonana*; le pyrénioïde a également été décrit chez beaucoup d'espèces de Pélagophytes mais pas chez *P. calceolata* (Andersen et al. 1993; Wetherbee et al. 2021; Burlacot and Peltier 2023). D'abord, le carbone inorganique traverse les membranes plasmiques et chloroplastiques. Le passage du  $\text{CO}_2$  se fait par diffusion, alors que celui des ions bicarbonate nécessite l'activité des transporteurs de bicarbonate ( $\text{HCO}_3^-$ ). Les transporteurs de formate/nitrite pourraient également être capables de transporter le  $\text{HCO}_3^-$ , avec une affinité plus faible (Burlacot and Peltier 2023). La régulation du pH dans les différents compartiments permet de stocker le carbone sous forme soit de  $\text{CO}_2$ , soit de  $\text{HCO}_3^-$ : une augmentation du pH induit la transformation du  $\text{CO}_2$  en  $\text{HCO}_3^-$  et inversement. L'interconversion de ces deux formes est également métabolisée par les anhydrases carboniques (CAH), des enzymes qui permettent l'équilibrage des espèces de carbone inorganique à un pH donné. Ces enzymes permettent de modifier les gradients de  $\text{CO}_2$  entre 2 compartiments, facilitant ainsi son passage par diffusion. Les bestrophines sont des protéines formant des canaux anioniques dans les membranes des thylakoïdes, permettant le transport du  $\text{HCO}_3^-$  depuis le stroma des chloroplastes jusqu'au lumen, à partir duquel il est converti en  $\text{CO}_2$  par une CAH avant d'entrer dans le pyrénioïde pour saturer la Rubisco et ainsi bloquer l'activation de la photorespiration. Ces mécanismes sont représentés

schématiquement sur la Figure 7 chez les diatomées *P. tricornutum* et *T. oceanica*. Chez *P. calceolata*, nous avons identifié 13 gènes codant pour des bestrophines, 5 transporteurs de formate/nitrite, 7 CAH et 2 transporteurs d'ions bicarbonates (Table Supplémentaire 5).

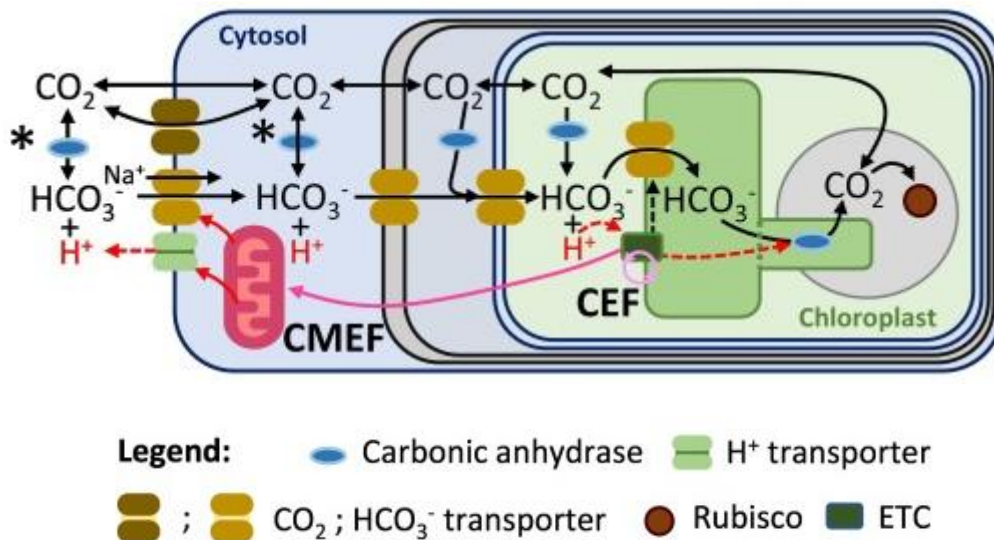


Figure 7. Les mécanismes de concentration du carbone (CCM) chez les diatomées *P. tricornutum* et *T. oceanica*. Appartenant aux Ochrophytes, il s'agit des organismes les plus proches de *P. calceolata* pour lesquels les CCM sont décrits. Les transporteurs de HCO<sub>3</sub><sup>-</sup> permettent l'entrée du carbone inorganique dans la cellule. Les anhydrases organiques transforment le CO<sub>2</sub> en HCO<sub>3</sub><sup>-</sup> et inversement selon le pH intracellulaire. Toutes les diatomées ne possèdent pas d'anhydrases carboniques extracellulaires ou cytoplasmiques. Les bestrophines permettent l'entrée du carbone inorganique du stroma au lumen (non représentés) et le gradient de pH (flèche rouge pointillé) fait passer le carbone dans le pyrénioïde. ETC=chaîne de transport des électrons, CMEF=flux d'électron du chloroplaste à la mitochondrie, CEF=flux d'électron cyclique. D'après Burlacot and Peltier 2023. La liste des gènes des CCM chez *P. calceolata* sont disponible dans la Table Supplémentaire 5.

Les CCM, la photorespiration et le cycle de Calvin sont liés : le premier limitant le second et optimisant le troisième. Par exemple, chez *Chlamydomonas reinhardtii*, la photorespiration est induite de manière temporaire en présence de faibles concentrations de CO<sub>2</sub>, correspondant au temps nécessaire à l'induction des CCM (Wang, Stessman, and Spalding 2015; Burlacot and Peltier 2023). Ainsi la surexpression de la photorespiration en surface pourrait être un signe que les CCM ne sont pas suffisamment efficaces pour saturer la Rubisco en CO<sub>2</sub> et empêcher son activité oxygénase, contrairement à la DCM. Également, la photorespiration pourrait agir comme une boucle rétroactive transitoire, permettant de contrôler le niveau d'activité des CCM (Burlacot and Peltier 2023). Les interactions entre ces métabolismes sont très étudiées, mais beaucoup de choses restent à comprendre à leurs sujets, particulièrement comment les conditions environnementales peuvent moduler leurs expressions.

Niveau d'expression des enzymes impliquées dans la photorespiration, le cycle de Calvin et des CCM dans l'environnement

En complément des résultats présentés ci-dessus, nous avons étudié les niveaux d'expression de la totalité de ces gènes dans les échantillons *Tara Oceans* (Figure 8). Sur l'ensemble des gènes étudiés, les plus exprimés en transcrits par millions sur l'ensemble des données codent pour une CAH, ou sont impliqués dans le cycle de Calvin : une phosphoglycérate kinase (PGK) et une phosphoribulokinase (PRK). Pour chaque catégorie d'enzymes, on constate des niveaux d'expression très variables. En général, ceux ayant une expression plus importante sont aussi ceux qui sont différentiellement exprimés selon la profondeur. C'est vrai pour les CAHs, les transporteurs de formate/nitrite et les protéines du GCS (Table Supplémentaire 5).

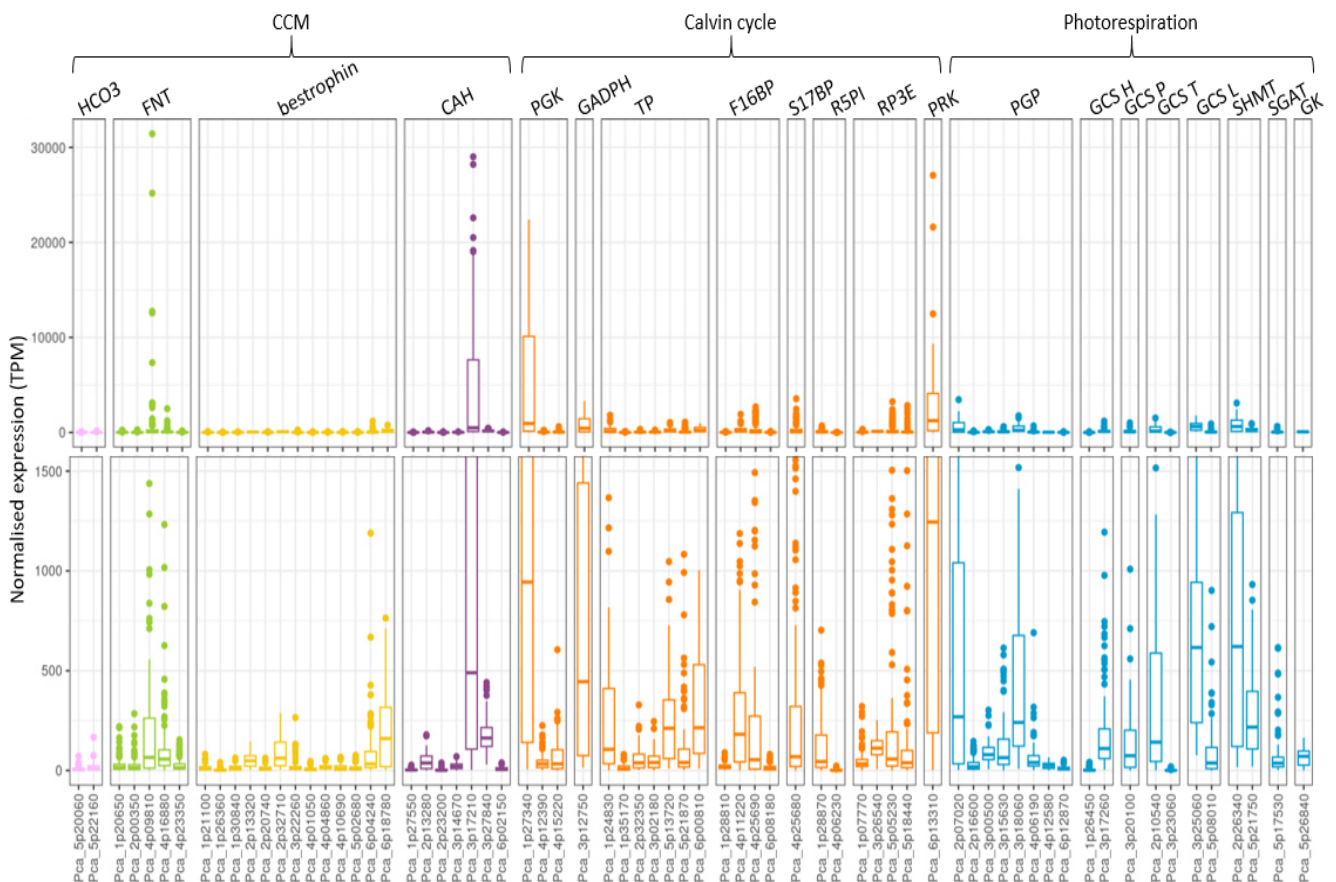


Figure 8. Expression normalisée en transcrit par millions (TPM) des gènes impliqués dans les CCM, le cycle de Calvin et la photorespiration chez *P. calceolata* dans les données *Tara Oceans*. Le volet du haut montre l'expression globale, et celui du bas l'expression entre 0 et 1500 TPM. Rose : transporteur d'ion bicarbonate ( $\text{HCO}_3^-$ ). Vert : transporteur de formate/nitrite (FNT). Violet : anhydrase carbonique (CAH). Orange : cycle de Calvin (PGK : phosphoglycérate kinase, GAPDH : glyceraldéhyde-3-phosphate déshydrogénase, TP : triosephosphate isomérase, F16BP : fructose-1,6-bisphosphatase, S17BP : sedoheptulose-1,7-bisphosphatase, R5PI : ribose 5-phosphate isomérase, RP3E : ribulose-phosphate 3-épimérase, PRK : phosphoribulokinase). Bleu : photorespiration (PGP : phosphoglycolate phosphatase, GCS\_H : protéine H du système de clivage de la glycine, GCS\_P : protéine P du système de clivage de la glycine ou glycine décarboxylase/déshydrogénase, GCS\_T : protéine T du système de clivage de la glycine ou aminométhyltransférase, GCS\_L : protéine L du système de clivage de la glycine ou dihydrolipoyl déshydrogénase, SMHT : sérine/glycine hydroxyméthyltransférase, SGAT : sérine glyoxylate transaminase, GK : glycérate kinase). Les valeurs d'expressions normalisés en TPM sont disponibles dans la Table Supplémentaire 5.

Nous avons ensuite voulu voir quels sont les profils d'expression de ces gènes d'intérêt dans l'ensemble des données *Tara Oceans*, nous les avons donc clusterisés. Il y a une séparation très nette des profils d'expression (Figure 9). Les échantillons environnementaux sont clusterisés en 2 groupes : le groupe 1 contient en majorité des échantillons de surface (51 en tout, dont 5 provenant de la DCM) tandis que le groupe 2 contient 22 échantillons de surface et 46 échantillons de DCM. Les gènes d'intérêt sont répartis en 4 groupes, les clusters A, B, C et D.

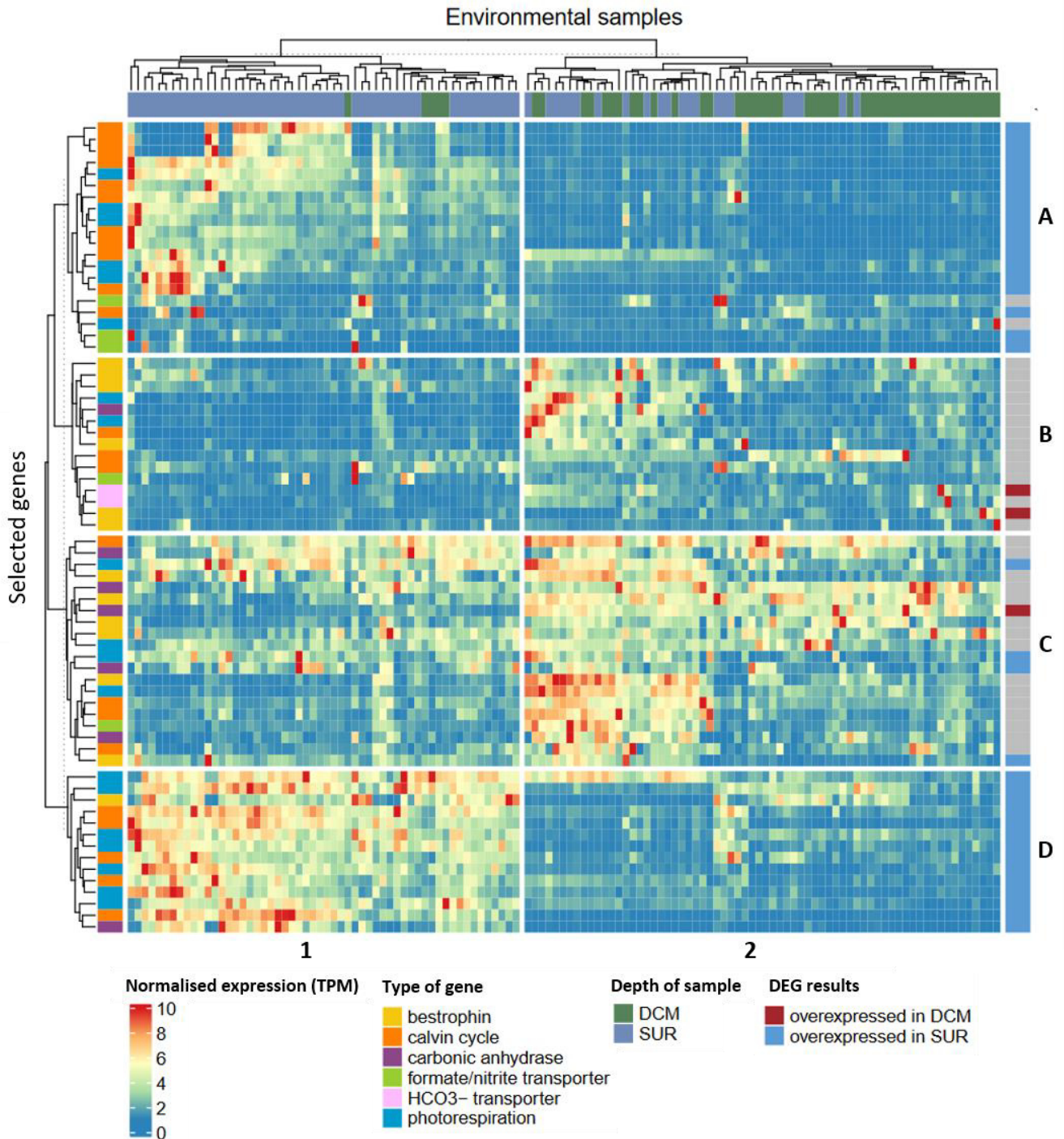


Figure 9. Heatmap représentant l'expression des gènes impliqués dans les CCM, le cycle de Calvin et la photorespiration dans l'ensemble des données *Tara Oceans*, normalisée en TPM entre 1 et 10. Le clustering a été réalisé avec une table de distance euclidienne et la méthode complète, et le calcul des k-moyennes a été fait avec la méthode Elbow (Figure Supplémentaire 1).

Les 35 gènes du cluster A et D sont plutôt exprimés dans le groupe d'échantillons 1 (surface), et assez peu exprimés dans le groupe 2. Ils sont principalement des gènes impliqués dans le cycle de Calvin et dans la photorespiration. Parmi les 20 gènes du cluster A, il y a respectivement 11 et 6 gènes de ces processus, et 3 transporteurs de formate/nitrite. Parmi les 15 gènes du cluster D, 5 codent pour des enzymes du cycle de Calvin, 8 pour des enzymes de la photorespiration, 1 pour une anhydrase carbonique et 1 bestrophine. Malgré un profil d'expression similaire, l'expression dans le groupe 1 est plus intense. Tous (à part un transporteur de formate/nitrite et une phosphoglycolate phosphatase) sont significativement plus exprimés en surface, et 16 ont un log2fold change supérieur à 2, donc sont au moins 4 fois plus exprimés en surface qu'en DCM. L'expression dans des échantillons de surface de gènes impliqués dans le cycle de Calvin et dans la photorespiration est cohérente avec les observations faites précédemment, ce qui confirme que ces deux métabolismes sont plus fortement exprimés en surface qu'à la DCM. Les 3 transporteurs de formate/nitrite présents dans ce cluster sont fortement modulés en fonction de la concentration et de la source d'azote, comme nous l'avons montré dans le chapitre 2. Ainsi, ces gènes pourraient être modulés par les différences de concentration en nutriments entre surface et DCM.

Les 14 gènes du cluster B sont plutôt exprimés dans le groupe 2, et faiblement dans le groupe 1. Dans l'ensemble, l'expression de ces gènes est plus faible que celle des autres clusters. On retrouve 6 des 13 bestrophines de *P. calceolata*, les 2 transporteurs d'ions bicarbonate, un transporteur de formate/nitrite et une anhydrase carbonique. Cette dernière, ainsi que l'une des bestrophines, sont 2 fois plus exprimées en DCM qu'en surface. On y retrouve également un gène de la photorespiration et 3 du cycle de Calvin, qui sont aussi très faiblement exprimés par rapport aux autres gènes de leurs cycles respectifs (Figure 8). L'expression des bestrophines est en miroir par rapport à l'expression des gènes de la photorespiration et du cycle de Calvin du groupe A : elle est plus faible dans les échantillons de surface par rapport à ceux de la DCM. Si les CCM ont une expression diminuée en surface, cela peut en partie expliquer l'intensification de la photorespiration. On peut se demander quelles sont les conditions environnementales qui induisent l'expression des bestrophines en DCM plutôt qu'en surface, alors que celle-ci est appauvrie en CO<sub>2</sub>.

Les 20 gènes du clusters C sont principalement des bestrophines (6 sur 13), des anhydrases carboniques (5 sur 7), 4 phosphoglycolate phosphatase, 4 enzymes du cycle de Calvin et 1 transporteur de formate/nitrite. Ces gènes sont exprimés assez fortement dans l'ensemble des échantillons de notre jeu de données, bien qu'un peu plus que dans le groupe 2. Deux des PGP sont différentiellement exprimés en surface, ainsi que l'une des bestrophines et une des anhydrases carboniques. Une autre est significativement plus exprimée à la DCM. Il semble que la moitié des bestrophines et la quasi-totalité des anhydrases carboniques ont une expression importante mais indépendante de la profondeur d'échantillonnage. Ainsi, ces CCM ne seraient pas véritablement impactées par la profondeur. Les paramètres environnementaux impactant leur expression restent à déterminer.

Ainsi, ces résultats confirment la régulation à la hausse de la plupart des enzymes impliquées dans le cycle de Calvin et la photorespiration à la surface. À l'inverse, les bestrophines, les transporteurs d'ions bicarbonates et les anhydrases carboniques suivent un schéma d'expression assez différent. Elles sont soit plutôt exprimées en DCM, soit exprimées dans l'ensemble des échantillons, sans distinction entre les différentes profondeurs. Ce profil d'expression des CCM semble indiquer qu'elles sont moins exprimées en surface.

## 4 - Conclusion

Ce travail préliminaire sur l'acclimatation de *P. calceolata* aux variations de profondeur ouvre la porte à un certain nombre de réflexions intéressantes. Bien que la profondeur ait un très fort impact sur l'expression des gènes, il est difficile de déterminer quels paramètres environnementaux précisément poussent *P. calceolata* à s'acclimater différemment entre ces profondeurs. La température, la concentration en nutriments et la luminosité sont les paramètres environnementaux les plus pertinents, bien qu'il soit probable que ce soit la synergie entre eux qui poussent *P. calceolata* à réguler son métabolisme. Les interactions biotiques sont sûrement d'une importance capitale aussi, mais elles ne peuvent pas être étudiées avec les données à notre disposition.

À la DCM, *P. calceolata* surexprime principalement des gènes codant pour des dynéines, ainsi que pour des gènes précédemment identifiés comme participant à l'acclimatation à la déplétion en nitrate et aux changements de source d'azote. On peut supposer que *P. calceolata* fait donc face à des conditions nécessitant une motilité accrue, par exemple comme stratégie pour faire face au manque de nitrate et pour assimiler d'autres sources d'azote. La DCM est en général plus riche en nutriments que la surface, ce profil d'expression est donc surprenant, mais pourrait être expliqué en partie par une intensification des interactions biotiques, comme par exemple une compétition pour l'accès aux ressources qui pousserait *P. calceolata* à limiter son assimilation de nitrate et à opter pour des sources d'azote organiques.

À la surface, *P. calceolata* exprime principalement des gènes codant pour la biosynthèse des acides gras, de la chlorophylle et pour la traduction de l'ARN. Également, une grande part des gènes différentiellement exprimés sont communs avec ceux qui sont différentiellement exprimés dans des conditions de culture standard par rapport à du stress nutritif. Il semble donc que *P. calceolata* présente en surface un métabolisme "en bonne santé" dans lequel l'azote est en quantité suffisante. Cela pourrait s'expliquer par l'efficacité propre aux picoeucaryotes photosynthétiques à faire face à des conditions oligotrophes en étant moins impactés que les espèces de plus grande taille. Si le stress azoté ne semble pas toucher *P. calceolata* dans les eaux de surface, on constate toutefois une surexpression des protéines impliquées dans la réponse au stress oxydatif, ce qui montre que *P. calceolata* souffre tout de même des conditions de vie en surface, potentiellement du stress lumineux ou thermique.

On constate une forte surexpression des enzymes impliquées dans le cycle de Calvin et la photorespiration lorsque *P. calceolata* est en surface. Les paramètres environnementaux régulant l'expression des gènes de ces métabolismes chez les microalgues sont encore peu décrits, bien que la température et la luminosité soient des candidats pertinents. La concentration de CO<sub>2</sub> et d'O<sub>2</sub> joue probablement un rôle aussi. Les gènes impliqués dans les CCM sont soit exprimés avec peu de distinction selon la profondeur, soit exprimés plus fortement dans les échantillons de DCM. Ce profil d'expression participe à expliquer l'importance de la photorespiration en surface, là où les CCM sont moins exprimés. Il reste cependant difficile à ce stade d'identifier avec précision et certitude les paramètres environnementaux modulant l'expression des CCM, de la photorespiration et du cycle de Calvin.

La difficulté principale pour tirer des conclusions biologiques fortes sur ces analyses est la complexité des données environnementales. D'une part, il est difficile de démêler l'impact de paramètres individuels sur l'expression d'un gène. Les variations sont multivariées, et intriquées de manière complexe. Seules des expérimentations menées en laboratoire permettront de confirmer ou d'infirmer les observations faites dans ce chapitre. D'autre part, si les résultats sont assez lisibles pour les gènes surexprimés en surface, il est plus difficile de voir un schéma clair émerger pour les gènes différentiellement exprimés à la DCM. En effet, il est difficile de considérer celle-ci comme un ensemble d'environnements uniformes alors que la profondeur d'échantillonnage varie entre 5 et 200 m (Figure Supplémentaire 2). Contrairement à la surface qui est relativement homogène, la DCM est plutôt une succession de couches différentes entre lesquelles les conditions environnementales et la composition de la communauté microbienne varient grandement (Cabello et al. 2016). De plus, les aléas du terrain font que pour certains échantillons, il était impossible de déterminer la profondeur de la DCM. Cela a probablement amené à nommer certains échantillons ainsi alors qu'ils ont été prélevés à une autre profondeur. Ainsi, il faudrait réexaminer les données pour déterminer si les échantillons de DCM avec lesquelles nous travaillons ont été identifiés comme tels.

La suite à donner à ces analyses est principalement expérimentale. L'étude précise de la réponse de *P. calceolata* à des variations de température, de luminosité ou de concentration de CO<sub>2</sub> devrait permettre de déterminer lequel de ces paramètres agit le plus sur l'expression des gènes. Cela permettrait de mieux comprendre quels paramètres environnementaux impactent la régulation du cycle de Calvin et de la photorespiration, et quelles conditions environnementales contraignent le métabolisme des picoeucaryotes photosynthétiques à la surface et à la DCM. Puisqu'il est prédit que la stratification de la colonne d'eau va s'intensifier avec le réchauffement climatique, il est pertinent de comprendre d'un point de vue fonctionnel quels sont les stress propres à ces profondeurs, et comment les micro-organismes s'y acclimatent.

## Discussion & Perspectives

---

Dans cette dernière partie, je vais discuter globalement les résultats de cette thèse dans le contexte plus large de la réponse du phytoplancton et des PPE au changement climatique. Ensuite, je reviendrai sur l'importance des interactions biotiques dans l'assimilation des nutriments, aspect peu exploré dans ce travail. Pour finir, je discuterai en quoi *Pelagomonas calceolata* a le potentiel de devenir un organisme modèle pertinent pour étudier l'acclimatation des PPE aux changements environnementaux, particulièrement dans le cadre du couplage des études *in situ* et *in vitro*, permettant de surmonter les biais inhérents à la génomique environnementale d'une part et à la biologie moléculaire et cellulaire traditionnelle d'autre part.

## 1 - L'impact du changement climatique sur le métabolisme de l'azote des PPE

L'étude de l'acclimatation de *P. calceolata* m'a permis de comprendre les stratégies développées par cette microalgue en réponse aux changements de disponibilités de l'azote. Une partie de ces mécanismes, comme la diminution des besoins en azote et le recyclage de l'azote intracellulaire par la dégradation des protéines et des acides aminés, étaient déjà connus et décrits chez des espèces relativement proches, comme les diatomées ou le Pélagophyte *Aureococcus anophagefferens*. D'autres, en revanche, sont relativement nouveaux. Ainsi, en croisant l'étude des gènes différentiellement exprimés dans l'environnement avec ceux identifiés dans des conditions expérimentales, j'ai pu suggérer un rôle détoxifiant de la cyanate lyase plutôt qu'une activité nécessaire au métabolisme d'une source d'azote organique extracellulaire. Ainsi, j'ai acquis au cours de cette thèse une meilleure compréhension de l'acclimatation de *P. calceolata* aux changements de quantité et de source d'azote.

On peut essayer d'anticiper comment *P. calceolata* va s'acclimater à ces futures conditions. L'étude de la biogéographie actuelle de *P. calceolata* a permis de prédire sa future distribution d'ici à 2100 (Chapitre 1, Figure 3). La méthodologie développée par Paul Frémont et al. prédit que son abondance relative devrait augmenter de 1.12 % aux latitudes les plus élevées (régions tempérées et subpolaires) et devrait diminuer de 0.8 % dans les eaux tropicales (Frémont et al. 2022). La température de l'eau dans les zones les plus froides est un facteur limitant pour la croissance du phytoplancton. Ainsi, son augmentation permettra à *P. calceolata* de coloniser ces eaux qui lui étaient auparavant inaccessible. Ce schéma d'évolution de l'abondance est aussi cohérent avec un changement de la concentration du nitrate disponible. Il devrait diminuer dans les zones tropicales à cause de l'intensification de la stratification, mais devrait augmenter dans les zones polaires, car l'augmentation de la température peut provoquer au contraire une inhibition de la stratification des eaux froides (Ardyna and Arrigo 2020).

On peut également essayer de prédire comment la modification du cycle de l'azote marin peut affecter la distribution de *P. calceolata*. Les modèles prédisent une diminution de la disponibilité des nutriments en réponse à l'intensification de la stratification, le réchauffement de l'eau et l'acidification. L'aire des zones oligotrophes a déjà augmenté de 0,8 à 4,3 % par an depuis 1998 (Polovina, Howell, and Abecassis 2008) et l'apport en nutriments depuis les eaux profondes vers la surface devrait diminuer de 9 à 14 % d'ici à la fin du siècle (Shulka et al. 2019). L'acidification devrait particulièrement affecter les cyanobactéries diazotrophes, intensifiant la fixation de l'azote (Wannicke et al. 2018). La tendance générale semble indiquer un changement dans la répartition de l'azote, favorisant l'ammonium grâce à l'augmentation de la fixation de l'azote due à l'acidification et au réchauffement, tandis que le nitrate et le nitrite sont désavantagés en raison de l'intensification de la stratification.

L'augmentation de la taille des zones oligotrophes devrait favoriser le développement

des PPE par rapport au phytoplancton de plus grande taille (Shuter 1978; Raven 1998; Litchman et al. 2007). La diminution de la quantité et du type d'azote disponible ne sera cependant pas sans conséquence sur la physiologie de ces PPE. Une déplétion en nitrate peut grandement affecter le taux de croissance de *P. calceolata*, car la biosynthèse de protéines est ralentie pour économiser l'azote. En termes de physiologie, le recours au recyclage de l'azote intracellulaire devrait être plus courant dans les eaux oligotrophes. Dans l'ensemble, si la quantité de nitrate disponible diminue comme le prévoient les modèles climatiques, et même si les PPE devraient augmenter en abondance par rapport aux autres phytoplanctons, leurs physiologies et leurs croissances seront tout de même impactées négativement.

Si l'augmentation de la fixation de l'azote amène à une augmentation de la concentration en ammonium, l'effet sur les PPE dépendra de l'espèce. En effet, la plupart des microalgues assimilent efficacement l'ammonium, mais pas les Pelagophyceae comme *P. calceolata* ou *A. anophagefferens* (Berg et al. 1997; Ou et al. 2018). Ainsi, cette classe en particulier pourrait être défavorisée par l'augmentation de l'ammonium comme source d'azote inorganique principale. Je suppose que, même si la diminution de l'azote inorganique devrait favoriser le développement des PPE par rapport aux organismes photosynthétiques plus gros, leur physiologie sera tout de même très touchée. De plus, l'augmentation possible des concentrations d'ammonium pourrait ne pas compenser l'effet de la perte de nitrate prévue pour l'ensemble des classes de PPE.

Des apports en azote organique pourraient contrebalancer ces effets. *P. calceolata* a une bonne croissance quand le nitrate est remplacé par du cyanate ou de l'urée, et même si nous ne comprenons pas encore l'ensemble des mécanismes à l'œuvre dans l'assimilation de l'azote organique chez les PPE. Ces sources alternatives pourraient avoir un rôle important dans l'environnement. Il est difficile de prévoir comment les concentrations d'urée et de cyanate vont évoluer en réponse au changement climatique. Si l'on exclut les sources anthropogéniques, les sources principales d'urée et de cyanate dans l'océan sont l'excrétion par les organismes de molécules azotées dissoutes (DON) ou la dégradation de la matière organique dissoute (DOM) (Sipler and Bronk 2015; Widner, Mulholland, and Mopper 2016). Des expériences en mésocosme ont montré que l'augmentation de la température pouvait amener à une augmentation de la quantité de DOM libérée par le phytoplancton (Morán et al. 2006; Wohlers et al. 2009; Engel et al. 2011). Les PPE, avec leurs petites tailles, ont un ratio surface:volume favorisant la diffusion de la DOM en dehors de la cellule (Thornton 2014). Donc, on pourrait constater une augmentation de la DON libérée par exsudation de matière organique en réponse à une augmentation de la température. Cependant, cette augmentation de la DON ne devrait pas avoir d'impact à l'échelle globale, puisque la biomasse phytoplanctonique totale devrait diminuer dans les prochaines décennies (Polovina, Howell, and Abecassis 2008; Boyd and Ellwood 2010; Shulka et al. 2019). Toutefois, cette augmentation de la production de DON pourrait permettre de compenser la diminution de la concentration en azote inorganique à l'échelle locale,

permettant à des PPE comme *P. calceolata* d'assimiler de l'urée, du cyanate ou des acides aminés en remplacement du nitrate ou de l'ammonium devenu rare. L'importance de la mixotrophie commence tout juste à être révélée, alors qu'elle pourrait avoir un rôle essentiel dans l'acclimatation du phytoplancton aux changements environnementaux (Faure et al. 2019).

## 2 - De l'importance des interactions biotiques

L'un des résultats majeurs de cette thèse est la découverte de la capacité à assimiler l'azote organique chez *P. calceolata*. Déjà décrite chez l'espèce *A. anophagefferens*, cette capacité permet à cette dernière de former des efflorescences malgré de faibles concentrations en azote inorganique. L'urée, les acides aminés et les purines sont autant de sources d'azote organique alternatives (Berg et al. 1997; Gry Mine Berg, Repeta, and Laroche 2002; Wurch et al. 2011; Wurch, Gobler, and Dyhrman 2014). Comme je l'ai montré dans le chapitre 1, *P. calceolata* possède un ensemble de gènes lui permettant d'assimiler l'urée (transporteur d'urée et uréase) et le cyanate (cyanate lyase). Dans le chapitre 2, j'ai montré que *P. calceolata* avait une croissance sur le cyanate et l'urée équivalente à celle dans la condition standard, riche en nitrate. J'ai découvert que la cyanate lyase était surexprimée dans des conditions de déplétion en nitrate, *in situ* et *in vitro*, mais sous-exprimée lorsque le cyanate était la source d'azote. Ainsi, j'ai pu en déduire que la cyanate lyase chez *P. calceolata* n'avait pas de rôle dans l'assimilation du cyanate comme source d'azote organique, mais plutôt dans la détoxification du cyanate intracellulaire généré lors de la dégradation des acides aminés en conséquence de la privation en azote. Des études récentes ayant montré la prévalence du gène codant pour la cyanate lyase, mais questionnant son rôle chez les PPE soutiennent nos conclusions (Mao et al. 2022; Sato, Hashihama, and Takeda 2023).

Cependant, j'ai éclipsé un élément pourtant essentiel au cours de cette analyse : le rôle potentiel des micro-organismes présents autour de *P. calceolata*, la phycosphère. En effet, la souche utilisée pendant cette thèse est non-axénique, c'est-à-dire qu'elle est accompagnée d'un consortium bactérien dont nous ignorons la composition. J'ai préalablement trié la souche au cytomètre en flux afin d'en diminuer la charge bactérienne, mais celle-ci demeure, et impacte possiblement la croissance et la physiologie de la microalgue. Ainsi, la phycosphère de *P. calceolata* pourrait jouer un rôle dans l'assimilation de sources d'azote organiques. Dans le cas où l'urée sert de source d'azote, nous constatons la surexpression de l'uréase, ce qui nous conforte dans l'idée que *P. calceolata* assimile l'urée par elle-même. En revanche, puisque la cyanate lyase semble ne pas avoir de rôle dans l'assimilation du cyanate, cela signifie que nous ne connaissons pas les gènes de cette voie chez *P. calceolata*, ni chez les autres PPE. Il serait possible que le phycosphère transforme le cyanate en une forme d'azote assimilable. J'exclus que le cyanate soit métabolisé en ammonium, car la croissance de *P. calceolata* sur l'ammonium est très diminuée, alors que celle sur le cyanate est peu impactée. Cependant, l'impact de la phycosphère sur la physiologie de *P. calceolata* pourrait être très important.

En plus de jouer potentiellement un rôle dans l'assimilation des nutriments, il pourrait ralentir ou stimuler la croissance de *P. calceolata* dans certaines conditions, lui apporter des métabolites, accaparer des métabolites importants, etc. Il est essentiel de rendre axénique la souche et de reproduire nos expériences afin de quantifier l'impact de cette phycosphère. Il est aussi important de caractériser les espèces du consortium bactérien, leurs rôles et leurs interactions, afin de comprendre comment cette phycosphère agit sur l'algue et comment ses partenaires communiquent et partagent ou non les ressources. L'alignement des lectures de séquençage sur le transcriptome de référence a permis d'éliminer les gènes bactériens de nos analyses, mais ceux-ci ont potentiellement eu un impact significatif sur l'assimilation de l'azote de *P. calceolata*. Identifier et caractériser les acteurs du consortium bactérien permettra de définir le rôle exact de ce phycosphère, et les interactions microalgues/bactéries et bactéries/bactéries qui y ont cours (Abby et al. 2014; Pinto et al. 2021; Vacant et al. 2022).

Dans le troisième chapitre, j'ai émis l'hypothèse que *P. calceolata* assimile préférentiellement l'azote organique à la DCM, en se basant sur la part de gènes différentiellement exprimé en commun entre ceux surexprimé à la DCM et ceux surexprimé dans les conditions de changements de quantité et de sources d'azote. Il serait possible qu'à la DCM, *P. calceolata* utilise des sources alternatives d'azote, notamment pour faire face à la compétition accrue liée à la plus grande abondance d'organismes phytoplanctoniques par rapport à la surface. Cette hypothèse permet de remettre *P. calceolata* dans son contexte environnemental : elle vit entourée d'autres cellules, avec qui elle peut être en compétition pour les ressources. Cette considération ouvre la porte à de nombreuses questions. Par exemple, on peut se demander si *P. calceolata* forme des associations avec d'autres cellules. Sa présence en plus faible abondance dans les fractions de taille supérieures 20-180  $\mu\text{m}$ , 180-2000  $\mu\text{m}$  et 0.8-2000  $\mu\text{m}$  peut laisser penser soit qu'elles forment des mucilages ou des colonies, comme en sont capables un certain nombre de Pelagophyceae (de plus, j'ai pu observer parfois chez *P. calceolata* des formations semblables à des mucilages), soit que les *P. calceolata* peuvent être physiquement associées à des espèces plus grandes. L'expression des gènes n'est pas grandement impactée par la fraction de taille dans les échantillons des fractions 0.8-5  $\mu\text{m}$  et 0.8-2000  $\mu\text{m}$ , on peut donc supposer que la physiologie de *P. calceolata* est similaire entre ces deux fractions de tailles. En revanche, je ne me suis pas intéressée à l'expression des gènes dans les fraction 20-180  $\mu\text{m}$  et 180-2000  $\mu\text{m}$ , je ne peux donc pas dire s'il y a une différence d'expression de ces cellules potentiellement en colonie par rapport aux cellules probablement sous forme libre.

Les interactions peuvent aussi être non physiques. Par exemple, *P. calceolata* pourrait profiter des métabolites rejetés par d'autres organismes. L'hypothèse de la Reine Noire (Black Queen Hypothesis) stipule que la perte d'un certains nombres de gènes, notamment ceux impliqués dans la voie de biosynthèses de métabolites solubles telles que les acides aminés, serait possible grâce à la mutualisation des métabolites.

L'application de cette théorie évolutive aux PPE permettrait d'expliquer la réduction du génome de certaines espèces sans passage à la vie symbiotique (Morris, Lenski, and Zinser 2012; Derilus et al. 2020; 2023). Pour que cette mutualisation des métabolites secondaires existe chez *P. calceolata*, il faudrait un mécanisme de *quorum-sensing*. La présence des gènes codant pour les senseurs de nitrate/nitrite (Chapitre 1, Figure 5) peut indiquer que *P. calceolata* est capable de ce *quorum-sensing*. On pourrait supposer qu'elle pourrait se concentrer dans les zones où il y a de l'azote disponible, profitant par exemple des molécules de DON relarguées par d'autres organismes. Si ce partage des métabolites a bien lieu entre les PPE, alors il peut en partie expliquer le succès écologique de *P. calceolata* dans les zones oligotrophes, pauvres en fer ou en azote.

La question des interactions entre cellules dans l'environnement prend de plus en plus d'importance, car elles sont longtemps passées sous les radars des études de génomique environnementales (Worden et al. 2015). Les réseaux de cooccurrence permettent de déterminer quelles sont les espèces qui sont co-abondantes à partir des données de méta-omiques comme celles de *Tara Oceans*, mais ces méthodes indirectes ne renseignent pas sur la nature de l'interaction. En revanche, l'utilisation de méthodes single-cell couplées au séquençage permettent d'étudier les interactions physiques directement à partir d'échantillons environnementaux.

### **3 - *Pelagomonas calceolata* comme modèle pour l'étude des PPE**

Au-delà des résultats scientifiques apportés sur l'acclimatation de *P. calceolata* aux changements environnementaux, ce manuscrit sert également de preuve de concept. En effet, ce travail est la première thèse portant entièrement et spécifiquement sur *Pelagomonas calceolata*. On peut se demander quel est l'intérêt d'étudier cet organisme en particulier alors que d'autres microalgues photosynthétiques sont déjà étudiées en laboratoires, et que plusieurs sont mieux caractérisées au niveau génétique, physiologique, métabolique ou écologique. L'étude de *P. calceolata* permet de mieux comprendre un groupe encore mal caractérisé, les Pélagophytes. Caractériser des espèces inconnues permet d'enrichir les ressources taxonomiques et fonctionnelles, qui font souvent défaut pour les eucaryotes marins (Worden and Allen 2010; del Campo et al. 2014). Chaque génome nouvellement séquencé amène son lot de nouvelles découvertes et enrichit les bases de données de nouvelles références, particulièrement avec des données parvenant de taxons peu décrits, permettant de diminuer les erreurs d'inférences taxonomiques ou fonctionnelles. J'ai mené au travers de cette thèse la première caractérisation globale d'un Pélagophyte pélagique et cosmopolite.

Ce travail de thèse montre que *P. calceolata* peut tout à fait servir de modèle pour étudier l'acclimatation des picoeucaryotes photosynthétiques. En effet, cette espèce présente un certain nombre de caractéristiques qui en font un organisme modèle

pertinent à l'ère des multi-omiques. Du point de vue de la génomique environnementale, *P. calceolata* remplit les critères définis par Worden et al. (2015) pour assurer le succès d'une étude méta-omique. D'abord, il faut un génome de référence de bonne qualité permettant d'identifier les séquences de cette espèce dans l'environnement. Ensuite, ce génome doit être de petite taille afin de ne pas nécessiter une profondeur de séquençage trop importante. Et finalement, l'espèce doit avoir une diversité assez faible au travers de l'environnement étudié (Worden et al. 2015). Avec son génome de 31 Mb assemblé à l'échelle du chromosome avec un séquençage en courtes et longues lectures, j'ai montré tout au long de cette thèse qu'il était possible d'étudier *P. calceolata* dans l'ensemble des données *Tara Oceans*, en métagénomique et en métatranscriptomique simultanément. L'abondance et le cosmopolitisme de *P. calceolata* en font une microalgue pertinente à étudier dans les jeux de données couvrant l'océan global, car elle est facile à détecter et présente dans une grande diversité de conditions environnementales.

La question de la diversité de *P. calceolata* dans l'environnement est un axe de recherche majeur à développer. L'un des premiers travaux que j'ai fait dans le cadre de ma thèse, et qui n'a pas été présenté ici, a été de rechercher des variants de l'ADNr 16S ou 18S dans l'ensemble des données de métabarcoding des expéditions *Tara Oceans* et *Tara Pacific*, afin d'identifier s'il y avait une ou plusieurs espèces de *Pelagomonas* dans l'environnement. Mon analyse m'a amenée à considérer qu'il n'y avait qu'une seule espèce. Cette diversité réduite est un des critères qui font de *P. calceolata* un bon modèle pour l'analyse des données de méta-omiques environnementales. Cependant, la diversité des *Pelagomonas* mérite d'être examinée plus longuement au travers d'une analyse de structure de population détaillée. Également, le choix de *P. calceolata* RCC100 comme souche de référence peut poser question. Échantillonnée dans le gyre central du Pacifique Nord en 1973, cette souche est en culture depuis 50 ans. Nous l'avons choisie car il s'agit de la souche historique à partir de laquelle l'espèce a été décrite (Andersen et al. 1993). Cependant, la mise en culture prolongée peut amener à une perte de fonction ou à une augmentation des modifications génétiques délétères. Le génome en culture de RCC100 présente de larges zones dupliquées qui semblent être en simple copie dans l'environnement. Malgré cela, le recrutement des séquences environnementales reste efficace. La comparaison du génome, de la physiologie et du métabolisme de RCC100 avec certaines des 72 autres souches de *P. calceolata* maintenues à la collection d'algues de Roscoff (Vaulot et al. 2004) permettra de déterminer si ce choix de souche était le plus pertinent. Comment la mise en culture a impacté le génome de chaque souche, et quelles différences physiologiques existent entre des souches échantillonnées à des sites différents en termes de température, luminosité, concentration en nutriments, etc. sont des questions nécessitant la comparaison de plusieurs souches d'une même espèce.

Du point de vue de la biologie cellulaire et moléculaire, *P. calceolata* a beaucoup de potentiel, mais des outils restent à développer afin d'en faire un organisme modèle performant. J'ai mis en place la culture de *P. calceolata* au Genoscope à la fin de l'année

2021, et l'entretien de cette espèce est assez facile. Les conditions de luminosité et de température permettant sa croissance peuvent être implémentées simplement. Avec son rythme de division d'environ 1 par jour, un repiquage tous les 15 jours suffit à maintenir une culture saine. L'extraction et le séquençage de l'ADN et de l'ARN ne rencontrent pas de difficulté particulière, un protocole standard du laboratoire de séquençage du Genoscope permet d'avoir de bons rendements avec un volume de culture d'environ 50 ml. Nous avons réussi en seulement quelques mois à optimiser les conditions de cultures et à mettre en place l'expérience complète dans différentes conditions de source et de quantité d'azote décrite dans le chapitre 2 (Chapitre 2, Materials and Methods).

Cependant, *P. calceolata* présente un certain nombre de caractéristiques qui font de son utilisation en laboratoire un défi. Elle présente une croissance assez lente, qui nécessite une attente de 7 à 10 jours pour obtenir une biomasse suffisante. On observe souvent plusieurs jours avec une croissance très faible après la dilution dans du milieu frais. Également, sa croissance semble être irrégulière par période, et impactée par des paramètres que nous ignorons. Beaucoup de choses restent à découvrir sur les paramètres impactant sa croissance et comment l'optimiser. La température et la luminosité choisies sont convenables pour les besoins de cette thèse, mais il est sûrement possible d'optimiser les conditions de culture. Comme évoqué précédemment, la souche utilisée dans le cadre de cette thèse n'est pas axénique, et la présence de cette phycosphère joue probablement un rôle sur la croissance et le métabolisme *P. calceolata*. C'est aussi une espèce assez petite (1.5 à 3 µm), ce qui la rend difficile à observer au microscope optique. À cause de cela, il est difficile d'observer ses caractéristiques morphologiques ou sa structure intracellulaire facilement.

Nous avons aussi fait face à un certain nombre de contretemps et d'échecs liés à des problèmes techniques plutôt que biologiques. *P. calceolata* étant un organisme peu étudié, il n'existe pas de protocole pour de multiples techniques de laboratoire. Bien que non décrites dans ce manuscrit, nous avons initié la mise en place de protocoles pour cultiver *P. calceolata* en bioréacteur, extraire ces métabolites, ou encore pour transformer cette souche afin faire de l'édition ou de la surexpression génique. Les microalgues pour lesquelles nous maîtrisons ces techniques sont de plus en plus nombreuses : on peut notamment citer *Ostreococcus tauri* (chlorophyte), *Phaeodactylum tricornutum* (diatomée) ou *Nannochloropsis gaditana* (eustigmatophyte) (Sproles et al. 2021). Ce travail a révélé un certain nombre de gènes de fonction inconnue qui sont de bon candidats pour des mécanismes importants dans l'acclimatation aux limitations en azote, en fer ou à la profondeur, mais il est impossible à ce jour de tester leurs fonctions dans *P. calceolata*. Le développement de protocoles pour permettre d'étudier *P. calceolata* au travers d'autres techniques va prendre un peu de temps, mais cela permettra d'en faire un organisme modèle pour des études de biologie moléculaire et cellulaire, à coupler avec des analyses de génomiques environnementales.

*P. calceolata* pourrait servir d'organisme modèle pour découvrir de nouvelles fonctions, mais aussi pour mieux comprendre le métabolisme central. Mon 3<sup>e</sup> chapitre ouvre la porte à des questions sur l'impact de l'environnement sur le métabolisme du carbone, en particulier le cycle de Calvin et la photorespiration. Bien que l'étude de ces mécanismes chez les plantes est ancienne, leur compréhension chez les microalgues est encore en construction. Comme évoqué dans le chapitre précédent, le cycle de Calvin est bien caractérisé chez les microalgues, mais les conditions environnementales régulant son expression sont encore peu connues. Pour la photosynthèse, il reste beaucoup de nouvelles découvertes à faire. On estime qu'une grande part de gènes impliqués dans ce métabolisme pourtant basique et très étudié sont encore non caractérisés (Li et al. 2019). La généralisation de nos connaissances des plantes terrestres et algues vertes à des groupes phylogénétiquement très distants comme les Straménopiles pourraient nous induire en erreur (Bailleul et al. 2015). *P. calceolata* pourrait donc servir de modèle pour étudier des métabolismes centraux et comprendre comment ils peuvent être régulés en réponse aux changements de conditions, en laboratoire et en multi-omiques environnementales.

## 4 - Conclusion

Lors de cette thèse, j'ai étudié les capacités d'acclimatation de *Pelagomonas calceolata* aux changements environnementaux, plus particulièrement sa réponse transcriptomique aux variations de nutriments. J'ai pu, au cours de ces trois chapitres, décrire les capacités d'acclimatation de cette espèce encore peu connue grâce à la métagénomique, la génomique comparative, la métatranscriptomique, la biologie cellulaire et la transcriptomique.

J'ai étudié les niveaux d'expression des gènes de *P. calceolata* *in situ* et en culture, ce qui m'a permis de découvrir comment cette espèce s'acclimate aux changements environnementaux. Particulièrement abondante dans les eaux tempérées et pauvres en fer, *P. calceolata* présente un ensemble de gènes permettant son acclimatation à des variations de disponibilités de fer et d'azote. L'étude de sa réponse transcriptomique aux manques d'azote *in situ* et *in vitro* ont mis en lumière une série de mécanismes d'acclimatation, notamment une certaine efficacité pour économiser et recycler son azote intracellulaire, comme les diatomées. *P. calceolata* présente aussi une forte capacité à utiliser des sources d'azote organiques, comme le pélagophyte *Aureococcus anophagefferens*. Cette étude, croisant données environnementales et expériences en laboratoire, a permis de déchiffrer plus précisément des mécanismes précis, notamment le fonctionnement de la cyanate lyase, remettant en question le rôle de cette enzyme dans l'assimilation de l'azote organique chez les eucaryotes. Finalement, j'ai exploré comment *P. calceolata* régule son métabolisme central en fonction de la profondeur, ouvrant la porte à l'étude de l'impact de l'environnement sur le cycle du carbone. Ainsi, ce travail donne la première étude complète sur l'acclimatation de ce PPE aux changements environnementaux.

Du fait de ses caractéristiques génomiques et physiologiques, *Pelagomonas calceolata* a le potentiel de devenir un nouvel organisme modèle, pour la compréhension des mécanismes moléculaires déterminant l'abondance et la répartition des PPE face aux changements environnementaux. Son petit génome et son abondance en font un modèle avantageux pour mener des études en génomique environnementale. Facile à cultiver dans des conditions variables, il est aisé de mener des études fonctionnelles nombreuses, afin de comparer l'expression des gènes *in situ* et *in vitro*.

Ainsi, on peut explorer la réponse de *P. calceolata* aux changements environnementaux, de l'écosystème jusqu'aux gènes. Bien que beaucoup de protocoles expérimentaux restent à mettre en place, cette thèse constitue une preuve que *Pelagomonas calceolata* peut permettre de faire de nouvelles découvertes sur le phytoplancton, afin de mieux comprendre sa réponse aux changements climatiques et sa place dans l'écosystème marin de demain.

# Bibliographie

---

- Abby, Sophie S., Marie Touchon, Aurelien De Jode, Nigel Grimsley, and Gwenael Piganeau. 2014. 'Bacteria in *Ostreococcus Tauri* Cultures – Friends, Foes or Hitchhikers?' *Frontiers in Microbiology* 5. <https://www.frontiersin.org/articles/10.3389/fmicb.2014.00505>.
- Agostoni, Marco, and Deana L. Erdner. 2011. 'Analysis of Ammonium Transporter and Urease Gene Expression in *Aureococcus Lagunensis*'. *Harmful Algae* 10 (5): 549–56. <https://doi.org/10.1016/j.hal.2011.04.007>.
- Aktan, Yelda, and B. Topaloğlu. 2011. 'First Record of *Chrysophaeum Taylorii* Lewis & Bryan and Their Benthic Mucilaginous Aggregates in the Aegean Sea (Eastern Mediterranean)'. *J Black Sea/Mediterr Environ* 17 (January): 159–70.
- Alberti, Adriana, Julie Poulain, Stefan Engelen, Karine Labadie, Sarah Romac, Isabel Ferrera, Guillaume Albin, et al. 2017. 'Viral to Metazoan Marine Plankton Nucleotide Sequences from the Tara Oceans Expedition'. *Scientific Data* 4 (1): 1–20. <https://doi.org/10.1038/sdata.2017.93>.
- Alipanah, Leila, Jens Rohloff, Per Winge, Atle M. Bones, and Tore Brembu. 2015. 'Whole-Cell Response to Nitrogen Deprivation in the Diatom *Phaeodactylum Tricornutum*'. *Journal of Experimental Botany* 66 (20): 6281–96. <https://doi.org/10.1093/jxb/erv340>.
- Allen, Andrew E., Christopher L. Dupont, Miroslav Oborník, Aleš Horák, Adriano Nunes-Nesi, John P. McCrow, Hong Zheng, et al. 2011. 'Evolution and Metabolic Significance of the Urea Cycle in Photosynthetic Diatoms'. *Nature* 473 (7346): 203–7. <https://doi.org/10.1038/nature10074>.
- Andersen, Robert A., Gary W. Saunders, Michael P. Paskind, and Julianne P. Sexton. 1993. 'Ultrastructure and 18S Rrna Gene Sequence for *Pelagomonas Calceolata* Gen. Et Sp. Nov. and the Description of a New Algal Class, the Pelagophyceae Classis Nov.1'. *Journal of Phycology* 29 (5): 701–15. <https://doi.org/10.1111/j.0022-3646.1993.00701.x>.
- Anderson, Paul M., Young-chul Sung, and James A. Fuchs. 1990. 'The Cyanase Operon and Cyanate Metabolism'. *FEMS Microbiology Reviews* 7 (3–4): 247–52. <https://doi.org/10.1111/j.1574-6968.1990.tb04920.x>.
- Ansorge, Wilhelm J. 2009. 'Next-Generation DNA Sequencing Techniques'. *New Biotechnology* 25 (4): 195–203. <https://doi.org/10.1016/j.nbt.2008.12.009>.
- Ardyna, Mathieu, and Kevin Robert Arrigo. 2020. 'Phytoplankton Dynamics in a Changing Arctic Ocean'. *Nature Climate Change* 10 (10): 892–903. <https://doi.org/10.1038/s41558-020-0905-y>.
- Ashworth, Justin, Sacha Coesel, Allison Lee, E. Virginia Armbrust, Mónica V. Orellana, and Nitin S. Baliga. 2013. 'Genome-Wide Diel Growth State Transitions in the Diatom *Thalassiosira Pseudonana*'. *Proceedings of the National Academy of Sciences* 110 (18): 7518–23. <https://doi.org/10.1073/pnas.1300962110>.
- Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. 2015. 'PISCES-v2: An Ocean Biogeochemical Model for Carbon and Ecosystem Studies'. *Geoscientific Model Development* 8 (8): 2465–2513. <https://doi.org/10.5194/gmd-8-2465-2015>.
- Bailleul, Benjamin, Nicolas Berne, Omer Murik, Dimitris Petroustos, Judit Prihoda, Atsuko Tanaka, Valeria Villanova, et al. 2015. 'Energetic Coupling between Plastids and Mitochondria Drives CO<sub>2</sub> Assimilation in Diatoms'. *Nature* 524 (7565): 366–69. <https://doi.org/10.1038/nature14599>.
- Behnke, Joerg, and Julie LaRoche. 2020. 'Iron Uptake Proteins in Algae and the Role of Iron Starvation-Induced Proteins (ISIPs)'. *European Journal of Phycology* 55 (3): 339–60. <https://doi.org/10.1080/09670262.2020.1744039>.
- Beman, J. Michael, Cheryl-Emiliane Chow, Andrew L. King, Yuanyuan Feng, Jed A. Fuhrman, Andreas Andersson, Nicholas R. Bates, Brian N. Popp, and David A. Hutchins. 2011. 'Global Declines in Oceanic Nitrification Rates as a Consequence of Ocean Acidification'. *Proceedings of the National Academy of Sciences* 108 (1): 208–13. <https://doi.org/10.1073/pnas.1011053108>.
- Bender, Sara, Colleen Durkin, Chris Berthiaume, Rhonda Morales, and E. Virginia Armbrust. 2014. 'Transcriptional Responses of Three Model Diatoms to Nitrate Limitation of Growth'. *Frontiers in Marine Science* 1. <https://www.frontiersin.org/articles/10.3389/fmars.2014.00003>.
- Bender, Sara J., Micaela S. Parker, and E. Virginia Armbrust. 2012. 'Coupled Effects of Light and Nitrogen Source on the Urea Cycle and Nitrogen Metabolism over a Diel Cycle in the Marine Diatom *Thalassiosira Pseudonana*'. *Protist* 163 (2): 232–51. <https://doi.org/10.1016/j.protis.2011.07.008>.
- Berg, G. M., P. M. Glibert, M. W. Lomas, and M. A. Burford. 1997. 'Organic Nitrogen Uptake and Growth by the Chrysophyte *Aureococcus Anophagefferens* during a Brown Tide Event'. *Marine Biology* 129 (2): 377–87. <https://doi.org/10.1007/s002270050178>.
- Berg, Gry Mine, Daniel J. Repeta, and Julie Laroche. 2002. 'Dissolved Organic Nitrogen Hydrolysis Rates in Axenic Cultures of *Aureococcus Anophagefferens* (Pelagophyceae): Comparison with Heterotrophic Bacteria'. *Applied and Environmental Microbiology* 68 (1): 401–4. <https://doi.org/10.1128/AEM.68.1.401-404.2002>.

- Berg, Gry Mine, Jeff Shrager, Gernot Glöckner, Kevin R. Arrigo, and Arthur R. Grossman. 2008. 'UNDERSTANDING NITROGEN LIMITATION IN AUREOCOCCUS ANOPHAGEFFERENS (PELAGOPHYCEAE) THROUGH cDNA AND qRT-PCR ANALYSIS'. *Journal of Phycology* 44 (5): 1235–49. <https://doi.org/10.1111/j.1529-8817.2008.00571.x>.
- Berges, J. A., D. O. Charlebois, D. C. Mauzerall, and P. G. Falkowski. 1996. 'Differential Effects of Nitrogen Limitation on Photosynthetic Efficiency of Photosystems I and II in Microalgae'. *Plant Physiology* 110 (2): 689–96. <https://doi.org/10.1104/pp.110.2.689>.
- Berthelot, Hugo, Solange Duhamel, Stéphane L'Helguen, Jean-Francois Maguer, Seaver Wang, Ivona Cetinić, and Nicolas Cassar. 2019. 'NanoSIMS Single Cell Analyses Reveal the Contrasting Nitrogen Sources for Small Phytoplankton'. *The ISME Journal* 13 (3): 651–62. <https://doi.org/10.1038/s41396-018-0285-8>.
- Botebol, Hugo, Gaele Lelandais, Christophe Six, Emmanuel Lesuisse, Arnaud Meng, Lucie Bittner, Stéphane Lecrom, et al. 2017. 'Acclimation of a Low Iron Adapted *Ostreococcus* Strain to Iron Limitation through Cell Biomass Lowering'. *Scientific Reports* 7 (1): 327. <https://doi.org/10.1038/s41598-017-00216-6>.
- Boyce, Daniel G., Marlon R. Lewis, and Boris Worm. 2010. 'Global Phytoplankton Decline over the Past Century'. *Nature* 466 (7306): 591–96. <https://doi.org/10.1038/nature09268>.
- Boyd, P. W., and M. J. Ellwood. 2010. 'The Biogeochemical Cycle of Iron in the Ocean'. *Nature Geoscience* 3 (10): 675–82. <https://doi.org/10.1038/ngeo964>.
- Boyd, P. W., T. Jickells, C. S. Law, S. Blain, E. A. Boyle, K. O. Buesseler, K. H. Coale, et al. 2007. 'Mesoscale Iron Enrichment Experiments 1993–2005: Synthesis and Future Directions'. *Science* 315 (5812): 612–17. <https://doi.org/10.1126/science.1131669>.
- Bricelj, V. Monica, and Scott P. MacQuarrie. 2007. 'Effects of Brown Tide (*Aureococcus Anophagefferens*) on Hard Clam *Mercenaria Mercenaria* Larvae and Implications for Benthic Recruitment'. *Marine Ecology Progress Series* 331 (February): 147–59. <https://doi.org/10.3354/meps331147>.
- Bristow, Laura A., Wiebke Mohr, Soeren Ahmerkamp, and Marcel M. M. Kuypers. 2017. 'Nutrients That Limit Growth in the Ocean'. *Current Biology* 27 (11): R474–78. <https://doi.org/10.1016/j.cub.2017.03.030>.
- Bronk, D. A., J. H. See, P. Bradley, and L. Killberg. 2007. 'DON as a Source of Bioavailable Nitrogen for Phytoplankton'. *Biogeosciences* 4 (3): 283–96. <https://doi.org/10.5194/bg-4-283-2007>.
- Browning, Thomas J., and C. Mark Moore. 2023. 'Global Analysis of Ocean Phytoplankton Nutrient Limitation Reveals High Prevalence of Co-Limitation'. *Nature Communications* 14 (1): 5014. <https://doi.org/10.1038/s41467-023-40774-0>.
- Brum, Jennifer R., J. Cesar Ignacio-Espinoza, Simon Roux, Guilhem Doucier, Silvia G. Acinas, Adriana Alberti, Samuel Chaffron, et al. 2015. 'Patterns and Ecological Drivers of Ocean Viral Communities'. *Science* 348 (6237): 1261498. <https://doi.org/10.1126/science.1261498>.
- Burlacot, Adrien, and Gilles Peltier. 2023. 'Energy Crosstalk between Photosynthesis and the Algal CO<sub>2</sub>-Concentrating Mechanisms'. *Trends in Plant Science* 28 (7): 795–807. <https://doi.org/10.1016/j.tplants.2023.03.018>.
- Busch, Florian A. 2020. 'Photorespiration in the Context of Rubisco Biochemistry, CO<sub>2</sub> Diffusion and Metabolism'. *The Plant Journal* 101 (4): 919–39. <https://doi.org/10.1111/tpj.14674>.
- Buskey, Edward J., Brian Wysor, and Cammie Hyatt. 1998. 'The Role of Hypersalinity in the Persistence of the Texas "Brown Tide" in the Laguna Madre'. *Journal of Plankton Research* 20 (8): 1553–65. <https://doi.org/10.1093/plankt/20.8.1553>.
- Cabello, Ana M., Mikel Latasa, Irene Forn, Xosé Anxelu G. Morán, and Ramon Massana. 2016. 'Vertical Distribution of Major Photosynthetic Picoeukaryotic Groups in Stratified Marine Waters'. *Environmental Microbiology* 18 (5): 1578–90. <https://doi.org/10.1111/1462-2920.13285>.
- Cabré, Anna, David Shields, Irina Marinov, and Tihomir S. Kostadinov. 2016. 'Phenology of Size-Partitioned Phytoplankton Carbon-Biomass from Ocean Color Remote Sensing and CMIP5 Models'. *Frontiers in Marine Science* 3. <https://www.frontiersin.org/articles/10.3389/fmars.2016.00039>.
- Campo, Javier del, Michael E. Sieracki, Robert Molestina, Patrick Keeling, Ramon Massana, and Iñaki Ruiz-Trillo. 2014. 'The Others: Our Biased Perspective of Eukaryotic Genomes'. *Trends in Ecology & Evolution* 29 (5): 252–59. <https://doi.org/10.1016/j.tree.2014.03.006>.
- Caputi, Luigi, Quentin Carradec, Damien Eveillard, Amos Kirilovsky, Eric Pelletier, Juan J. Pierella Karlusich, Fabio Rocha Jimenez Vieira, et al. 2019. 'Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems'. *Global Biogeochemical Cycles* 33 (3): 391–419. <https://doi.org/10.1029/2018GB006022>.
- Caronni, Sarah, Alessia Bresciani, Maria Anna Delaria, Federico Meloni, Augusto Navone, Pieraugusto Panzalis, Kirsten Heimann, and Giulia Ceccherelli. 2015. 'Ecology of the Benthic Mucilage-Forming Microalga *Chrysophaeum Taylorii* in the W Mediterranean Sea: Substratum and Depth Preferences'. *Estuarine, Coastal and Shelf Science* 161 (August): 38–45. <https://doi.org/10.1016/j.ecss.2015.04.012>.
- Caronni, Sarah, Chiara Calabretti, Gianluca Cavagna, Giulia Ceccherelli, Maria Anna Delaria, Giovanni Macri, Augusto Navone, and Pieraugusto Panzalis. 2017. 'The Invasive Microalga *Chrysophaeum Taylorii*: Interactive Stressors

- Regulate Cell Density and Mucilage Production'. *Marine Environmental Research* 129 (August): 156–65. <https://doi.org/10.1016/j.marenvres.2017.05.005>.
- Caronni, Sarah, Maria Anna Delaria, Kirsten Heimann, Giovanni Macri, Augusto Navone, Pieraugusto Panzalis, and Giulia Ceccherelli. 2016. 'The Role of Floating Mucilage in the Invasive Spread of the Benthic Microalga *Chrysophaeum Taylorii*'. *Marine Ecology* 37 (4): 867–76. <https://doi.org/10.1111/maec.12365>.
- Carradec, Quentin, Eric Pelletier, Corinne Da Silva, Adriana Alberti, Yoann Seeleuthner, Romain Blanc-Mathieu, Gipsi Lima-Mendez, et al. 2018. 'A Global Ocean Atlas of Eukaryotic Genes'. *Nature Communications* 9 (1): 373. <https://doi.org/10.1038/s41467-017-02342-1>.
- Cerón García, M. C., A. Sánchez Mirón, J. M. Fernández Sevilla, E. Molina Grima, and F. García Camacho. 2005. 'Mixotrophic Growth of the Microalga *Phaeodactylum Tricornutum*: Influence of Different Nitrogen and Organic Carbon Sources on Productivity and Biomass Composition'. *Process Biochemistry* 40 (1): 297–305. <https://doi.org/10.1016/j.procbio.2004.01.016>.
- Choi, Chang Jae, Valeria Jimenez, David M. Needham, Camille Poirier, Charles Bachy, Harriet Alexander, Susanne Wilken, et al. 2020. 'Seasonal and Geographical Transitions in Eukaryotic Phytoplankton Community Structure in the Atlantic and Pacific Oceans'. *Frontiers in Microbiology* 11. <https://doi.org/10.3389/fmicb.2020.542372>.
- Choi, Dong H., Sung M. An, Sungjun Chun, Eun C. Yang, Karen E. Selph, Charity M. Lee, and Jae H. Noh. 2016. 'Dynamic Changes in the Composition of Photosynthetic Picoeukaryotes in the Northwestern Pacific Ocean Revealed by High-Throughput Tag Sequencing of Plastid 16S rRNA Genes'. *FEMS Microbiology Ecology* 92 (2): fiv170. <https://doi.org/10.1093/femsec/fiv170>.
- Clode, Peta L., Martin Saunders, Garth Maker, Martha Ludwig, and Craig A. Atkins. 2009. 'Uric Acid Deposits in Symbiotic Marine Algae'. *Plant, Cell & Environment* 32 (2): 170–77. <https://doi.org/10.1111/j.1365-3040.2008.01909.x>.
- Coale, Kenneth H., Kenneth S. Johnson, Steve E. Fitzwater, R. Michael Gordon, Sara Tanner, Francisco P. Chavez, Laurie Ferioli, et al. 1996. 'A Massive Phytoplankton Bloom Induced by an Ecosystem-Scale Iron Fertilization Experiment in the Equatorial Pacific Ocean'. *Nature* 383 (6600): 495–501. <https://doi.org/10.1038/383495a0>.
- Collos, Yves, and John Berges. 2002. 'Nitrogen Metabolism in Phytoplankton'. In .
- Cullen, John J. 2015. 'Subsurface Chlorophyll Maximum Layers: Enduring Enigma or Mystery Solved?' *Annual Review of Marine Science* 7 (1): 207–39. <https://doi.org/10.1146/annurev-marine-010213-135111>.
- Davison, Jack R., and Carole A. Bewley. 2019. 'Antimicrobial Chrysophaentin Analogs Identified from Laboratory Cultures of the Marine Microalga *Chrysophaeum Taylorii*'. *Journal of Natural Products* 82 (1): 148–53. <https://doi.org/10.1021/acs.jnatprod.8b00858>.
- Davison, Jack R., Rahim Rajwani, Gengxiang Zhao, and Carole A. Bewley. 2023. 'The Genome of Antibiotic-Producing Colonies of the Pelagophyte Alga *Chrysophaeum Taylorii* Reveals a Diverse and Non-Canonical Capacity for Secondary Metabolism'. *Scientific Reports* 13 (July): 11944. <https://doi.org/10.1038/s41598-023-38042-8>.
- Delmont, Tom O., Morgan Gaia, Damien D. Hingsinger, Paul Frémont, Chiara Vanni, Antonio Fernandez-Guerra, A. Murat Eren, et al. 2022. 'Functional Repertoire Convergence of Distantly Related Eukaryotic Plankton Lineages Abundant in the Sunlit Ocean'. *Cell Genomics* 2 (5): 100123. <https://doi.org/10.1016/j.xgen.2022.100123>.
- Delmont, Tom O., Christopher Quince, Alon Shaiber, Özcan C. Esen, Sonny TM Lee, Michael S. Rappé, Sandra L. McLellan, Sebastian Lucker, and A. Murat Eren. 2018. 'Nitrogen-Fixing Populations of Planctomycetes and Proteobacteria Are Abundant in Surface Ocean Metagenomes'. *Nature Microbiology* 3 (7): 804. <https://doi.org/10.1038/s41564-018-0176-9>.
- Derbès, Auguste-Alphonse, and Antoine-Joseph-Jean Solier. 1856. 'Mémoire Sur Quelques Points de La Physiologie Des Algues'. *Comptes Rendus Hebdomadaires Des Séances de l'Académie Des Sciences*.
- Derilus, D., C. E. Burdyslaw, F. Pinero, and S. E. Massey. 2023. 'Photosynthetic Picoeukaryotes Significantly Affect Global Oceanic Meta-Metabolism: Cellular and Genome Streamlining Confer Ecological Success'. *Aquatic Microbial Ecology* 89 (March): 23–41. <https://doi.org/10.3354/ame01995>.
- Derilus, D., M. Z. Rahman, F. Pinero, and S. E. Massey. 2020. 'Synergism between the Black Queen Effect and the Proteomic Constraint on Genome Size Reduction in the Photosynthetic Picoeukaryotes'. *Scientific Reports* 10 (1): 8918. <https://doi.org/10.1038/s41598-020-65476-1>.
- DeYoe, Hudson R., Dean A. Stockwell, Robert R. Biologare, Mikel Latasa, Paul W. Johnson, Paul E. Hargraves, and Curtis A. Suttle. 1997. 'Description and Characterization of the Algal Species *Aureoombra Lagunensis* Gen. Et Sp. Nov. and Referral of *Aureoombra* and *Aureococcus* to the Pelagophyceae<sup>1</sup>'. *Journal of Phycology* 33 (6): 1042–48. <https://doi.org/10.1111/j.0022-3646.1997.01042.x>.
- DeYoe, Hudson R., and Curtis A. Suttle. 1994. 'The Inability of the Texas "Brown Tide" Alga to Use Nitrate and the Role of Nitrogen in the Initiation of a Persistent Bloom of This Organism<sup>1</sup>'. *Journal of Phycology* 30 (5): 800–806. <https://doi.org/10.1111/j.0022-3646.1994.00800.x>.
- Dimier, Céline, Christophe Brunet, Richard Geider, and John Raven. 2009. 'Growth and Photoregulation Dynamics of the Picoeukaryote *Pelagomonas Calceolata* in Fluctuating Light'. *Limnology and Oceanography* 54 (3): 823–36.

- <https://doi.org/10.4319/lo.2009.54.3.0823>.
- Dimier, Céline, Saviello Giovanni, Tramontano Ferdinando, and Christophe Brunet. 2009. 'Comparative Ecophysiology of the Xanthophyll Cycle in Six Marine Phytoplanktonic Species'. *Protist* 160 (3): 397–411. <https://doi.org/10.1016/j.protis.2009.03.001>.
- Dirnhuber, P., and F. Schütz. 1948. 'The Isomeric Transformation of Urea into Ammonium Cyanate in Aqueous Solutions'. *Biochemical Journal* 42 (4): 628–32.
- Dong, Hong-Po, Kai-Xuan Huang, Hua-Long Wang, Song-Hui Lu, Jing-Yi Cen, and Yue-Lei Dong. 2014. 'Understanding Strategy of Nitrate and Urea Assimilation in a Chinese Strain of *Aureococcus Anophagefferens* through RNA-Seq Analysis'. *PLOS ONE* 9 (10): e111069. <https://doi.org/10.1371/journal.pone.0111069>.
- Dunne, John P. 2022. 'Fall and Rise of the Phytoplankton'. *Nature Climate Change* 12 (8): 708–9. <https://doi.org/10.1038/s41558-022-01439-w>.
- Dupont, Chris L., John P. McCrow, Ruben Valas, Ahmed Moustafa, Nathan Walworth, Ursula Goodenough, Robyn Roth, et al. 2015. 'Genomes and Gene Expression across Light and Productivity Gradients in Eastern Subtropical Pacific Microbial Communities'. *The ISME Journal* 9 (5): 1076–92. <https://doi.org/10.1038/ismej.2014.198>.
- Engel, Anja, Nicole Händel, Julia Wohlers, Mirko Lunau, Hans-Peter Grossart, Ulrich Sommer, and Ulf Riebesell. 2011. 'Effects of Sea Surface Warming on the Production and Composition of Dissolved Organic Matter during Phytoplankton Blooms: Results from a Mesocosm Study'. *Journal of Plankton Research* 33 (3): 357–72. <https://doi.org/10.1093/plankt/fbq122>.
- Estrada, Marta, Maximino Delgado, Dolores Blasco, Mikel Latasa, Ana María Cabello, Verónica Benítez-Barrios, Eugenio Fraile-Nuez, Patricija Mozetič, and Montserrat Vidal. 2016. 'Phytoplankton across Tropical and Subtropical Regions of the Atlantic, Indian and Pacific Oceans'. *PLOS ONE* 11 (3): e0151699. <https://doi.org/10.1371/journal.pone.0151699>.
- Faure, Emile, Fabrice Not, Anne-Sophie Benoiston, Karine Labadie, Lucie Bittner, and Sakina-Dorothee Ayata. 2019. 'Mixotrophic Protists Display Contrasted Biogeographies in the Global Ocean'. *The ISME Journal* 13 (4): 1072–83. <https://doi.org/10.1038/s41396-018-0340-5>.
- Field, Christopher B., Michael J. Behrenfeld, James T. Randerson, and Paul Falkowski. 1998. 'Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components'. *Science* 281 (5374): 237–40. <https://doi.org/10.1126/science.281.5374.237>.
- Fischer, Woodward W., James Hemp, and Jena E. Johnson. 2016. 'Evolution of Oxygenic Photosynthesis'. *Annual Review of Earth and Planetary Sciences* 44 (1): 647–83. <https://doi.org/10.1146/annurev-earth-060313-054810>.
- Frémont, Paul, Marion Gehlen, Mathieu Vrac, Jade Leconte, Tom O. Delmont, Patrick Wincker, Daniele Iudicone, and Olivier Jaillon. 2022. 'Restructuring of Plankton Genomic Biogeography in the Surface Ocean under Climate Change'. *Nature Climate Change* 12 (4): 393–401. <https://doi.org/10.1038/s41558-022-01314-8>.
- Freyria, Nastasia J., Alan Kuo, Mansi Chovatia, Jenifer Johnson, Anna Lipzen, Kerrie W. Barry, Igor V. Grigoriev, and Connie Lovejoy. 2022. 'Salinity Tolerance Mechanisms of an Arctic Pelagophyte Using Comparative Transcriptomic and Gene Expression Analysis'. *Communications Biology* 5 (1): 1–13. <https://doi.org/10.1038/s42003-022-03461-2>.
- Fu, Weiwei, James T. Randerson, and J. Keith Moore. 2016. 'Climate Change Impacts on Net Primary Production (NPP) and Export Production (EP) Regulated by Increasing Stratification and Phytoplankton Community Structure in the CMIP5 Models'. *Biogeosciences* 13 (18): 5151–70. <https://doi.org/10.5194/bg-13-5151-2016>.
- Fuller, Nicholas J., Colin Campbell, David J. Allen, Frances D. Pitt, Katrin Zwirgmaier, Florence Le Gall, Daniel Vaultot, and David J. Scanlan. 2006. 'Analysis of Photosynthetic Picoeukaryote Diversity at Open Ocean Sites in the Arabian Sea Using a PCR Biased towards Marine Algal Plastids'. *Aquatic Microbial Ecology* 43 (1): 79–93. <https://doi.org/10.3354/ame043079>.
- Gann, Eric R., Alexander R. Truchon, Spiridon E. Papoulis, Sonya T. Dyhrman, Christopher J. Gobler, and Steven W. Wilhelm. 2022. '*Aureococcus Anophagefferens* (Pelagophyceae) Genomes Improve Evaluation of Nutrient Acquisition Strategies Involved in Brown Tide Dynamics'. *Journal of Phycology* 58 (1): 146–60. <https://doi.org/10.1111/jpy.13221>.
- García-Portela, María, Beatriz Reguera, Jesús Gago, Mickael Le Gac, and Francisco Rodríguez. 2020. 'Uptake of Inorganic and Organic Nitrogen Sources by *Dinophysis Acuminata* and *D. Acuta*'. *Microorganisms* 8 (2): 187. <https://doi.org/10.3390/microorganisms8020187>.
- Geider, Richard J., and Julie La Roche. 1994. 'The Role of Iron in Phytoplankton Photosynthesis, and the Potential for Iron-Limitation of Primary Productivity in the Sea'. *Photosynthesis Research* 39 (3): 275–301. <https://doi.org/10.1007/BF00014588>.
- Geitler, L. 1930. 'Ein Grunes Filarplasmodium Und Andere Neue Protisten'. *Arch. Protistenk* 69: 615–37.
- Gerwick, W. H. 1989. '6-Desmethoxyhormothamnione, a New Cytotoxic Styrylchromone from the Marine Cryptophyte *Chrysothamnium Taylori*'. *Journal of Natural Products* 52 (2): 252–56. <https://doi.org/10.1021/np50062a005>.

- Gerwick, William H, Albert Lopez, Gregory D Van Duyne, Jon Clardy, Wanda Ortiz, and Adriana Baez. 1986. 'Hormothamnion, a Novel Cytotoxic Styrylchromone from the Marine Cyanophyte Hormothamnion Enteromorphaoides Grunow'. *Tetrahedron Letters* 27 (18): 1979–82. [https://doi.org/10.1016/S0040-4039\(00\)84426-3](https://doi.org/10.1016/S0040-4039(00)84426-3).
- Giner, Caterina R., Massimo C. Pernice, Vanessa Balagué, Carlos M. Duarte, Josep M. Gasol, Ramiro Logares, and Ramon Massana. 2020. 'Marked Changes in Diversity and Relative Activity of Picoeukaryotes with Depth in the World Ocean'. *The ISME Journal* 14 (2): 437–49. <https://doi.org/10.1038/s41396-019-0506-9>.
- Gobler, C. J., D. J. Lonsdale, and G. L. Boyer. 2005. 'A Review of the Causes, Effects, and Potential Management of Harmful Brown Tide Blooms Caused by *Aureococcus Anophagefferens* (Hargraves et Sieburth)'. *Estuaries* 28 (5): 726–49. <https://doi.org/10.1007/BF02732911>.
- Gobler, Christopher J., Dianna L. Berry, Sonya T. Dyhrman, Steven W. Wilhelm, Asaf Salamov, Alexei V. Lobanov, Yan Zhang, et al. 2011. 'Niche of Harmful Alga *Aureococcus Anophagefferens* Revealed through Ecogenomics'. *Proceedings of the National Academy of Sciences* 108 (11): 4352–57. <https://doi.org/10.1073/pnas.1016106108>.
- Gobler, Christopher J., George E. Boneillo, Casey J. Debenham, and David A. Caron. 2004. 'Nutrient Limitation, Organic Matter Cycling, and Plankton Dynamics during an *Aureococcus Anophagefferens* Bloom'. *Aquatic Microbial Ecology* 35 (1): 31–43. <https://doi.org/10.3354/ame035031>.
- Gobler, Christopher J., Alexei V. Lobanov, Ying-Zhong Tang, Anton A. Turanov, Yan Zhang, Martina Doblin, Gordon T. Taylor, Sergio A. Sañudo-Wilhelmy, Igor V. Grigoriev, and Vadim N. Gladyshev. 2013. 'The Central Role of Selenium in the Biochemistry and Ecology of the Harmful Pelagophyte, *Aureococcus Anophagefferens*'. *The ISME Journal* 7 (7): 1333–43. <https://doi.org/10.1038/ismej.2013.25>.
- Gobler, Christopher J., Marc J. Renaghan, and Nathaniel J. Buck. 2002. 'Impacts of Nutrients and Grazing Mortality on the Abundance of *Aureococcus Anophagefferens* during a New York Brown Tide Bloom'. *Limnology and Oceanography* 47 (1): 129–41. <https://doi.org/10.4319/lo.2002.47.1.0129>.
- Gobler, Christopher J., and William G. Sunda. 2012. 'Ecosystem Disruptive Algal Blooms of the Brown Tide Species, *Aureococcus Anophagefferens* and *Aureombra Lagunensis*'. *Harmful Algae*, Harmful Algae--The requirement for species-specific information, 14 (February): 36–45. <https://doi.org/10.1016/j.hal.2011.10.013>.
- Grossman, Ryan D. 2022. 'Resolving the Molecular Ecology of Marine Microbial Eukaryotes with Metatranscriptomes'. University of Washington.
- Guérin, Nina, Marta Ciccarella, Elisa Flamant, Paul Frémont, Sophie Mangenot, Benjamin Istace, Benjamin Noel, et al. 2022. 'Genomic Adaptation of the Picoeukaryote *Pelagomonas Calceolata* to Iron-Poor Oceans Revealed by a Chromosome-Scale Genome Sequence'. *Communications Biology* 5 (1): 1–14. <https://doi.org/10.1038/s42003-022-03939-z>.
- Guidi, Lionel, Samuel Chaffron, Lucie Bittner, Damien Eveillard, Abdelhalim Larhlimi, Simon Roux, Youssef Darzi, et al. 2016. 'Plankton Networks Driving Carbon Export in the Oligotrophic Ocean'. *Nature* 532 (7600): 465–70. <https://doi.org/10.1038/nature16942>.
- Guillard, R. R. L., and P. E. Hargraves. 1993. '*Stichochrysis Immobilis* Is a Diatom, Not a Chrysophyte'. *Phycologia* 32 (3): 234–36. <https://doi.org/10.2216/i0031-8884-32-3-234.1>.
- Guilloton, Michel, and Francis Karst. 1987. 'Isolation and Characterization of *Escherichia Coli* Mutants Lacking Inducible Cyanase'. *Microbiology* 133 (3): 645–53. <https://doi.org/10.1099/00221287-133-3-645>.
- Han, Kwi Young, Louis Graf, Carolina P. Reyes, Barbara Melkonian, Robert A. Andersen, Hwan Su Yoon, and Michael Melkonian. 2018. 'A Re-Investigation of *Sarcinochrysis Marina* (*Sarcinochrysidales*, *Pelagophyceae*) from Its Type Locality and the Descriptions of *Arachnochrysis*, *Pelagospilus*, *Sargassococcus* and *Sungminbooa* Genera Nov.'. *Protist* 169 (1): 79–106. <https://doi.org/10.1016/j.protis.2017.12.004>.
- Handelsman, Jo, Michelle R. Rondon, Sean F. Brady, Jon Clardy, and Robert M. Goodman. 1998. 'Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products'. *Chemistry & Biology* 5 (10): R245–49. [https://doi.org/10.1016/S1074-5521\(98\)90108-9](https://doi.org/10.1016/S1074-5521(98)90108-9).
- Henson, Stephanie A., B. B. Cael, Stephanie R. Allen, and Stephanie Dutkiewicz. 2021. 'Future Phytoplankton Diversity in a Changing Climate'. *Nature Communications* 12 (1): 1–8. <https://doi.org/10.1038/s41467-021-25699-w>.
- Hildebrand, Mark. 2005. 'Cloning and Functional Characterization of Ammonium Transporters from the Marine Diatom *Cylindrotheca Fusiformis* (*Bacillariophyceae*)<sup>1</sup>'. *Journal of Phycology* 41 (1): 105–13. <https://doi.org/10.1111/j.1529-8817.2005.04108.x>.
- Hockin, Nicola Louise, Thomas Mock, Francis Mulholland, Stanislav Kopriva, and Gill Malin. 2012. 'The Response of Diatom Central Carbon Metabolism to Nitrogen Starvation Is Different from That of Green Algae and Higher Plants'. *Plant Physiology* 158 (1): 299–312. <https://doi.org/10.1104/pp.111.184333>.
- Hoffmann, L., C. Billard, M. Janssens, M. Leruth, and V. Demoulin. 2000. 'Mass Development of Marine Benthic *Sarcinochrysidales* (*Chrysophyceae* s.l.) in Corsica' 43 (3): 223–31. <https://doi.org/10.1515/BOT.2000.024>.
- Honda, Daisuke, and Isao Inouye. 1995. 'Ultrastructure and Reconstruction of the Flagellar Apparatus Architecture in *Ankylochrysis Lutea* (*Chrysophyceae*, *Sarcinochrysidales*)'. *Phycologia* 34 (3): 215–27.

- <https://doi.org/10.2216/i0031-8884-34-3-215.1>.
- Hu, Zhangxi, Shunshan Duan, Ning Xu, and Margaret R. Mulholland. 2014. 'Growth and Nitrogen Uptake Kinetics in Cultured *Prorocentrum Donghaiense*'. *PLOS ONE* 9 (4): e94030. <https://doi.org/10.1371/journal.pone.0094030>.
- Hu, Zhangxi, Margaret R. Mulholland, Shunshan Duan, and Ning Xu. 2012. 'Effects of Nitrogen Supply and Its Composition on the Growth of *Prorocentrum Donghaiense*'. *Harmful Algae* 13 (January): 72–82. <https://doi.org/10.1016/j.hal.2011.10.004>.
- Hutchins, D. A., and P. W. Boyd. 2016. 'Marine Phytoplankton and the Changing Ocean Iron Cycle'. *Nature Climate Change* 6 (12): 1072–79. <https://doi.org/10.1038/nclimate3147>.
- Hutchins, David A., and Douglas G. Capone. 2022. 'The Marine Nitrogen Cycle: New Developments and Global Change'. *Nature Reviews Microbiology* 20 (7): 401–14. <https://doi.org/10.1038/s41579-022-00687-z>.
- Isaji, Yuta, C. Yoshikawa, N. O. Ogawa, K. Matsumoto, A. Makabe, S. Toyoda, N. F. Ishikawa, et al. 2022. 'Nitrogen Sources for Phytoplankton in the Eastern Indian Ocean Determined From  $\delta^{15}\text{N}$  of Chlorophyll a and Divinylchlorophyll a'. *Geochemistry, Geophysics, Geosystems* 23 (4): e2021GC010057. <https://doi.org/10.1029/2021GC010057>.
- Jabre, Loay J., Andrew E. Allen, J. Scott P. McCain, John P. McCrow, Nancy Tenenbaum, Jenna L. Spackeen, Rachel E. Sipler, et al. 2021. 'Molecular Underpinnings and Biogeochemical Consequences of Enhanced Diatom Growth in a Warming Southern Ocean'. *Proceedings of the National Academy of Sciences* 118 (30): e2107238118. <https://doi.org/10.1073/pnas.2107238118>.
- Jannasch, Holger W., and Galen E. Jones. 1959. 'Bacterial Populations in Sea Water as Determined by Different Methods of Enumeration1'. *Limnology and Oceanography* 4 (2): 128–39. <https://doi.org/10.4319/lo.1959.4.2.0128>.
- Jensen, Erik, Romain Clément, Stephen C. Maberly, and Brigitte Gontero. 2017. 'Regulation of the Calvin–Benson–Bassham Cycle in the Enigmatic Diatoms: Biochemical and Evolutionary Variations on an Original Theme'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1728): 20160401. <https://doi.org/10.1098/rstb.2016.0401>.
- Jickells, T. D., Z. S. An, K. K. Andersen, A. R. Baker, G. Bergametti, N. Brooks, J. J. Cao, et al. 2005. 'Global Iron Connections Between Desert Dust, Ocean Biogeochemistry, and Climate'. *Science* 308 (5718): 67–71. <https://doi.org/10.1126/science.1105959>.
- Johnson, W. V., and P. M. Anderson. 1987. 'Bicarbonate Is a Recycling Substrate for Cyanase'. *The Journal of Biological Chemistry* 262 (19): 9021–25.
- Jouenne, Fabien, Wenche Eikrem, Florence Le Gall, Dominique Marie, Geir Johnsen, and Daniel Vaultot. 2011. '*Prasinoderma Singularis* Sp. Nov. (Prasinophyceae, Chlorophyta), a Solitary Coccoid Prasinophyte from the South-East Pacific Ocean'. *Protist* 162 (1): 70–84. <https://doi.org/10.1016/j.protis.2010.04.005>.
- Kamennaya, Nina A., Mark Chernihovsky, and Anton F. Post. 2008. 'The Cyanate Utilization Capacity of Marine Unicellular Cyanobacteria'. *Limnology and Oceanography* 53 (6): 2485–94. <https://doi.org/10.4319/lo.2008.53.6.2485>.
- Kamennaya, Nina A., and Anton F. Post. 2011. 'Characterization of Cyanate Metabolism in Marine *Synechococcus* and *Prochlorococcus* Spp'. *Applied and Environmental Microbiology* 77 (1): 291–301. <https://doi.org/10.1128/AEM.01272-10>.
- Kamp, Anja, Signe Høgslund, Nils Risgaard-Petersen, and Peter Stief. 2015. 'Nitrate Storage and Dissimilatory Nitrate Reduction by Eukaryotic Microbes'. *Frontiers in Microbiology* 6. <https://www.frontiersin.org/articles/10.3389/fmicb.2015.01492>.
- Kang, Yoonja, Matthew J. Harke, Dianna L. Berry, Jackie L. Collier, Steven W. Wilhelm, Sonya T. Dyhrman, and Christopher J. Gobler. 2021. 'Transcriptomic Responses of Four Pelagophytes to Nutrient (N, P) and Light Stress'. *Frontiers in Marine Science* 8. <https://doi.org/10.3389/fmars.2021.636699>.
- Kang, Yoonja, Florian Koch, and Christopher J. Gobler. 2015. 'The Interactive Roles of Nutrient Loading and Zooplankton Grazing in Facilitating the Expansion of Harmful Algal Blooms Caused by the Pelagophyte, *Aureoumbra Lagunensis*, to the Indian River Lagoon, FL, USA'. *Harmful Algae* 49 (November): 162–73. <https://doi.org/10.1016/j.hal.2015.09.005>.
- Karsenti, Eric, Silvia G. Acinas, Peer Bork, Chris Bowler, Colomban De Vargas, Jeroen Raes, Matthew Sullivan, et al. 2011. 'A Holistic Approach to Marine Eco-Systems Biology'. *PLOS Biology* 9 (10): e1001177. <https://doi.org/10.1371/journal.pbio.1001177>.
- Kazamia, Elena, Jan Mach, Jeffrey B. McQuaid, Xia Gao, Tyler H. Coale, Ronald Malych, Jean-Michel Camadro, et al. 2022. 'In Vivo Localization of Iron Starvation Induced Proteins under Variable Iron Supplementation Regimes in *Phaeodactylum Tricornutum*'. *Plant Direct* 6 (12): e472. <https://doi.org/10.1002/pld3.472>.
- Kazamia, Elena, Robert Sutak, Javier Paz-Yepes, Richard G. Dorrell, Fabio Rocha Jimenez Vieira, Jan Mach, Joe Morrissey, et al. 2018. 'Endocytosis-Mediated Siderophore Uptake as a Strategy for Fe Acquisition in Diatoms'. *Science Advances* 4 (5): eaar4536. <https://doi.org/10.1126/sciadv.aar4536>.

- Keeling, Patrick J., Fabien Burki, Heather M. Wilcox, Bassem Allam, Eric E. Allen, Linda A. Amaral-Zettler, E. Virginia Armbrust, et al. 2014. 'The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing'. *PLoS Biology* 12 (6): e1001889. <https://doi.org/10.1371/journal.pbio.1001889>.
- Keeling, Patrick J., and Javier del Campo. 2017. 'Marine Protists Are Not Just Big Bacteria'. *Current Biology* 27 (11): R541–49. <https://doi.org/10.1016/j.cub.2017.03.075>.
- Keffer, Jessica L., Sonia Huecas, Jared T. Hammill, Peter Wipf, Jose M. Andreu, and Carole A. Bewley. 2013. 'Chrysophaeonts Are Competitive Inhibitors of FtsZ and Inhibit Z-Ring Formation in Live Bacteria'. *Bioorganic & Medicinal Chemistry* 21 (18): 5673–78. <https://doi.org/10.1016/j.bmc.2013.07.033>.
- Kemp, Alan E. S., and Tracy A. Villareal. 2018. 'The Case of the Diatoms and the Muddled Mandalas: Time to Recognize Diatom Adaptations to Stratified Waters'. *Progress in Oceanography* 167 (October): 138–49. <https://doi.org/10.1016/j.pocean.2018.08.002>.
- Kirkham, Amy R., Cécile Lepère, Ludwig E. Jardillier, Fabrice Not, Heather Bouman, Andrew Mead, and David J. Scanlan. 2013. 'A Global Perspective on Marine Photosynthetic Picoeukaryote Community Structure'. *The ISME Journal* 7 (5): 922–36. <https://doi.org/10.1038/ismej.2012.166>.
- Kopp, C., M. Pernice, I. Domart-Coulon, C. Djediat, J. E. Spangenberg, D. T. L. Alexander, M. Hignette, T. Meziane, and A. Meibom. 2013. 'Highly Dynamic Cellular-Level Response of Symbiotic Coral to a Sudden Increase in Environmental Nitrogen'. *mBio* 4 (3): 10.1128/mbio.00052-13. <https://doi.org/10.1128/mbio.00052-13>.
- Kopylova, Evguenia, Laurent Noé, and Hélène Touzet. 2012. 'SortMeRNA: Fast and Accurate Filtering of Ribosomal RNAs in Metatranscriptomic Data'. *Bioinformatics (Oxford, England)* 28 (24): 3211–17. <https://doi.org/10.1093/bioinformatics/bts611>.
- Kumar, Ajay, and Surojit Bera. 2020. 'Revisiting Nitrogen Utilization in Algae: A Review on the Process of Regulation and Assimilation'. *Bioresource Technology Reports* 12 (December): 100584. <https://doi.org/10.1016/j.biteb.2020.100584>.
- Kwiatkowski, Lester, Olivier Torres, Laurent Bopp, Olivier Aumont, Matthew Chamberlain, James R. Christian, John P. Dunne, et al. 2020. 'Twenty-First Century Ocean Warming, Acidification, Deoxygenation, and Upper-Ocean Nutrient and Primary Production Decline from CMIP6 Model Projections'. *Biogeosciences* 17 (13): 3439–70. <https://doi.org/10.5194/bg-17-3439-2020>.
- Lampe, Robert H., Seaver Wang, Nicolas Cassar, and Adrian Marchetti. 2019. 'Strategies among Phytoplankton in Response to Alleviation of Nutrient Stress in a Subtropical Gyre'. *The ISME Journal* 13 (12): 2984–97. <https://doi.org/10.1038/s41396-019-0489-6>.
- Langmead, Ben, and Steven L. Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- Lelandais, Gaëlle, Ivo Scheiber, Javier Paz-Yepes, Jean-Claude Lozano, Hugo Botebol, Jana Pilátová, Vojtěch Žárský, et al. 2016. 'Ostreococcus Tauri Is a New Model Green Alga for Studying Iron Metabolism in Eukaryotic Phytoplankton'. *BMC Genomics* 17 (1): 319. <https://doi.org/10.1186/s12864-016-2666-6>.
- Levitán, Orly, Jorge Dinamarca, Ehud Zelzion, Desmond S. Lun, L. Tiago Guerra, Min Kyung Kim, Joomi Kim, Benjamin A. S. Van Mooy, Debashish Bhattacharya, and Paul G. Falkowski. 2015. 'Remodeling of Intermediate Metabolism in the Diatom Phaeodactylum Tricornutum under Nitrogen Stress'. *Proceedings of the National Academy of Sciences* 112 (2): 412–17. <https://doi.org/10.1073/pnas.1419818112>.
- Levitus, Sydney, Margarita E. Conkright, Joseph L. Reid, Raymond G. Najjar, and Arnold Mantyla. 1993. 'Distribution of Nitrate, Phosphate and Silicate in the World Oceans'. *Progress in Oceanography* 31 (3): 245–73. [https://doi.org/10.1016/0079-6611\(93\)90003-V](https://doi.org/10.1016/0079-6611(93)90003-V).
- Lewin, Joyce, Richard E. Norris, S. W. Jeffrey, and Barbara E. Pearson. 1977. 'An Aberrant Chrysophycean Alga Pelagococcus Subviridis Gen. Nov. Et Sp. Nov. from the North Pacific Ocean<sup>1,2</sup>'. *Journal of Phycology* 13 (3): 259–66. <https://doi.org/10.1111/j.1529-8817.1977.tb02925.x>.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, Xiaobo, Weronika Patena, Friedrich Fauser, Robert E. Jinkerson, Shai Saroussi, Moritz T. Meyer, Nina Ivanova, et al. 2019. 'A Genome-Wide Algal Mutant Library and Functional Screen Identifies Genes Required for Eukaryotic Photosynthesis'. *Nature Genetics* 51 (4): 627–35. <https://doi.org/10.1038/s41588-019-0370-6>.
- Linder, Tomas. 2018. 'Cyanase-Independent Utilization of Cyanate as a Nitrogen Source in Ascomycete Yeasts'. *World Journal of Microbiology and Biotechnology* 35 (1): 3. <https://doi.org/10.1007/s11274-018-2579-4>.
- Litchman, Elena, Christopher A. Klausmeier, Oscar M. Schofield, and Paul G. Falkowski. 2007. 'The Role of Functional Traits and Trade-Offs in Structuring Phytoplankton Communities: Scaling from Cellular to Ecosystem Level'. *Ecology Letters* 10 (12): 1170–81. <https://doi.org/10.1111/j.1461-0248.2007.01117.x>.
- Liu, Hongbin, and Edward J. Buskey. 2000. 'The Exopolymer Secretions (EPS) Layer Surrounding Aureoumbra

- Lagunensis Cells Affects Growth, Grazing, and Behavior of Protozoa'. *Limnology and Oceanography* 45 (5): 1187–91. <https://doi.org/10.4319/lo.2000.45.5.1187>.
- Lobban, Christopher, Daisuke Honda, Chihara M, and Scheffter M. 1995. 'Chrysocystis Fragilis Gen. Nov., Sp. Nov. (Chrysophyceae, Sarcinochrysidales), with Notes on Other Macroscopic Chrysophytes (Golden Algae) on Guam Reefs'. *Micronesica* 28 (January): 91–102.
- Loos, Luna M. van der, and Reindert Nijland. 2021. 'Biases in Bulk: DNA Metabarcoding of Marine Communities and the Methodology Involved'. *Molecular Ecology* 30 (13): 3270–88. <https://doi.org/10.1111/mec.15592>.
- López-García, Purificación, Francisco Rodríguez-Valera, Carlos Pedrós-Alió, and David Moreira. 2001. 'Unexpected Diversity of Small Eukaryotes in Deep-Sea Antarctic Plankton'. *Nature* 409 (6820): 603–7. <https://doi.org/10.1038/35054537>.
- Louca, Stilianos, Laura Wegener Parfrey, and Michael Doebeli. 2016. 'Decoupling Function and Taxonomy in the Global Ocean Microbiome'. *Science* 353 (6305): 1272–77. <https://doi.org/10.1126/science.aaf4507>.
- Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. 'Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2'. *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Madoui, Mohammed-Amin, Julie Poulain, Kevin Sugier, Marc Wessner, Benjamin Noel, Leo Berline, Karine Labadie, et al. 2017. 'New Insights into Global Biogeography, Population Structure and Natural Selection from the Genome of the Epipelagic Copepod Oithona'. *Molecular Ecology* 26 (17): 4467–82. <https://doi.org/10.1111/mec.14214>.
- Mäki, Anita, Pauliina Salmi, Anu Mikkonen, Anke Kremp, and Marja Tirola. 2017. 'Sample Preservation, DNA or RNA Extraction and Data Analysis for High-Throughput Phytoplankton Community Sequencing'. *Frontiers in Microbiology* 8. <https://www.frontiersin.org/articles/10.3389/fmicb.2017.01848>.
- Mao, Xuewei, Jianwei Chen, Cock van Oosterhout, Huan Zhang, Guangxing Liu, Yunyun Zhuang, and Thomas Mock. 2022. 'Diversity, Prevalence, and Expression of Cyanase Genes (cynS) in Planktonic Marine Microorganisms'. *The ISME Journal* 16 (2): 602–5. <https://doi.org/10.1038/s41396-021-01081-y>.
- Martin, John H., and Steve E. Fitzwater. 1988. 'Iron Deficiency Limits Phytoplankton Growth in the North-East Pacific Subarctic'. *Nature* 331 (6154): 341–43. <https://doi.org/10.1038/331341a0>.
- McCarthy, James K., Sarah R. Smith, John P. McCrow, Maxine Tan, Hong Zheng, Karen Beeri, Robyn Roth, et al. 2017. 'Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom'. *The Plant Cell* 29 (8): 2047–70. <https://doi.org/10.1105/tpc.16.00910>.
- McLaren, Michael R, Amy D Willis, and Benjamin J Callahan. 2019. 'Consistent and Correctable Bias in Metagenomic Sequencing Experiments'. Edited by Peter Turnbaugh, Wendy S Garrett, Peter Turnbaugh, Christopher Quince, and Sean Gibbons. *eLife* 8 (September): e46923. <https://doi.org/10.7554/eLife.46923>.
- McQuaid, Jeffrey B., Adam B. Kustka, Miroslav Oborník, Aleš Horák, John P. McCrow, Bogumil J. Karas, Hong Zheng, et al. 2018. 'Carbonate-Sensitive Phytotransferrin Controls High-Affinity Iron Uptake in Diatoms'. *Nature* 555 (7697): 534–37. <https://doi.org/10.1038/nature25982>.
- Meyer, Nils, Aljoscha Rydzyk, and Georg Pohnert. 2022. 'Pronounced Uptake and Metabolism of Organic Substrates by Diatoms Revealed by Pulse-Labeling Metabolomics'. *Frontiers in Marine Science* 9. <https://www.frontiersin.org/articles/10.3389/fmars.2022.821167>.
- Miller, Anthony G., and George S. Espie. 1994. 'Photosynthetic Metabolism of Cyanate by the Cyanobacterium Synechococcus UTEX 625'. *Archives of Microbiology* 162 (3): 151–57. <https://doi.org/10.1007/BF00314468>.
- Mock, Thomas, Manoj Pratim Samanta, Vaughn Iverson, Chris Berthiaume, Matthew Robison, Karie Holtermann, Colleen Durkin, et al. 2008. 'Whole-Genome Expression Profiling of the Marine Diatom Thalassiosira Pseudonana Identifies Genes Involved in Silicon Bioprocesses'. *Proceedings of the National Academy of Sciences* 105 (5): 1579–84. <https://doi.org/10.1073/pnas.0707946105>.
- Moon-van der Staay, Seung Yeo, Rupert De Wachter, and Daniel Vaultot. 2001. 'Oceanic 18S rDNA Sequences from Picoplankton Reveal Unsuspected Eukaryotic Diversity'. *Nature* 409 (6820): 607–10. <https://doi.org/10.1038/35054541>.
- Moore, C. M., M. M. Mills, K. R. Arrigo, I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, et al. 2013. 'Processes and Patterns of Oceanic Nutrient Limitation'. *Nature Geoscience* 6 (9): 701–10. <https://doi.org/10.1038/ngeo1765>.
- Morán, Xosé Anxelu G., Marta Sebastián, Carlos Pedrís-Alió, and Marta Estrada. 2006. 'Response of Southern Ocean Phytoplankton and Bacterioplankton Production to Short-Term Experimental Warming'. *Limnology and Oceanography* 51 (4): 1791–1800. <https://doi.org/10.4319/lo.2006.51.4.1791>.
- Morris, J. Jeffrey, Richard E. Lenski, and Erik R. Zinser. 2012. 'The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss'. *mBio* 3 (2): 10.1128/mbio.00036-12. <https://doi.org/10.1128/mbio.00036-12>.
- Morrissey, Joe, and Chris Bowler. 2012. 'Iron Utilization in Marine Cyanobacteria and Eukaryotic Algae'. *Frontiers in Microbiology* 3. <https://doi.org/10.3389/fmicb.2012.00043>.
- Muhlstein, Harris I., and Tracy A. Villareal. 2007. 'Organic and Inorganic Nutrient Effects on Growth Rate–Irradiance

- Relationships in the Texas Brown-Tide Alga *Aureoumbra Lagunensis* (Pelagophyceae)<sup>1</sup>. *Journal of Phycology* 43 (6): 1223–26. <https://doi.org/10.1111/j.1529-8817.2007.00412.x>.
- Not, Fabrice, Javier del Campo, Vanessa Balagué, Colomban de Vargas, and Ramon Massana. 2009. 'New Insights into the Diversity of Marine Picoeukaryotes'. *PLOS ONE* 4 (9): e7143. <https://doi.org/10.1371/journal.pone.0007143>.
- Nouri, Mohammad-Zaman, Ali Moumeni, and Setsuko Komatsu. 2015. 'Abiotic Stresses: Insight into Gene Regulation and Protein Expression in Photosynthetic Pathways of Plants'. *International Journal of Molecular Sciences* 16 (9): 20392–416. <https://doi.org/10.3390/ijms160920392>.
- Olsen, G J, D J Lane, S J Giovannoni, N R Pace, and D A Stahl. 1986. 'Microbial Ecology and Evolution: A Ribosomal RNA Approach'. *Annual Review of Microbiology* 40 (1): 337–65. <https://doi.org/10.1146/annurev.mi.40.100186.002005>.
- Ong, Han Chuan, Steven W. Wilhelm, Christopher J. Gobler, George Bullerjahn, Michael A. Jacobs, John McKay, Elizabeth H. Sims, et al. 2010. 'Analyses of the Complete Chloroplast Genome Sequences of Two Members of the Pelagophyceae: *Aureococcus Anophagefferens* Ccmp1984 and *Aureoumbra Lagunensis* Ccmp15071'. *Journal of Phycology* 46 (3): 602–15. <https://doi.org/10.1111/j.1529-8817.2010.00841.x>.
- Ou, Linjian, Yangyang Cai, Wenyu Jin, Zhou Wang, and Songhui Lu. 2018. 'Understanding the Nitrogen Uptake and Assimilation of the Chinese Strain of *Aureococcus Anophagefferens* (Pelagophyceae)'. *Algal Research* 34 (September): 182–90. <https://doi.org/10.1016/j.algal.2018.07.019>.
- Palatinszky, Marton, Craig Herbold, Nico Jehmlich, Mario Pogoda, Ping Han, Martin von Bergen, Ilias Lagkouvardos, et al. 2015. 'Cyanate as an Energy Source for Nitrifiers'. *Nature* 524 (7563): 105–8. <https://doi.org/10.1038/nature14856>.
- Palenik, Brian, and Sarah E. Henson. 1997. 'The Use of Amides and Other Organic Nitrogen Sources by the Phytoplankton *Emiliana Huxleyi*'. *Limnology and Oceanography* 42 (7): 1544–51. <https://doi.org/10.4319/lo.1997.42.7.1544>.
- Pesant, Stéphane, Fabrice Not, Marc Picheral, Stefanie Kandels-Lewis, Noan Le Bescot, Gabriel Gorsky, Daniele Iudicone, et al. 2015. 'Open Science Resources for the Discovery and Analysis of Tara Oceans Data'. *Scientific Data* 2 (1): 150023. <https://doi.org/10.1038/sdata.2015.23>.
- Peter, Kalista Higinis, and Ulrich Sommer. 2013. 'Phytoplankton Cell Size Reduction in Response to Warming Mediated by Nutrient Limitation'. *PLOS ONE* 8 (9): e71528. <https://doi.org/10.1371/journal.pone.0071528>.
- Pierella Karlusich, Juan José, Federico M. Ibarbalz, and Chris Bowler. 2020. 'Phytoplankton in the Tara Ocean'. *Annual Review of Marine Science* 12 (1): 233–65. <https://doi.org/10.1146/annurev-marine-010419-010706>.
- Pilátová, Jana, Tomáš Pánek, Miroslav Oborník, Ivan Čepička, and Peter Mojžeš. 2022. 'Revisiting Biocrystallization: Purine Crystalline Inclusions Are Widespread in Eukaryotes'. *The ISME Journal* 16 (9): 2290–94. <https://doi.org/10.1038/s41396-022-01264-1>.
- Pinto, Jordan, Raphaël Lami, Marc Krasovec, Régis Grimaud, Laurent Urios, Josselin Lupette, Marie-Line Escande, et al. 2021. 'Features of the Opportunistic Behaviour of the Marine Bacterium *Marinobacter Algicola* in the Microalga *Ostreococcus Tauri* Phycosphere'. *Microorganisms* 9 (8): 1777. <https://doi.org/10.3390/microorganisms9081777>.
- Polovina, Jeffrey J., Evan A. Howell, and Melanie Abecassis. 2008. 'Ocean's Least Productive Waters Are Expanding'. *Geophysical Research Letters* 35 (3). <https://doi.org/10.1029/2007GL031745>.
- Poretzky, Rachel S., Nasreen Bano, Alison Buchan, Gary LeCleir, Jutta Kleikemper, Maria Pickering, Whitney M. Pate, Mary Ann Moran, and James T. Hollibaugh. 2005. 'Analysis of Microbial Gene Transcripts in Environmental Samples'. *Applied and Environmental Microbiology* 71 (7): 4121–26. <https://doi.org/10.1128/AEM.71.7.4121-4126.2005>.
- Price, Neil M., and Paul J. Harrison. 1988. 'Uptake of Urea C and Urea N by the Coastal Marine Diatom *Thalassiosira Pseudonana*'. *Limnology and Oceanography* 33 (4): 528–37. <https://doi.org/10.4319/lo.1988.33.4.0528>.
- Props, Ruben, Frederiek-Maarten Kerckhof, Peter Rubbens, Jo De Vrieze, Emma Hernandez Sanabria, Willem Waegeman, Pieter Monsieurs, Frederik Hammes, and Nico Boon. 2017. 'Absolute Quantification of Microbial Taxon Abundances'. *The ISME Journal* 11 (2): 584–87. <https://doi.org/10.1038/ismej.2016.117>.
- Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. 'The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools'. *Nucleic Acids Research* 41 (D1): D590–96. <https://doi.org/10.1093/nar/gks1219>.
- Raven, J. A. 1998. 'The Twelfth Tansley Lecture. Small Is Beautiful: The Picophytoplankton'. *Functional Ecology* 12 (4): 503–13. <https://doi.org/10.1046/j.1365-2435.1998.00233.x>.
- Raven, J. A., B. Wollenweber, and L. L. Handley. 1993. 'The Quantitative Role of Ammonia/Ammonium Transport and Metabolism by Plants in the Global Nitrogen Cycle'. *Physiologia Plantarum* 89 (3): 512–18. <https://doi.org/10.1111/j.1399-3054.1993.tb05207.x>.
- Raven, John A. 2013. 'Iron Acquisition and Allocation in Stramenopile Algae'. *Journal of Experimental Botany* 64 (8):

- 2119–27. <https://doi.org/10.1093/jxb/ert121>.
- Raven, John A., Bernd Wollenweber, and Linda L. Handley. 1992. 'A Comparison of Ammonium and Nitrate as Nitrogen Sources for Photolithotrophs'. *New Phytologist* 121 (1): 19–32. <https://doi.org/10.1111/j.1469-8137.1992.tb01088.x>.
- Remmers, Ilse M., Sarah D'Adamo, Dirk E. Martens, Ric C. H. de Vos, Roland Mumm, Antoine H. P. America, Jan H. G. Cordewener, et al. 2018. 'Orchestration of Transcriptome, Proteome and Metabolome in the Diatom *Phaeodactylum Tricornutum* during Nitrogen Limitation'. *Algal Research* 35 (November): 33–49. <https://doi.org/10.1016/j.algal.2018.08.012>.
- Rii, Yoshimi M., Solange Duhamel, Robert R. Bidigare, David M. Karl, Daniel J. Repeta, and Matthew J. Church. 2016. 'Diversity and Productivity of Photosynthetic Picoeukaryotes in Biogeochemically Distinct Regions of the South East Pacific Ocean'. *Limnology and Oceanography* 61 (3): 806–24. <https://doi.org/10.1002/lno.10255>.
- Rogato, Alessandra, Alberto Amato, Daniele Iudicone, Maurizio Chiurazzi, Maria Immacolata Ferrante, and Maurizio Ribera d'Alcalà. 2015. 'The Diatom Molecular Toolkit to Handle Nitrogen Uptake'. *Marine Genomics, SI:StructureRegulationEvolution*, 24 (December): 95–108. <https://doi.org/10.1016/j.margen.2015.05.018>.
- Rusch, Douglas B., Aaron L. Halpern, Granger Sutton, Karla B. Heidelberg, Shannon Williamson, Shibu Yooseph, Dongying Wu, et al. 2007. 'The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific'. *PLOS Biology* 5 (3): e77. <https://doi.org/10.1371/journal.pbio.0050077>.
- Sahlin, Kristoffer, and Paul Medvedev. 2021. 'Error Correction Enables Use of Oxford Nanopore Technology for Reference-Free Transcriptome Analysis'. *Nature Communications* 12 (1): 2. <https://doi.org/10.1038/s41467-020-20340-8>.
- Santi, Ioulia, Panagiotis Kasapidis, Ioannis Karakassis, and Paraskevi Pitta. 2021. 'A Comparison of DNA Metabarcoding and Microscopy Methodologies for the Study of Aquatic Microbial Eukaryotes'. *Diversity* 13 (5): 180. <https://doi.org/10.3390/d13050180>.
- Santin, Anna, Luigi Caputi, Antonella Longo, Maurizio Chiurazzi, Maurizio Ribera d'Alcalà, Monia Teresa Russo, Maria Immacolata Ferrante, and Alessandra Rogato. 2021. 'Integrative Omics Identification, Evolutionary and Structural Analysis of Low Affinity Nitrate Transporters in Diatoms, diNPFs'. *Open Biology* 11 (4): 200395. <https://doi.org/10.1098/rsob.200395>.
- Sanz-Luque, Emanuel, Alejandro Chamizo-Ampudia, Angel Llamas, Aurora Galvan, and Emilio Fernandez. 2015. 'Understanding Nitrate Assimilation and Its Regulation in Microalgae'. *Frontiers in Plant Science* 6. <https://www.frontiersin.org/articles/10.3389/fpls.2015.00899>.
- Sato, Mitsuhide, Fuminori Hashihama, and Shigenobu Takeda. 2023. 'Effects of Cyanate Enrichment on Growth of Natural Phytoplankton Populations in the Subtropical Pacific'. *Journal of Oceanography* 79 (1): 1–12. <https://doi.org/10.1007/s10872-022-00658-1>.
- Scarsini, Matteo, Stanislas Thiriet-Rupert, Brigitte Veidl, Florence Mondeguer, Hanhua Hu, Justine Marchand, and Benoît Schoefs. 2022. 'The Transition Toward Nitrogen Deprivation in Diatoms Requires Chloroplast Stand-By and Deep Metabolic Reshuffling'. *Frontiers in Plant Science* 12. <https://doi.org/10.3389/fpls.2021.760516>.
- Schaffelke, B., K. Heimann, P. A. Marshall, and A. M. Ayling. 2004. 'Blooms of *Chrysochromulina fragilis* on the Great Barrier Reef'. *Coral Reefs* 23 (4): 514–514. <https://doi.org/10.1007/s00338-004-0439-1>.
- Segura Broncano, Laia, Krzysztof Robin Pukacz, Vanessa Reichel-Deland, Urte Schlüter, Sebastian Triesch, and Andreas P. M. Weber. 2023. 'Photorespiration Is the Solution, Not the Problem'. *Journal of Plant Physiology* 282 (March): 153928. <https://doi.org/10.1016/j.jplph.2023.153928>.
- Shaked, Yeala, and Hagar Lis. 2012. 'Disassembling Iron Availability to Phytoplankton'. *Frontiers in Microbiology* 3. <https://www.frontiersin.org/articles/10.3389/fmicb.2012.00123>.
- Shulka, P. R., J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, et al. 2019. 'IPCC, 2019: Summary for Policymakers. In: Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems'. IPCC.
- Shuter, B. J. 1978. 'Size Dependence of Phosphorus and Nitrogen Subsistence Quotas in Unicellular Microorganisms 1'. *Limnology and Oceanography* 23 (6): 1248–55. <https://doi.org/10.4319/lo.1978.23.6.1248>.
- Sibbald, Shannon J, Maggie Lawton, and John M Archibald. 2021. 'Mitochondrial Genome Evolution in Pelagophyte Algae'. *Genome Biology and Evolution* 13 (3). <https://doi.org/10.1093/gbe/evab018>.
- Sipler, Rachel E., and Deborah A. Bronk. 2015. 'Chapter 4 - Dynamics of Dissolved Organic Nitrogen'. In *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)*, edited by Dennis A. Hansell and Craig A. Carlson, 127–232. Boston: Academic Press. <https://doi.org/10.1016/B978-0-12-405940-5.00004-2>.
- Smith, Sarah R., Chris L. Dupont, James K. McCarthy, Jared T. Brodrick, Miroslav Oborník, Aleš Horák, Zoltán Füssy, et al. 2019. 'Evolution and Regulation of Nitrogen Flux through Compartmentalized Metabolic Networks in a Marine Diatom'. *Nature Communications* 10 (1): 4552. <https://doi.org/10.1038/s41467-019-12407-y>.
- Sproles, Ashley E., Francis J. Fields, Tressa N. Smalley, Chau H. Le, Amr Badary, and Stephen P. Mayfield. 2021. 'Recent

- Advancements in the Genetic Engineering of Microalgae'. *Algal Research* 53 (March): 102158. <https://doi.org/10.1016/j.algal.2020.102158>.
- Staley, James T., and Allan Konopka. 1985. 'Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats'. *Annual Review of Microbiology* 39 (1): 321–46. <https://doi.org/10.1146/annurev.mi.39.100185.001541>.
- Stoecker, Diane K. 1999. 'Mixotrophy among Dinoflagellates'. *Journal of Eukaryotic Microbiology* 46 (4): 397–401. <https://doi.org/10.1111/j.1550-7408.1999.tb04619.x>.
- Stoler, Nicholas, and Anton Nekrutenko. 2021. 'Sequencing Error Profiles of Illumina Sequencing Instruments'. *NAR Genomics and Bioinformatics* 3 (1): lqab019. <https://doi.org/10.1093/nargab/lqab019>.
- Sun, Liqin, Ling Wang, Jing Li, and Honghui Liu. 2013. 'Immunomodulation and Antitumor Activities of Degraded Polysaccharide from Marine Microalgae *Sarcinochrysis Marina* Geitler'. *International Journal Bioautomation* 17 (July): 107–16.
- Sun, Ning, Ligeng Ma, Deyun Pan, Hongyu Zhao, and Xing Wang Deng. 2003. 'Evaluation of Light Regulatory Potential of Calvin Cycle Steps Based on Large-Scale Gene Expression Profiling Data'. *Plant Molecular Biology* 53 (4): 467–78. <https://doi.org/10.1023/B:PLAN.0000019071.12878.9e>.
- Sunagawa, Shinichi, Silvia G. Acinas, Peer Bork, Chris Bowler, Damien Eveillard, Gabriel Gorsky, Lionel Guidi, et al. 2020. 'Tara Oceans: Towards Global Ocean Ecosystems Biology'. *Nature Reviews Microbiology* 18 (8): 428–45. <https://doi.org/10.1038/s41579-020-0364-5>.
- Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, et al. 2015. 'Structure and Function of the Global Ocean Microbiome'. *Science* 348 (6237): 1261359. <https://doi.org/10.1126/science.1261359>.
- Tamoi, Masahiro, and Shigeru Shigeoka. 2015. 'Diversity of Regulatory Mechanisms of Photosynthetic Carbon Metabolism in Plants and Algae'. *Bioscience, Biotechnology, and Biochemistry* 79 (6): 870–76. <https://doi.org/10.1080/09168451.2015.1020754>.
- Tang, Weiyi, Joan Llorc, Jakob Weis, Morgane M. G. Perron, Sara Basart, Zuchuan Li, Shubha Sathyendranath, et al. 2021. 'Widespread Phytoplankton Blooms Triggered by 2019–2020 Australian Wildfires'. *Nature* 597 (7876): 370–75. <https://doi.org/10.1038/s41586-021-03805-8>.
- Ter-Ovanesian, Louis M. P., Baptiste Rigaud, Alberto Mezzetti, Jean-François Lambert, and Marie-Christine Maurel. 2021. 'Carbamoyl Phosphate and Its Substitutes for the Uracil Synthesis in Origins of Life Scenarios'. *Scientific Reports* 11 (1): 19356. <https://doi.org/10.1038/s41598-021-98747-6>.
- Terrado, Ramon, Adam Monier, Robyn Edgar, and Connie Lovejoy. 2015. 'Diversity of Nitrogen Assimilation Pathways among Microbial Photosynthetic Eukaryotes'. *Journal of Phycology* 51 (3): 490–506. <https://doi.org/10.1111/jpy.12292>.
- Thompson, Anne W., Rachel A. Foster, Andreas Krupke, Brandon J. Carter, Niculina Musat, Daniel Vault, Marcel M. M. Kuypers, and Jonathan P. Zehr. 2012. 'Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga'. *Science* 337 (6101): 1546–50. <https://doi.org/10.1126/science.1222700>.
- Thornton, Daniel C.O. 2014. 'Dissolved Organic Matter (DOM) Release by Phytoplankton in the Contemporary and Future Ocean'. *European Journal of Phycology* 49 (1): 20–46. <https://doi.org/10.1080/09670262.2013.875596>.
- Timmermans, K. R., B. van der Wagt, M. J. W. Veldhuis, A. Maatman, and H. J. W. de Baar. 2005. 'Physiological Responses of Three Species of Marine Pico-Phytoplankton to Ammonium, Phosphate, Iron and Light Limitation'. *Journal of Sea Research, Iron Resources and Oceanic Nutrients - Advancement of Global Environmental Simulations*, 53 (1): 109–20. <https://doi.org/10.1016/j.seares.2004.05.003>.
- Toulza, Eve, Alessandro Tagliabue, Stéphane Blain, and Gwenael Piganeau. 2012. 'Analysis of the Global Ocean Sampling (GOS) Project for Trends in Iron Uptake by Surface Ocean Microbes'. *PLOS ONE* 7 (2): e30931. <https://doi.org/10.1371/journal.pone.0030931>.
- Trefault, Nicole, Rodrigo De la Iglesia, Mario Moreno-Pino, Adriana Lopes dos Santos, Catherine Gérikas Ribeiro, Génesis Parada-Pozo, Antonia Cristi, Dominique Marie, and Daniel Vault. 2021. 'Annual Phytoplankton Dynamics in Coastal Waters from Fildes Bay, Western Antarctic Peninsula'. *Scientific Reports* 11 (1): 1368. <https://doi.org/10.1038/s41598-020-80568-8>.
- Tuo, Sing-how, Yuh-ling Lee Chen, and Houng-Yung Chen. 2014. 'Low Nitrate Availability Promotes Diatom Diazotroph Associations in the Marginal Seas of the Western Pacific'. *Aquatic Microbial Ecology* 73 (2): 135–50. <https://doi.org/10.3354/ame01715>.
- Turk-Kubo, Kendra A., Matthew M. Mills, Kevin R. Arrigo, Gert van Dijken, Britt A. Henke, Brittany Stewart, Samuel T. Wilson, and Jonathan P. Zehr. 2021. 'UCYN-A/Haptophyte Symbioses Dominate N<sub>2</sub> Fixation in the Southern California Current System'. *ISME Communications* 1 (1): 1–13. <https://doi.org/10.1038/s43705-021-00039-7>.
- Tyson, Gene W., Jarrod Chapman, Philip Hugenholtz, Eric E. Allen, Rachna J. Ram, Paul M. Richardson, Victor V. Solovyev, Edward M. Rubin, Daniel S. Rokhsar, and Jillian F. Banfield. 2004. 'Community Structure and Metabolism through Reconstruction of Microbial Genomes from the Environment'. *Nature* 428 (6978): 37–43.

- <https://doi.org/10.1038/nature02340>.
- Uwizeye, Clarisse, Johan Decelle, Pierre-Henri Jouneau, Serena Flori, Benoit Gallet, Jean-Baptiste Keck, Davide Dal Bo, et al. 2021. 'Morphological Bases of Phytoplankton Energy Management and Physiological Responses Unveiled by 3D Subcellular Imaging'. *Nature Communications* 12 (1): 1049. <https://doi.org/10.1038/s41467-021-21314-0>.
- Vacant, Sophie, L. Felipe Benites, Christophe Salmeron, Laurent Intertaglia, Manon Norest, Adrien Cadoudal, Frederic Sanchez, Carlos Caceres, and Gwenael Piganeau. 2022. 'Long-Term Stability of Bacterial Associations in a Microcosm of *Ostreococcus Tauri* (Chlorophyta, Mamiellophyceae)'. *Frontiers in Plant Science* 13. <https://www.frontiersin.org/articles/10.3389/fpls.2022.814386>.
- Van de Waal, Dedmer B., and Elena Litchman. 2020. 'Multiple Global Change Stressor Effects on Phytoplankton Nutrient Acquisition in a Future Ocean'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1798): 20190706. <https://doi.org/10.1098/rstb.2019.0706>.
- Vandeputte, Doris, Gunter Kathagen, Kevin D'hoë, Sara Vieira-Silva, Mireia Valles-Colomer, João Sabino, Jun Wang, et al. 2017. 'Quantitative Microbiome Profiling Links Gut Community Variation to Microbial Load'. *Nature* 551 (7681): 507–11. <https://doi.org/10.1038/nature24460>.
- Vannier, Thomas, Jade Leconte, Yoann Seeleuthner, Samuel Mondy, Eric Pelletier, Jean-Marc Aury, Colomban de Vargas, et al. 2016. 'Survey of the Green Picoalga Bathycoccus Genomes in the Global Ocean'. *Scientific Reports* 6 (1): 37900. <https://doi.org/10.1038/srep37900>.
- Varasteh, Tooba, Diogo Tschoeke, Arthur W. Silva-Lima, Cristiane Thompson, and Fabiano Thompson. 2022. 'Transcriptome of the Coral *Mussismilia Braziliensis* Symbiont *Sargassococcus Simulans*'. *Marine Genomics* 61 (February): 100912. <https://doi.org/10.1016/j.margen.2021.100912>.
- Varela, Marta M., A. Bode, E. Fernández, N. González, V. Kitidis, M. Varela, and E. M. S. Woodward. 2005. 'Nitrogen Uptake and Dissolved Organic Nitrogen Release in Planktonic Communities Characterised by Phytoplankton Size–Structure in the Central Atlantic Ocean'. *Deep Sea Research Part I: Oceanographic Research Papers* 52 (9): 1637–61. <https://doi.org/10.1016/j.dsr.2005.03.005>.
- Vargas, Colomban de, Stéphane Audic, Nicolas Henry, Johan Decelle, Frédéric Mahé, Ramiro Logares, Enrique Lara, et al. 2015. 'Eukaryotic Plankton Diversity in the Sunlit Ocean'. *Science* 348 (6237). <https://doi.org/10.1126/science.1261605>.
- Vaulot, Daniel, Wenche Eikrem, Manon Viprey, and Hervé Moreau. 2008. 'The Diversity of Small Eukaryotic Phytoplankton ( $\leq 3$  Mm) in Marine Ecosystems'. *FEMS Microbiology Reviews* 32 (5): 795–820. <https://doi.org/10.1111/j.1574-6976.2008.00121.x>.
- Vaulot, Daniel, Florence Gall, Dominique Le Marie, Laure Guillou, and Frédéric Partensky. 2004. 'The Roscoff Culture Collection (RCC): A Collection Dedicated to Marine Picoplankton'. *Nova Hedwigia*, August, 49–70. <https://doi.org/10.1127/0029-5035/2004/0079-0049>.
- Veer, J. van der. 1970. 'Ankylonoton Luteum (Chrysophyta), a New Species from the Tamar Estuary, Cornwall'. *Acta Botanica Neerlandica* 19 (5): 616–36.
- Venter, J. Craig, Karin Remington, John F. Heidelberg, Aaron L. Halpern, Doug Rusch, Jonathan A. Eisen, Dongying Wu, et al. 2004. 'Environmental Genome Shotgun Sequencing of the Sargasso Sea'. *Science* 304 (5667): 66–74. <https://doi.org/10.1126/science.1093857>.
- Vesk, Maret, and S. W. Jeffrey. 1987. 'Ultrastructure and Pigments of Two Strains of the Picoplanktonic Alga *Pelagococcus Subviridis* (Chrysophyceae)1'. *Journal of Phycology* 23 (s2): 322–36. <https://doi.org/10.1111/j.1529-8817.1987.tb04141.x>.
- Villar, Emilie, Gregory K. Farrant, Michael Follows, Laurence Garczarek, Sabrina Speich, Stéphane Audic, Lucie Bittner, et al. 2015. 'Environmental Characteristics of Agulhas Rings Affect Interocean Plankton Transport'. *Science* 348 (6237): 1261447. <https://doi.org/10.1126/science.1261447>.
- Vorobev, Alexey, Marion Dupouy, Quentin Carradec, Tom O. Delmont, Anita Annamalé, Patrick Wincker, and Eric Pelletier. 2020. 'Transcriptome Reconstruction and Functional Analysis of Eukaryotic Marine Plankton Communities via High-Throughput Metagenomics and Metatranscriptomics'. *Genome Research* 30 (4): 647–59. <https://doi.org/10.1101/gr.253070.119>.
- Voss, Maren, Hermann W. Bange, Joachim W. Dippner, Jack J. Middelburg, Joseph P. Montoya, and Bess Ward. 2013. 'The Marine Nitrogen Cycle: Recent Discoveries, Uncertainties and the Potential Relevance of Climate Change'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1621): 20130121. <https://doi.org/10.1098/rstb.2013.0121>.
- Walworth, Nathan G., Fei-Xue Fu, Eric A. Webb, Mak A. Saito, Dawn Moran, Matthew R. McIlvin, Michael D. Lee, and David A. Hutchins. 2016. 'Mechanisms of Increased *Trichodesmium* Fitness under Iron and Phosphorus Co-Limitation in the Present and Future Ocean'. *Nature Communications* 7 (1): 12081. <https://doi.org/10.1038/ncomms12081>.
- Walworth, Nathan G., Michael D. Lee, Fei-Xue Fu, David A. Hutchins, and Eric A. Webb. 2016. 'Molecular and

- Physiological Evidence of Genetic Assimilation to High CO<sub>2</sub> in the Marine Nitrogen Fixer *Trichodesmium*'. *Proceedings of the National Academy of Sciences* 113 (47): E7367–74. <https://doi.org/10.1073/pnas.1605202113>.
- Wang, Yingjun, Dan J. Stessman, and Martin H. Spalding. 2015. 'The CO<sub>2</sub> Concentrating Mechanism and Photosynthetic Carbon Assimilation in Limiting CO<sub>2</sub>: How Chlamydomonas Works against the Gradient'. *The Plant Journal* 82 (3): 429–48. <https://doi.org/10.1111/tpj.12829>.
- Wannicke, Nicola, Claudia Frey, Cliff S. Law, and Maren Voss. 2018. 'The Response of the Marine Nitrogen Cycle to Ocean Acidification'. *Global Change Biology* 24 (11): 5031–43. <https://doi.org/10.1111/gcb.14424>.
- Webb, Paul. 2021. *Introduction to Oceanography*. Roger Williams University Pressbooks. <https://rwu.pressbooks.pub/webboceanography/>.
- Weissenbach, Jean, and Abdelghani Sghir. 2016. 'Microbiotes et métagénomique'. *médecine/sciences* 32 (11): 937–43. <https://doi.org/10.1051/medsci/20163211008>.
- Wetherbee, Richard, Trevor T. Bringloe, Joana F. Costa, Allison van de Meene, Robert A. Andersen, and Heroen Verbruggen. 2021. 'New Pelagophytes Show a Novel Mode of Algal Colony Development and Reveal a Perforated Theca That May Define the Class'. *Journal of Phycology* 57 (2): 396–411. <https://doi.org/10.1111/jpy.13074>.
- Wetherbee, Richard, Trevor T. Bringloe, Allison van de Meene, Robert A. Andersen, and Heroen Verbruggen. 2023. 'Structure and Formation of the Perforated Theca Defining the Pelagophyceae (Heterokonta), and Three New Genera That Substantiate the Diverse Nature of the Class'. *Journal of Phycology* 59 (1): 126–51. <https://doi.org/10.1111/jpy.13294>.
- Widner, Brittany, Calvin W. Mordy, and Margaret R. Mulholland. 2018. 'Cyanate Distribution and Uptake above and within the Eastern Tropical South Pacific Oxygen Deficient Zone'. *Limnology and Oceanography* 63 (S1): S177–92. <https://doi.org/10.1002/lno.10730>.
- Widner, Brittany, and Margaret R. Mulholland. 2017. 'Cyanate Distribution and Uptake in North Atlantic Coastal Waters'. *Limnology and Oceanography* 62 (6): 2538–49. <https://doi.org/10.1002/lno.10588>.
- Widner, Brittany, Margaret R. Mulholland, and Kenneth Mopper. 2016. 'Distribution, Sources, and Sinks of Cyanate in the Coastal North Atlantic Ocean'. *Environmental Science & Technology Letters* 3 (8): 297–302. <https://doi.org/10.1021/acs.estlett.6b00165>.
- Wirtz, Kai, S. Lan Smith, Moritz Mathis, and Jan Taucher. 2022. 'Vertically Migrating Phytoplankton Fuel High Oceanic Primary Production'. *Nature Climate Change* 12 (8): 750–56. <https://doi.org/10.1038/s41558-022-01430-5>.
- Wohlers, Julia, Anja Engel, Eckart Zöllner, Petra Breithaupt, Klaus Jürgens, Hans-Georg Hoppe, Ulrich Sommer, and Ulf Riebesell. 2009. 'Changes in Biogenic Carbon Flow in Response to Sea Surface Warming'. *Proceedings of the National Academy of Sciences* 106 (17): 7067–72. <https://doi.org/10.1073/pnas.0812743106>.
- Worden, Alexandra Z., and Andrew E. Allen. 2010. 'The Voyage of the Microbial Eukaryote'. *Current Opinion in Microbiology, Antimicrobials/Genomics*, 13 (5): 652–60. <https://doi.org/10.1016/j.mib.2010.08.001>.
- Worden, Alexandra Z., Michael J. Follows, Stephen J. Giovannoni, Susanne Wilken, Amy E. Zimmerman, and Patrick J. Keeling. 2015. 'Rethinking the Marine Carbon Cycle: Factoring in the Multifarious Lifestyles of Microbes'. *Science* 347 (6223): 1257594. <https://doi.org/10.1126/science.1257594>.
- Worden, Alexandra Z., Jan Janouskovec, Darcy McRose, Augustin Engman, Rory M. Welsh, Stephanie Malfatti, Susannah G. Tringe, and Patrick J. Keeling. 2012. 'Global Distribution of a Wild Alga Revealed by Targeted Metagenomics'. *Current Biology* 22 (17): R675–77. <https://doi.org/10.1016/j.cub.2012.07.054>.
- Worden, Alexandra Z., and Fabrice Not. 2008. 'Ecology and Diversity of Picoeukaryotes'. In *Microbial Ecology of the Oceans*, 159–205. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470281840.ch6>.
- Wurch, Louie L., Christopher J. Gobler, and Sonya T. Dyhrman. 2014. 'Expression of a Xanthine Permease and Phosphate Transporter in Cultures and Field Populations of the Harmful Alga *Aureococcus Anophagefferens*: Tracking Nutritional Deficiency during Brown Tides'. *Environmental Microbiology* 16 (8): 2444–57. <https://doi.org/10.1111/1462-2920.12374>.
- Wurch, Louie L., Sheean T. Haley, Elizabeth D. Orchard, Christopher J. Gobler, and Sonya T. Dyhrman. 2011. 'Nutrient-Regulated Transcriptional Responses in the Brown Tide-Forming Alga *Aureococcus Anophagefferens*'. *Environmental Microbiology* 13 (2): 468–81. <https://doi.org/10.1111/j.1462-2920.2010.02351.x>.
- Wynne, Michael J., Robert A. Andersen, Louis Graf, and Hwan Su Yoon. 2014. 'Aureoscheda, a New Genus of Marine Pelagophyceae from the Bahamas, Caribbean Sea'. *Phycologia* 53 (5): 513–22. <https://doi.org/10.2216/14-055.1>.
- Yooseph, Shibu, Granger Sutton, Douglas B. Rusch, Aaron L. Halpern, Shannon J. Williamson, Karin Remington, Jonathan A. Eisen, et al. 2007. 'The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families'. *PLOS Biology* 5 (3): e16. <https://doi.org/10.1371/journal.pbio.0050016>.
- Zehr, Jonathan P. 2011. 'Nitrogen Fixation by Marine Cyanobacteria'. *Trends in Microbiology* 19 (4): 162–73. <https://doi.org/10.1016/j.tim.2010.12.004>.
- Zehr, Jonathan P., and Raphael M. Kudela. 2011. 'Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems'. *Annual Review of Marine Science* 3 (1): 197–225. <https://doi.org/10.1146/annurev-marine-120709-142819>.



# Annexes

---

## 1 - Annexes du chapitre 1

### Supplementary Tables

Supplementary tables are available : <https://doi.org/10.1038/s42003-022-03939-z>

### Supplementary Figures

See following pages.



## Supplementary information

### Genomic adaptation of the picoeukaryote *Pelagomonas calceolata* to iron-poor oceans revealed by a chromosome-scale genome sequence.

Nina Guérin<sup>1,2</sup>, Marta Ciccarella<sup>1</sup>, Elisa Flamant<sup>1,2</sup>, Paul Frémont<sup>1,2</sup>, Sophie Mangenot<sup>1,2</sup>, Benjamin Istace<sup>1</sup>, Benjamin Noel<sup>1</sup>, Caroline Belser<sup>1</sup>, Laurie Bertrand<sup>1,2</sup>, Karine Labadie<sup>2,3</sup>, Corinne Cruaud<sup>2,3</sup>, Sarah Romac<sup>4</sup>, Charles Bachy<sup>4,5</sup>, Martin Gachenot<sup>5</sup>, Eric Pelletier<sup>1,2</sup>, Adriana Alberti<sup>1,2,6</sup>, Olivier Jaillon<sup>1,2</sup>, Patrick Wincker<sup>1,2</sup>, Jean-Marc Aury<sup>1</sup>, Quentin Carradec<sup>1,2\*</sup>

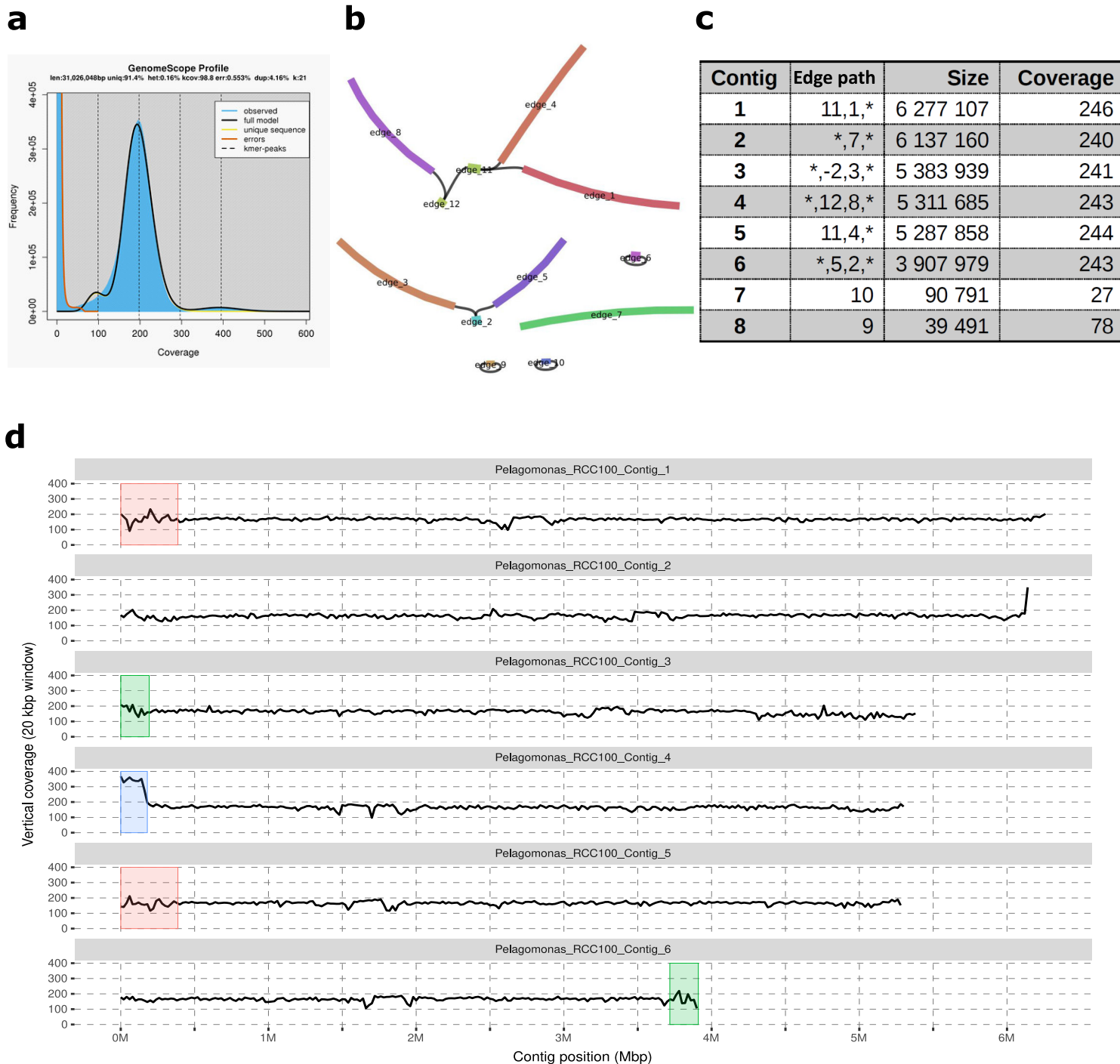
\*Correspondence and requests for materials should be addressed to [qcarrade@genoscope.cns.fr](mailto:qcarrade@genoscope.cns.fr)

#### Supplementary Figures

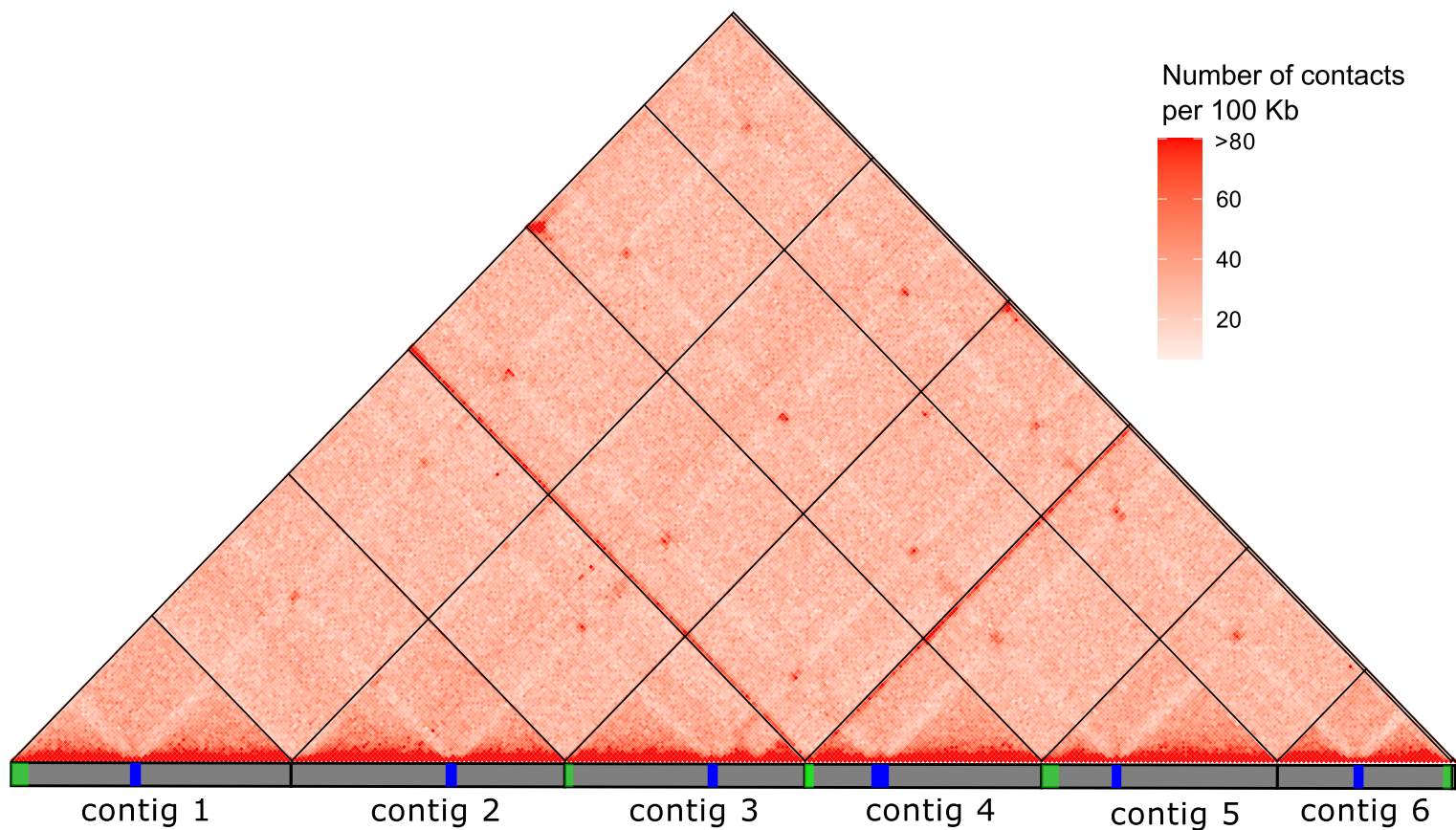
Figure S1: Assembly and duplicated regions of the <i>P. calceolata</i> genome.	2
Figure S2: Hi-C interaction map for <i>P. calceolata</i> genome.	3
Figure S3: Identification of introner elements from repeat families.	4
Figure S4: <i>P. calceolata</i> proteins shared with other stramenopiles.	5
Figure S5: Estimation of the relative abundance of <i>P. calceolata</i> with 2 different methods.	6
Figure S6: Relative abundance and ecological niche of <i>P. calceolata</i> in the 0.8-2000 $\mu\text{m}$ size-fraction.	7
Figure S7: Modelled genomic relative abundance of <i>P. calceolata</i> in present day and at the end of the century.	8
Figure S8: Expression levels of genes involved in iron transport and coding for superoxide dismutases.	9
Figure S9: Expression levels of genes involved in nitrogen compounds uptake.	10
Figure S10: Expression levels of genes involved in nitrogen metabolism in <i>P. calceolata</i> .	11
Figure S11: Nitrate and nitrite sensing proteins in <i>P. calceolata</i> and other eukaryotes.	12

#### Supplementary Notes

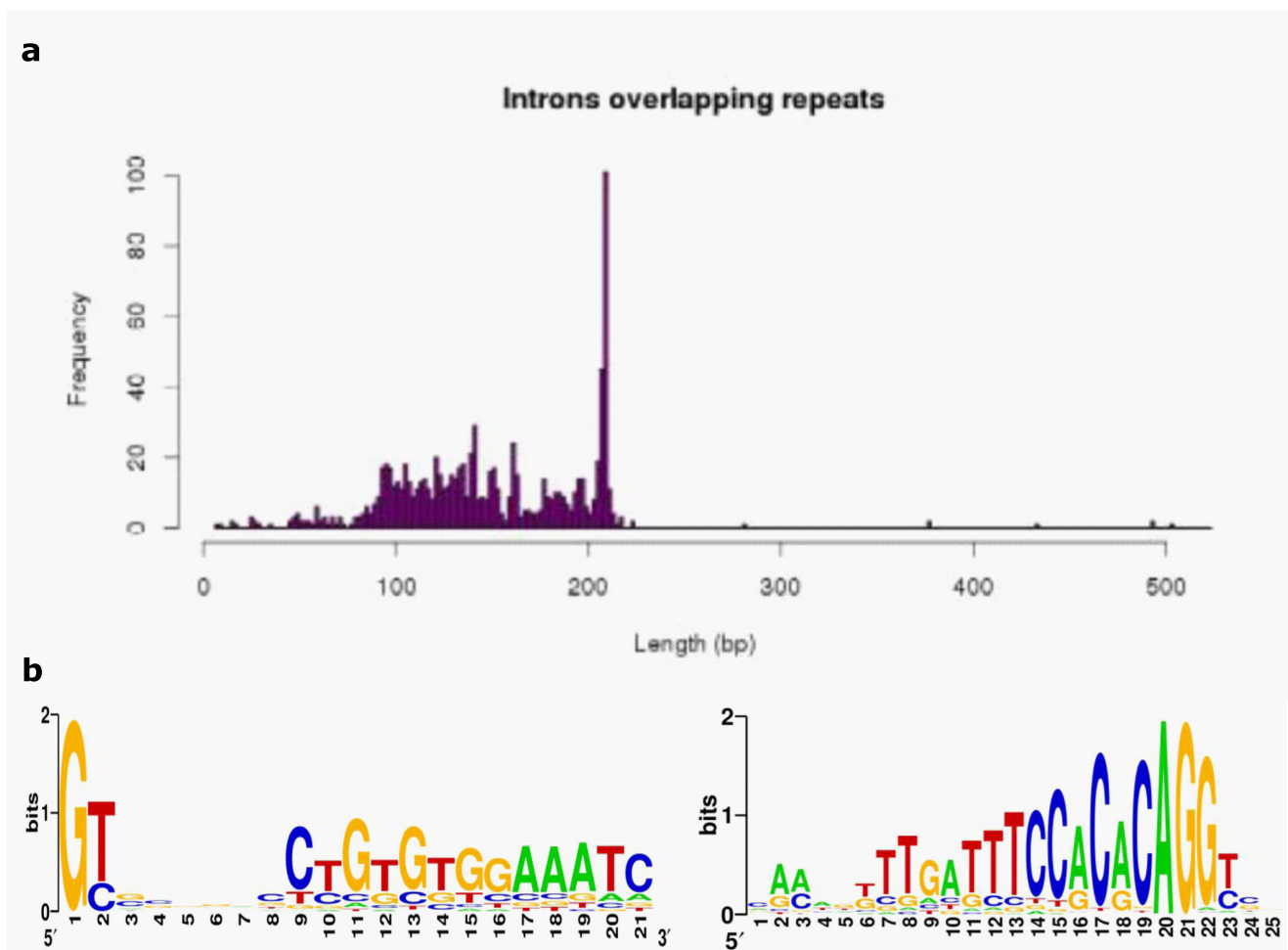
Supplementary Note 1: Duplicated regions in <i>P. calceolata</i> genome	13
Supplementary Note 2: Description of the Hi-C result	13
Supplementary Note 3: Introns and Introner Elements	13
Supplementary Note 4: GC content distribution along <i>P. calceolata</i> chromosomes	14
Supplementary Note 5: Centromeres, recombination and meiosis in <i>P. calceolata</i>	15
Supplementary References	16



**Figure S1: Assembly and duplicated regions of the *P. calceolata* genome.** a) K-mer profile of *P. calceolata* Illumina reads. The graph shows the fit of the GenomeScope model (black) to the observed k-mer frequencies (blue). The k-mers presenting high frequency and low coverage (red) are considered as sequencing errors. b) Graphical representation of the Flye assembly graph generated with Bandage tool (Wick et al. 2015). Each colored box represents a sequence (edge) of the assembly. Edges connected with one or several black lines indicate unresolved repeats at the extremity of the contig (e.g. edge<sub>2</sub> is connected to edge<sub>3</sub> and/or edge<sub>5</sub>). Based on the vertical coverage, Flye chose to duplicate edge<sub>2</sub> to form contig 3 and contig 6). c) Table of the assembly paths and metrics of the assembled contigs detected by Flye. Numbers in the edge path column correspond to the edges in panel A. Telomeric repeats located at the ending edges of the assembly graph are reported as a star in the path column. d) Vertical coverage of Illumina reads in the *P. calceolata* genome. The dark line represents the number of Illumina short reads mapped over a window of 20 Kb on the 6 nuclear contigs of *P. calceolata*. Red and green boxes indicate highly similar regions (>99% of identity). The instability of the read coverage in these regions may be explained by the presence of variants in one of the two copies. The blue box is the only large region with a vertical coverage twice as high as the genome average.



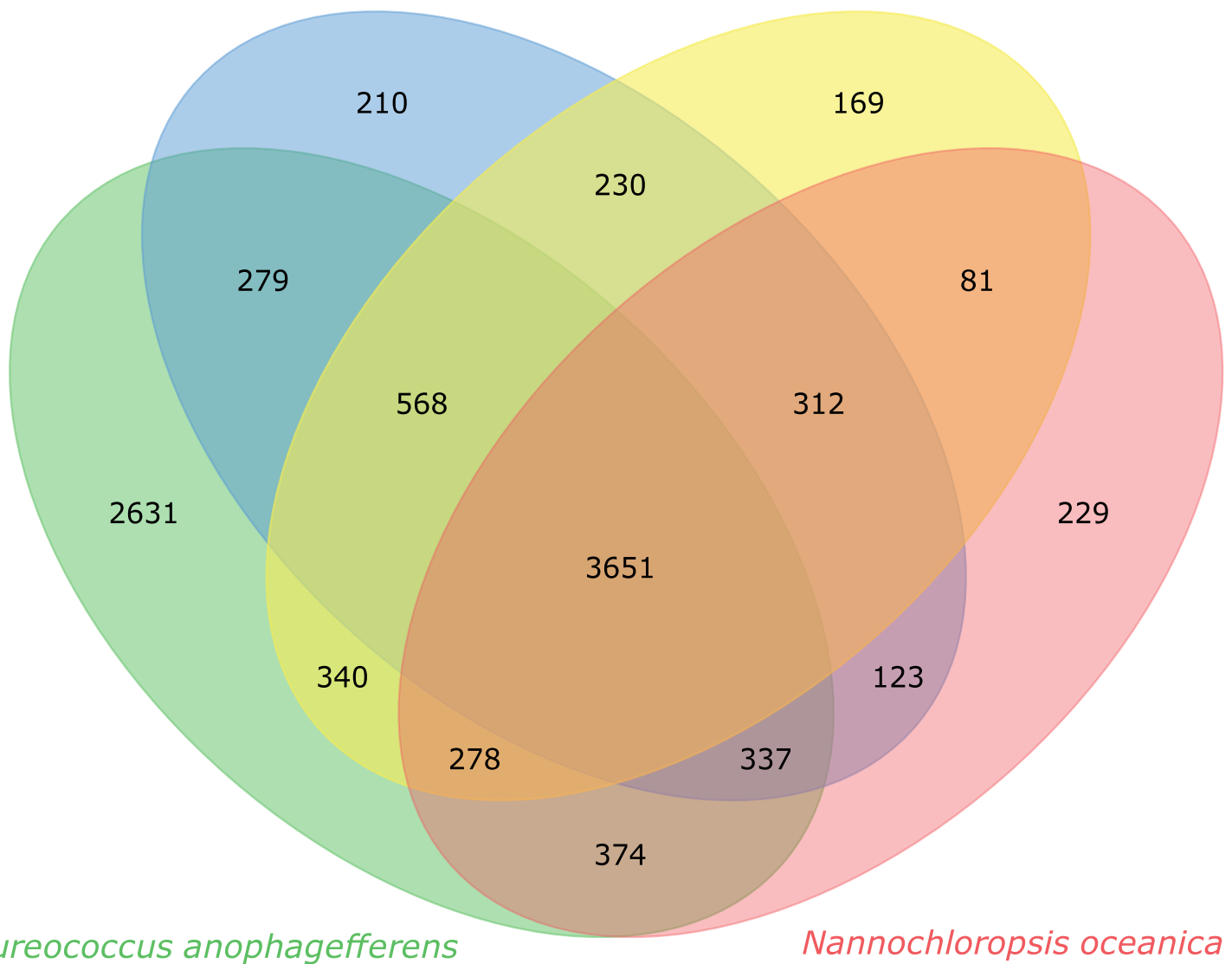
**Figure S2: Hi-C interaction map for *P. calceolata* genome.** The localisation of interaction points in the *P. calceolata* genome sequence are indicated by red dots (window of 100Kb). The red intensity is proportional to the number of reads covering this interaction. The low-GC region in each contig is indicated in blue. Red dots corresponding to these regions indicate physical proximity between centromeres. Duplicated regions are indicated in green. Red dots corresponding to these regions are due to the random alignment of reads on the duplicated regions. The interaction background is more intense at the beginning of contig 4 because this duplicated region was collapsed in the assembly resulting in a double read coverage.



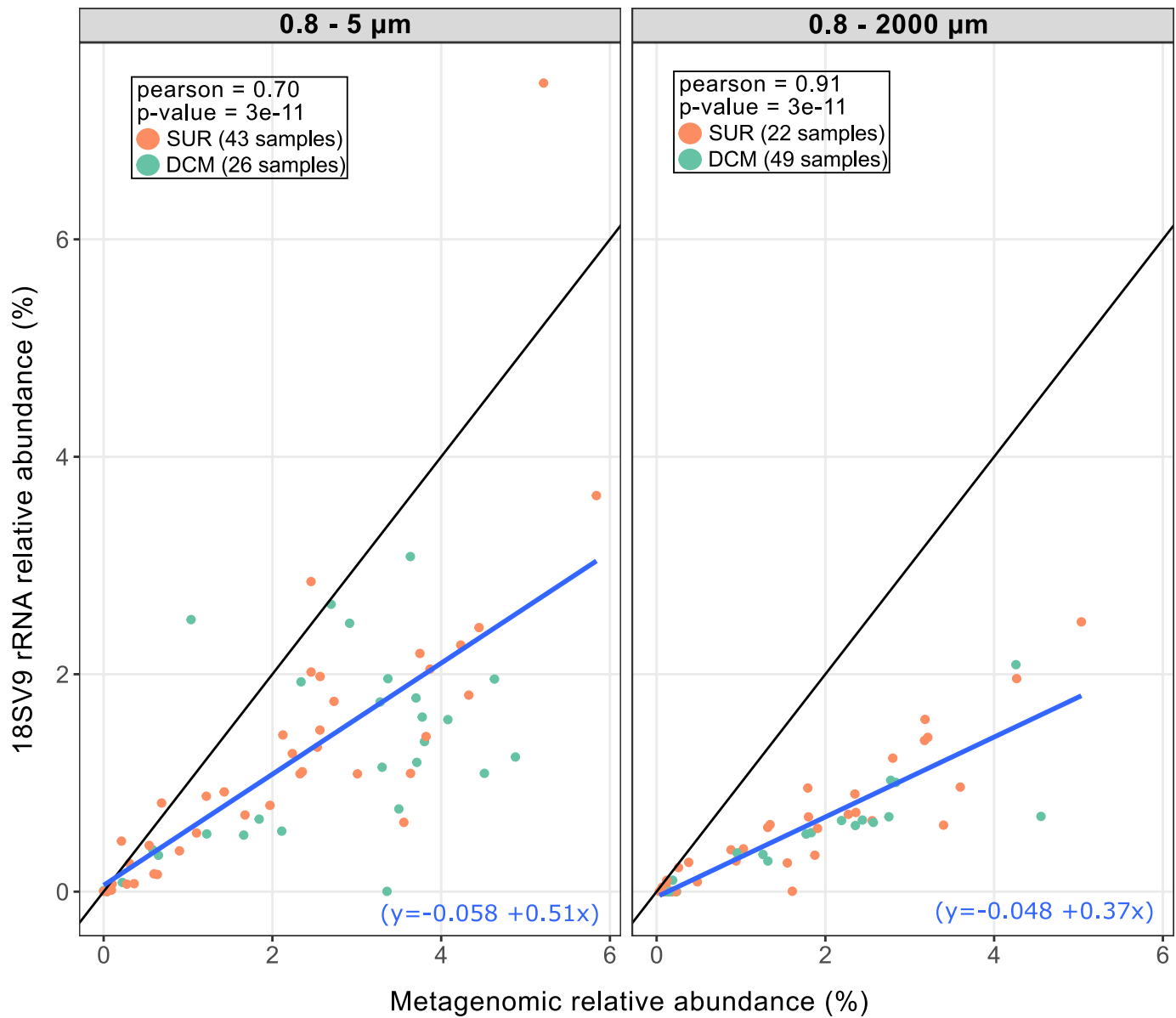
**Figure S3: Identification of Introner Elements from repeat families.** a) Distribution of putative IE lengths shows a pick around 200 bp. b) Logo representations of the starting and ending sequences of putative IE, revealing the presence of GT or GC donor splicing sites at the 5' ending, AG acceptor splicing site at the 3' ending, and conserved Terminal Inverted Repeats at the flanking regions.

*Phaeodactylum tricornutum*

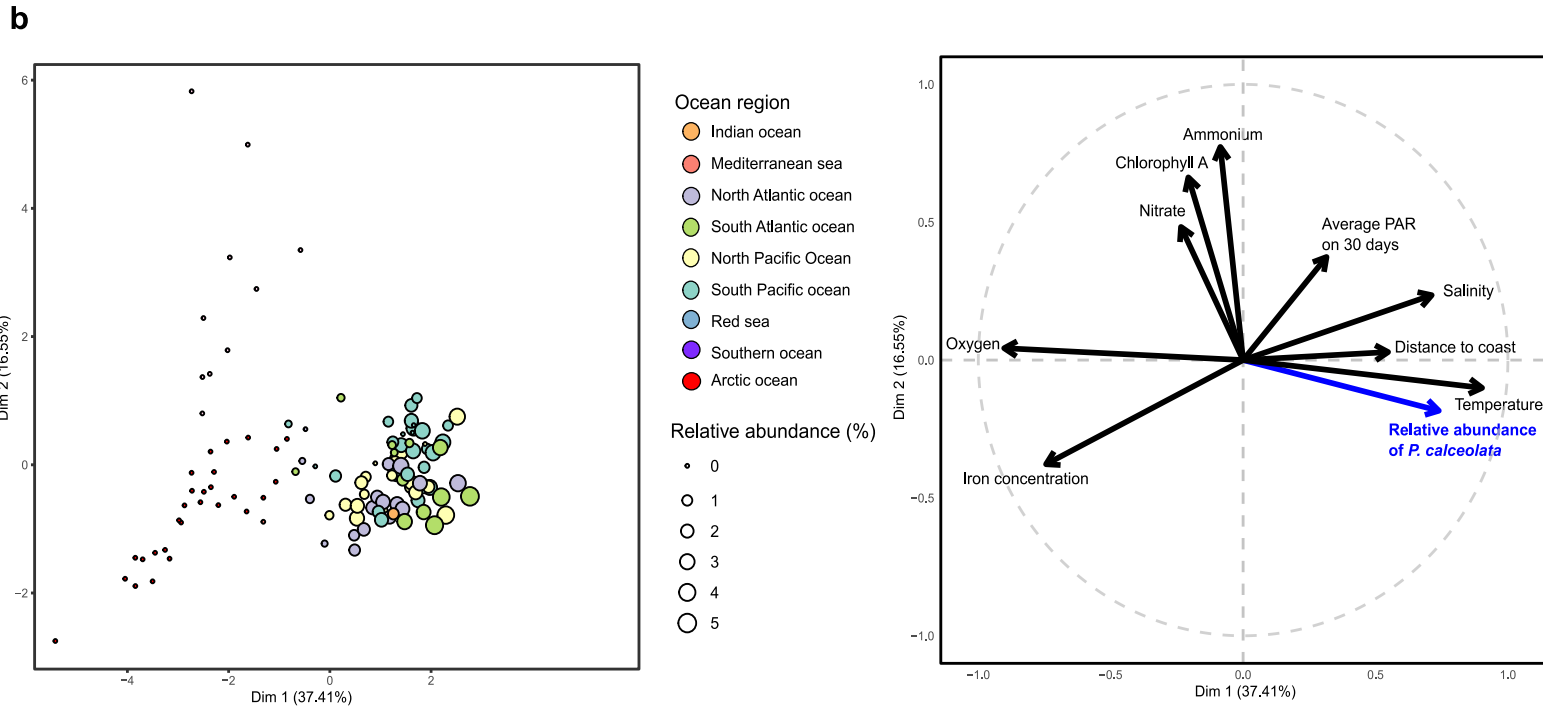
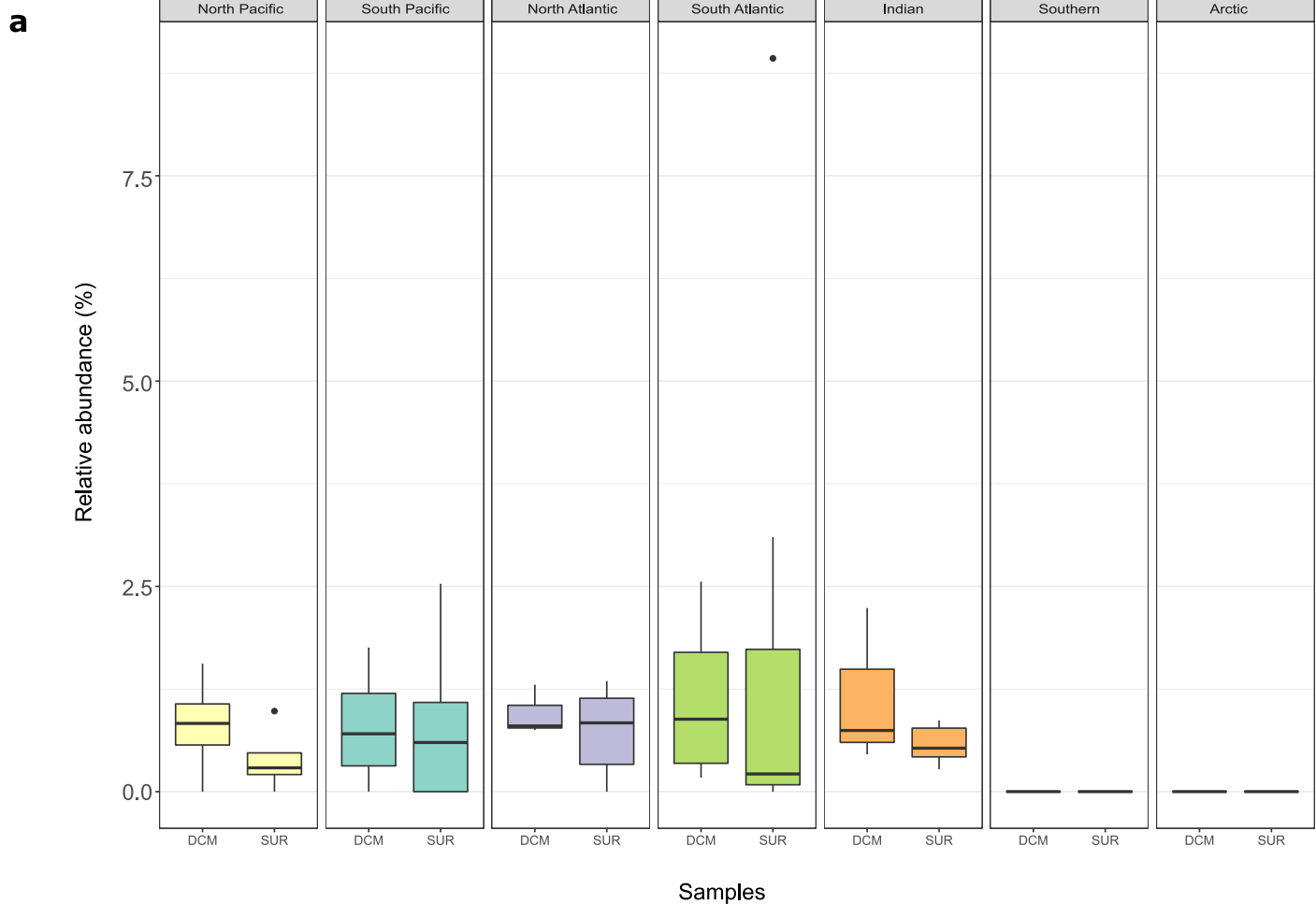
*Thalassiosira pseudonana*



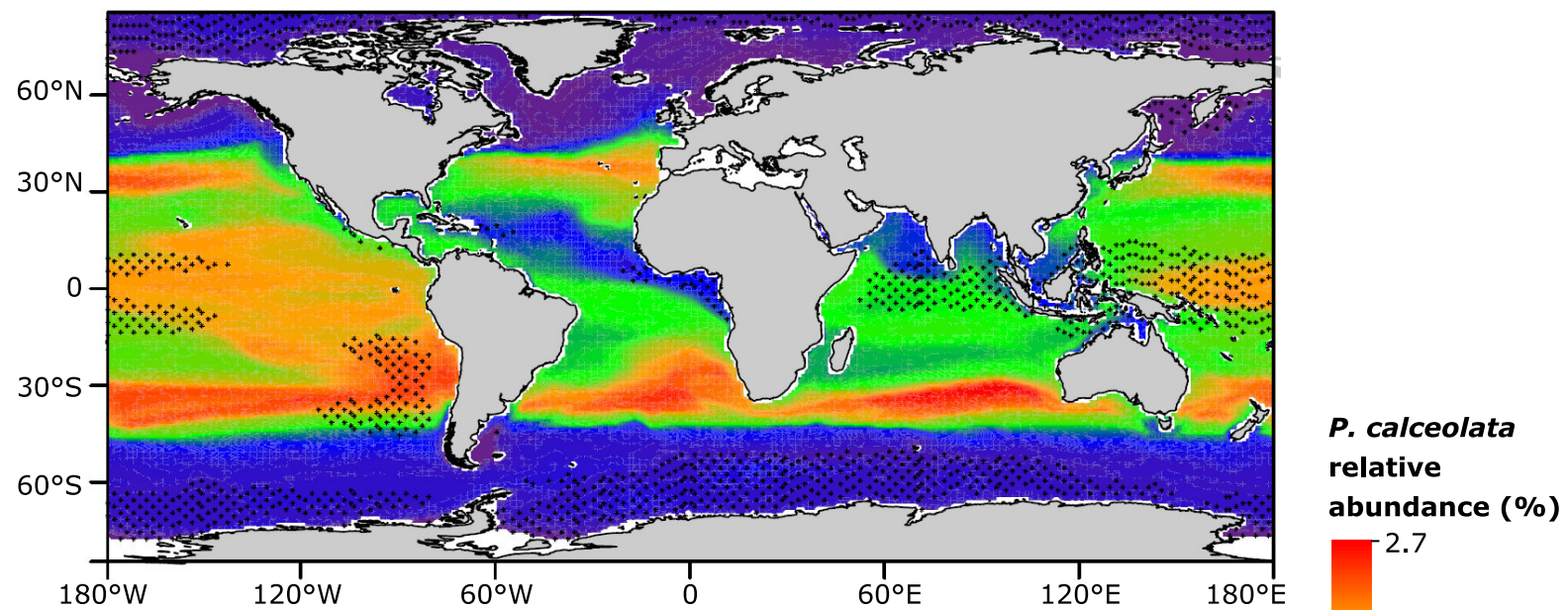
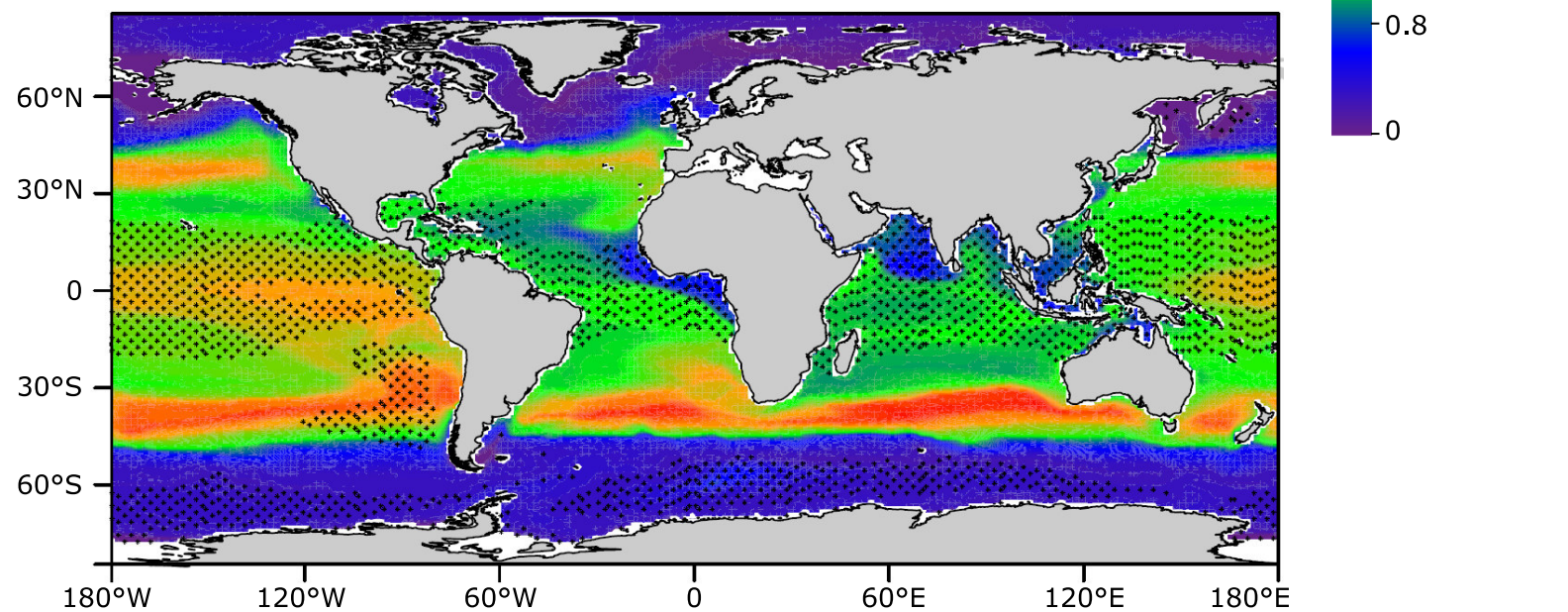
**Figure S4: *P. calceolata* proteins shared with other Stramenopiles.** The venn diagram displays the number of *P. calceolata* proteins homologous with at least one protein in another Stramenopile genome (alignment length >80% of the shortest protein and evalue <math>10^{-5}</math>).



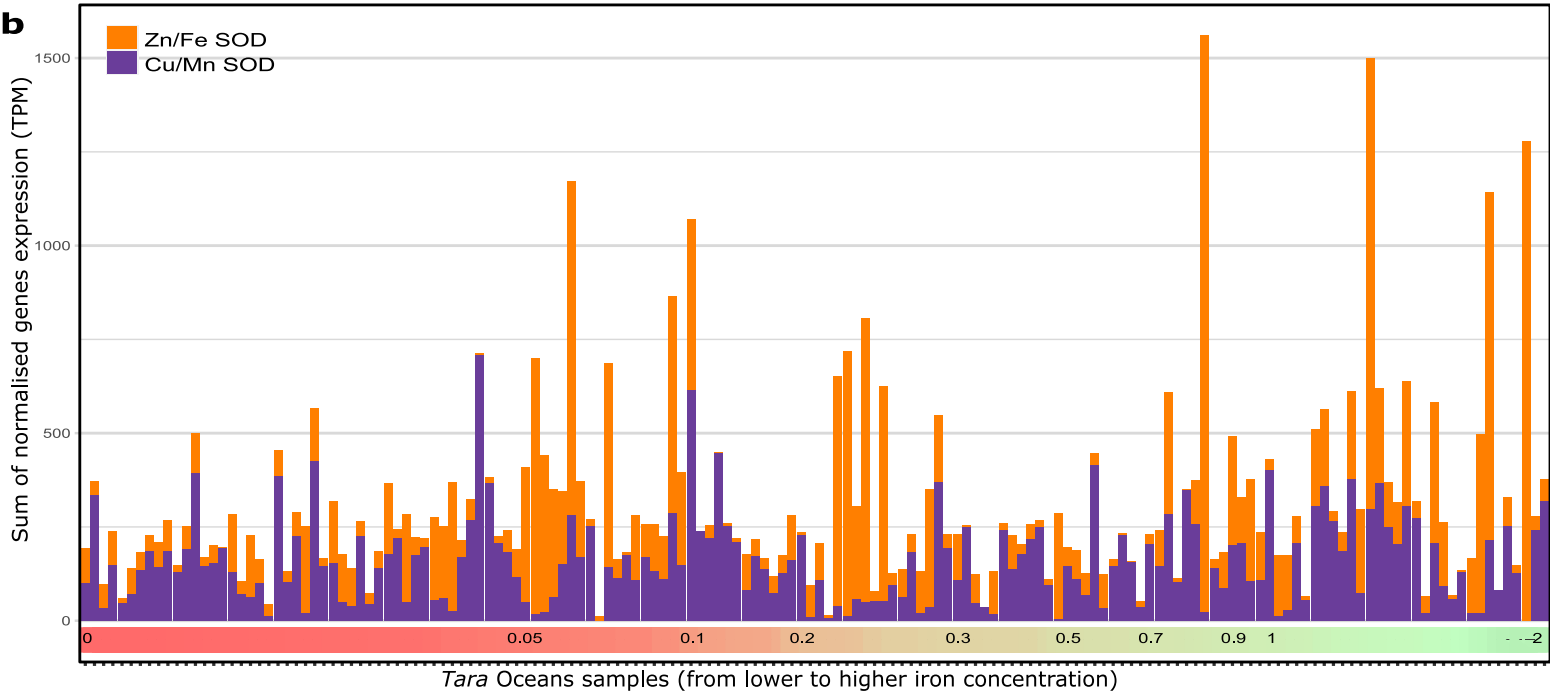
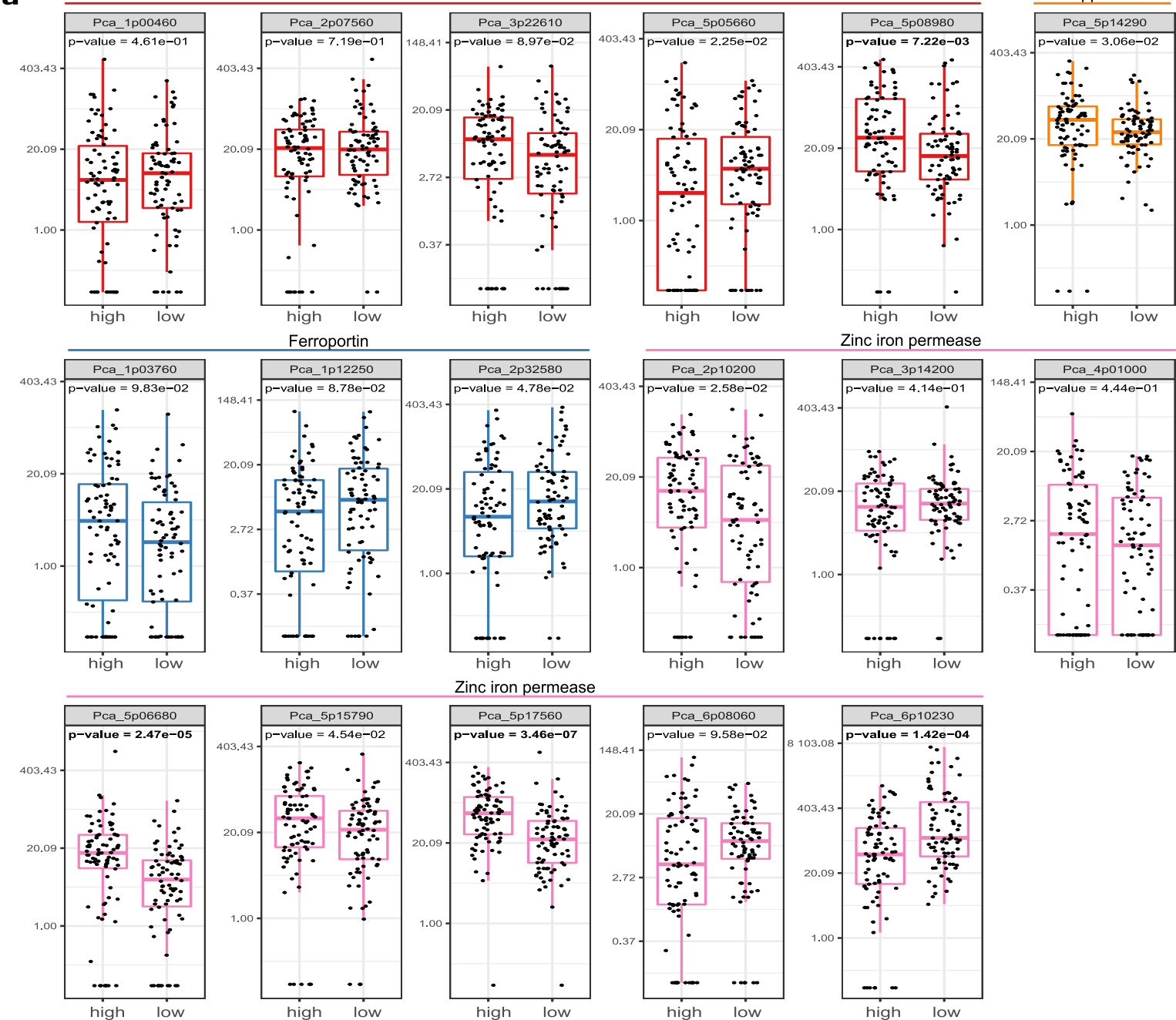
**Figure S5: Relative abundance of *P. calceolata* with 2 different methods.** Each dot represents a surface (orange) or DCM (green) sample. The relative abundance of *P. calceolata* is calculated from metagenomic reads aligned on the genome (x-axis) and with the amplified 18S rRNA sequence (y-axis) for size-fractions 0.8 - 5 μm (left) and 0.8 - 2000 μm (right). Linear regressions of observed values and the equation are indicated in blue. Black lines are expected trends.



**Figure S6: Relative abundance and ecological niche of *P. calceolata* in the 0.8-2000  $\mu\text{m}$  size fraction.** a) Boxplot of the relative abundance of *P. calceolata* in each oceanic region in surface and DCM samples in the 0.8 - 2000  $\mu\text{m}$  size-fraction. b) Principal component analysis of the relative abundance of *P. calceolata* and the 9 environmental parameters in the 0.8-2000  $\mu\text{m}$  size fraction. Each dot represents a sample, with a size proportional to the relative abundance of *P. calceolata*. The colours indicate the oceanic basins.

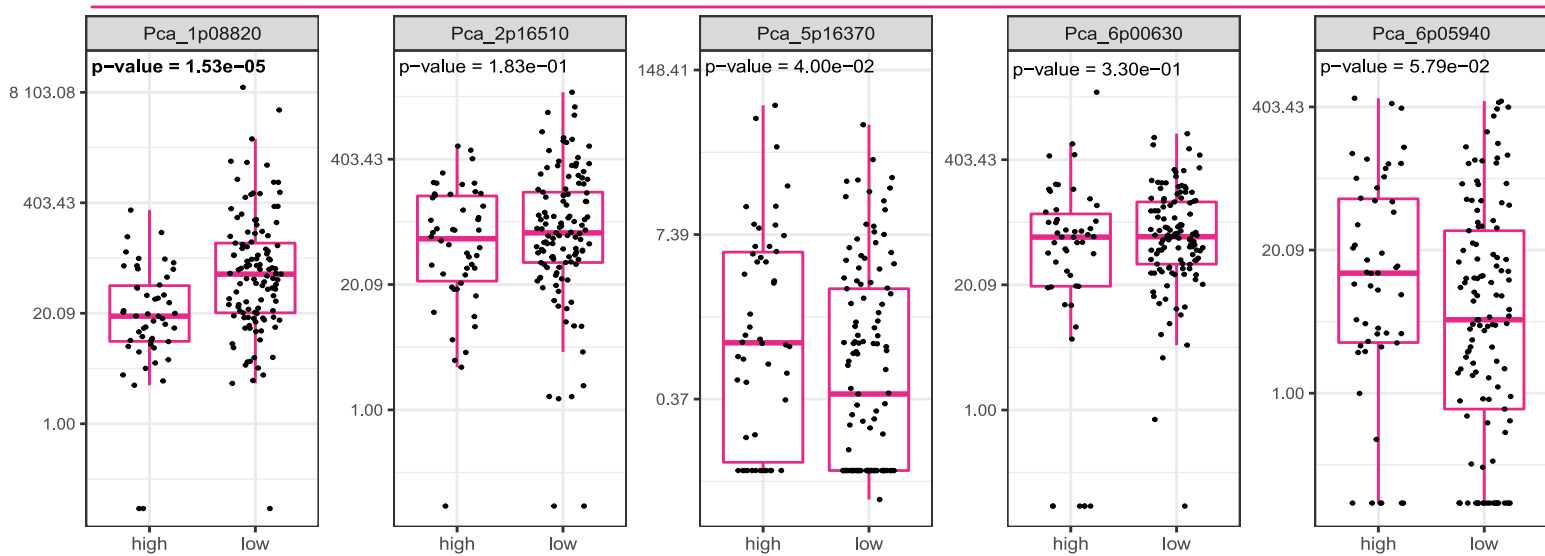
**a****b**

**Figure S7: Modelled relative abundance of *P. calceolata* in present day (a) and at the end of the century (b).** The relative abundance of *P. calceolata* measured in *Tara* samples (0.8-2000  $\mu\text{m}$  size-fraction) is modeled using the combination of 4 machine learning techniques trained on WOA18 environmental parameters at months, depths and sampling locations of *Tara* Oceans samples (a) or trained on projected climatology in 2099 under the RCP8.5 scenario (b).

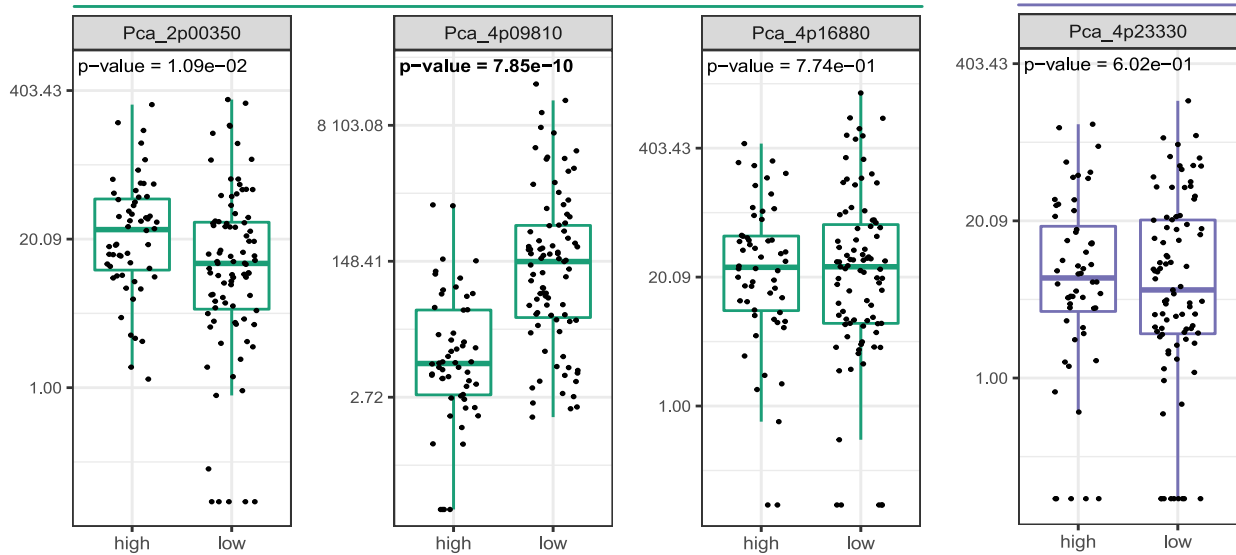


**Figure S8: Expression levels of genes involved in iron transport and coding for superoxide dismutases (SOD).** a) Relative expression levels (TPM) of genes involved in iron uptake compared between low-iron (<0.2 nM) and high-iron (>0.2 nM) environment. P-values of Wilcoxon statistical tests between low- and high-iron conditions are indicated for each gene. Significant p-values (<0.01) are in bold. b) Relative expression levels (TPM) of genes coding for zinc/iron SOD (orange) and its non-ferrous equivalent copper/manganese SOD (purple) in each *Tara Oceans* sample. Samples are sorted from low-iron (left) to high-iron (right) conditions. Iron concentrations are indicated in nM on the horizontal coloured gradient.

Ammonium transporter

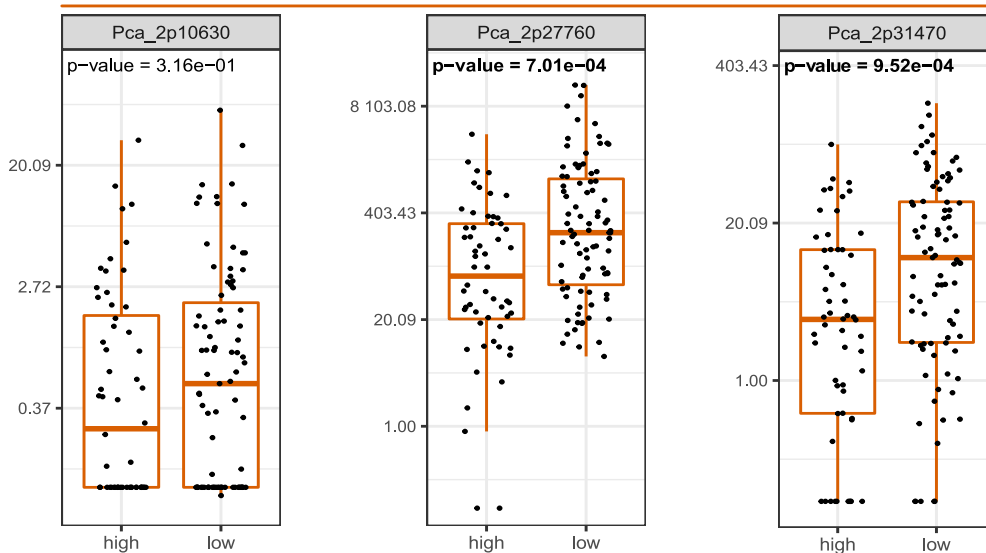


Formate/nitrite transporter



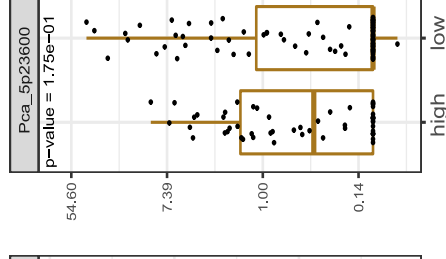
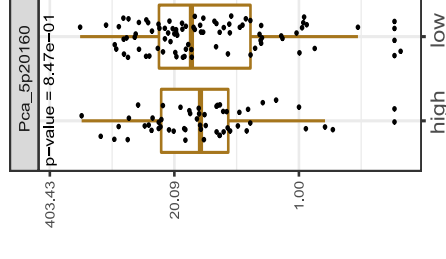
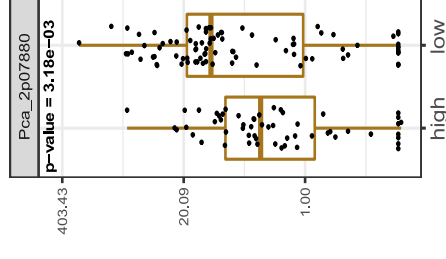
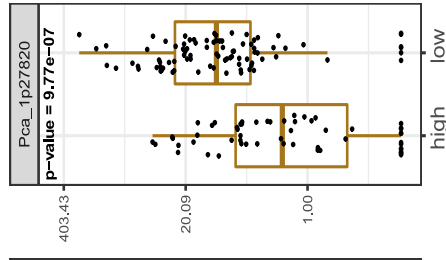
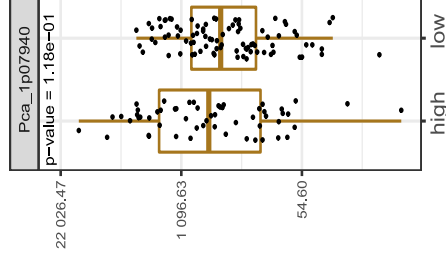
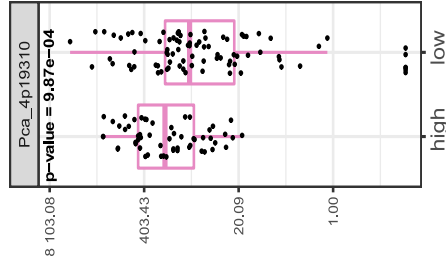
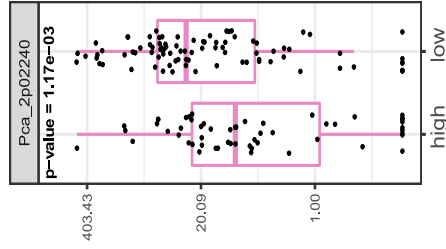
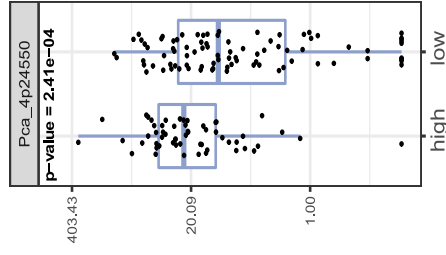
Urea transporter

Nitrate/nitrite transporter

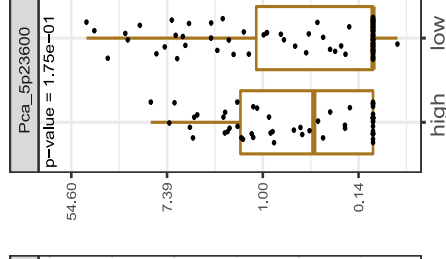
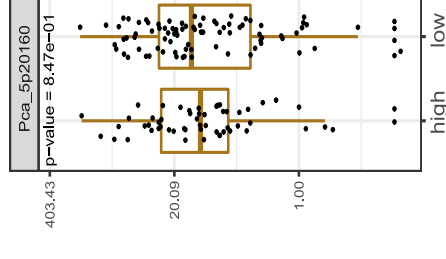
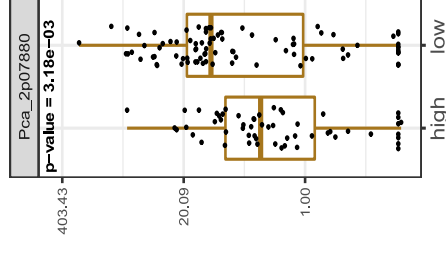
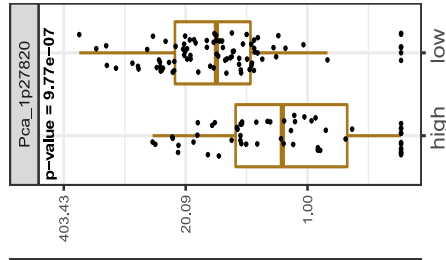
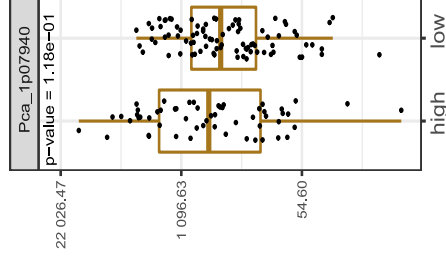
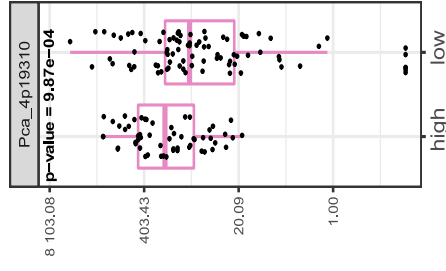
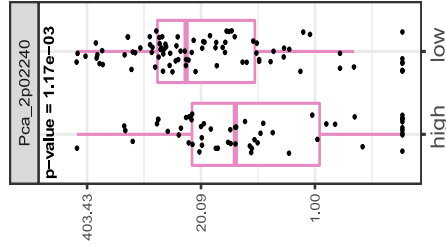
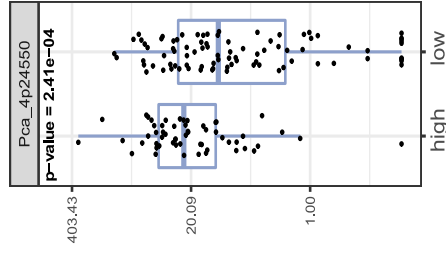


**Figure S9: Expression levels of genes involved in nitrogen compounds transport.** Relative expression levels (TPM) of genes involved in ammonium uptake compared between low-ammonium (<0.025  $\mu\text{M}$ ) and high-ammonium (>0.025  $\mu\text{M}$ ) environments, and of genes involved in nitrate, nitrite, formate and urea uptake compared between low-nitrate (<2  $\mu\text{M}$ ) and high-nitrate (>2  $\mu\text{M}$ ) environments, in the 0.8-5  $\mu\text{m}$  and 0.8-2000  $\mu\text{m}$  size fractions. P-values of Wilcoxon statistical tests between low- and high-ammonium/nitrate conditions are indicated for each gene. Significant p-values (<0.01) are in bold.

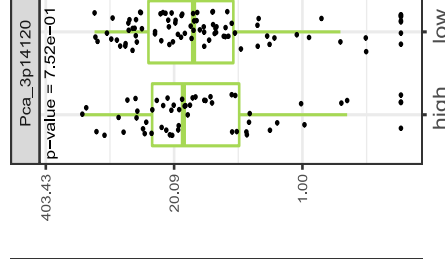
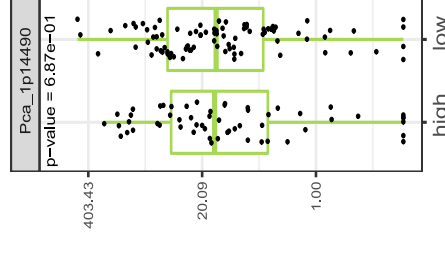
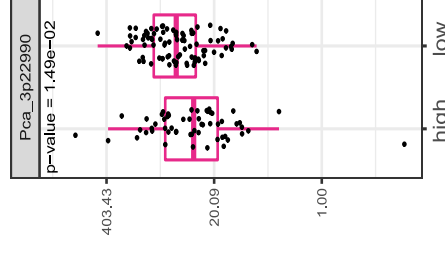
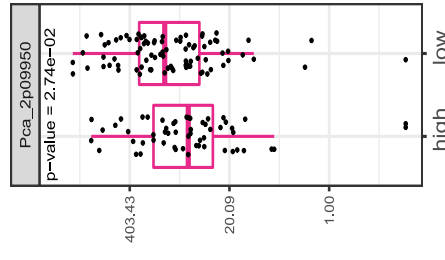
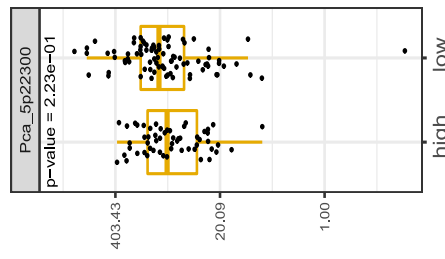
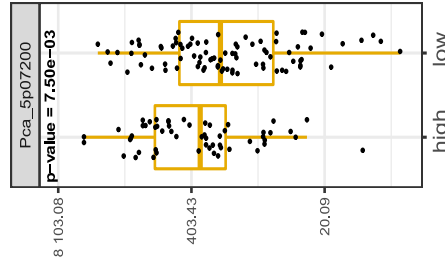
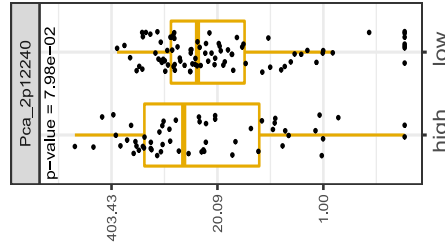
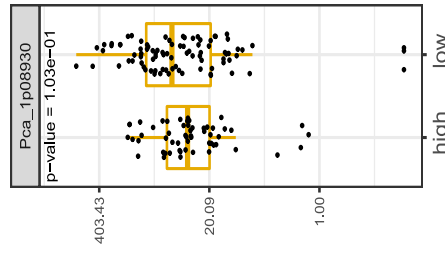
## Nitrate reductase



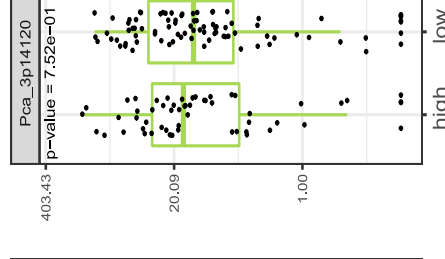
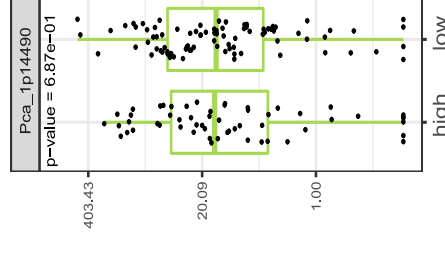
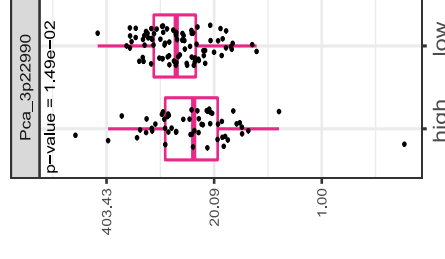
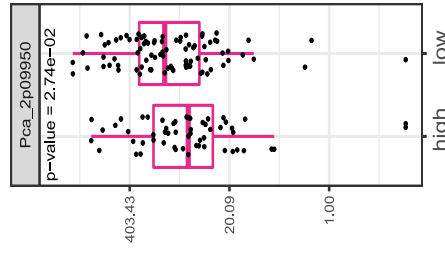
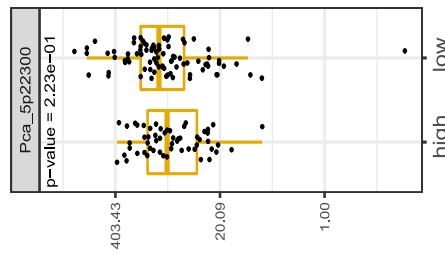
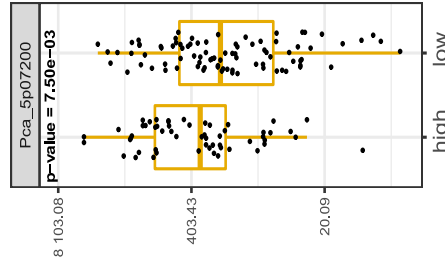
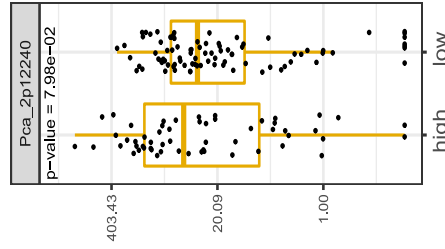
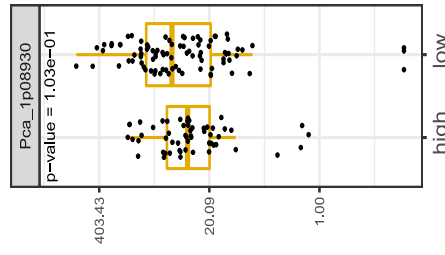
## Glutamine synthetase (GS)



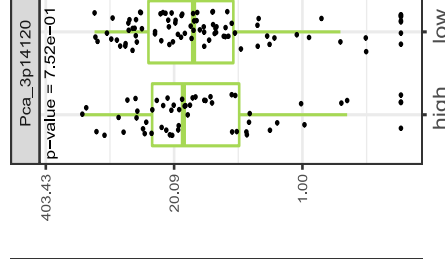
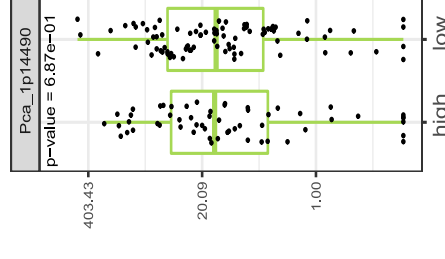
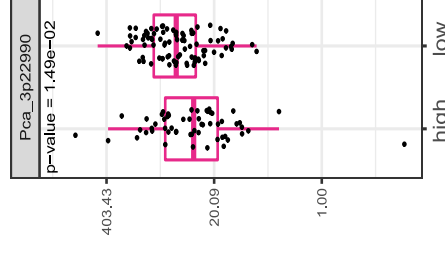
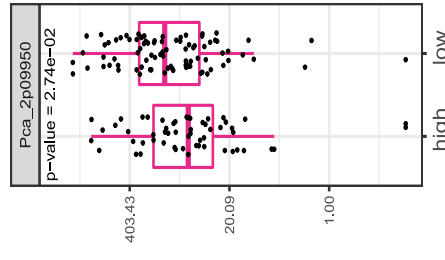
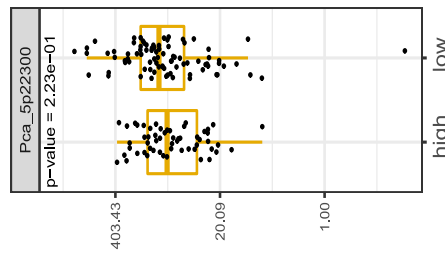
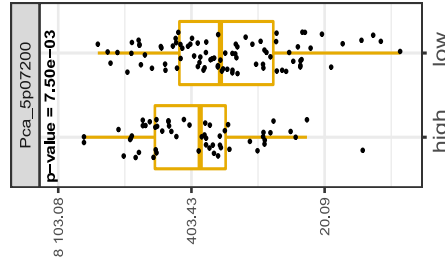
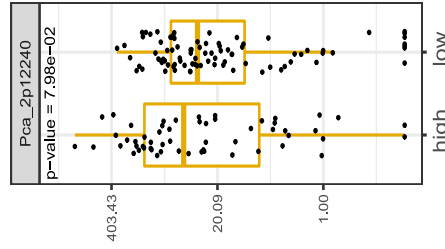
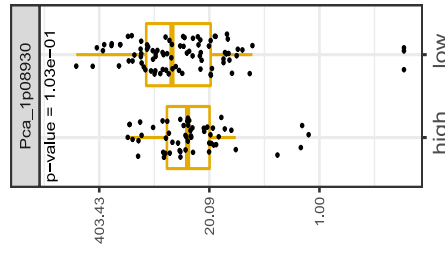
## Glutamate synthase (GOGAT)



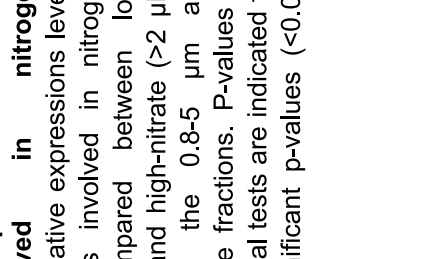
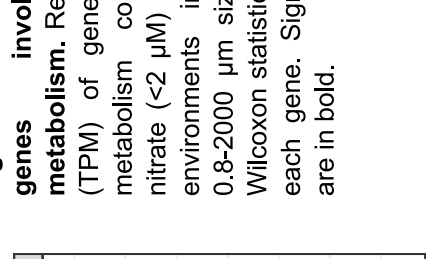
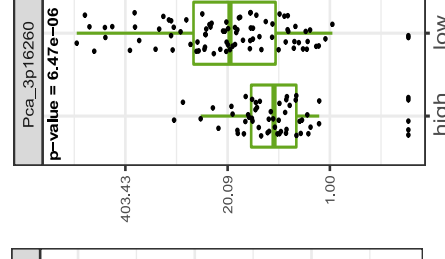
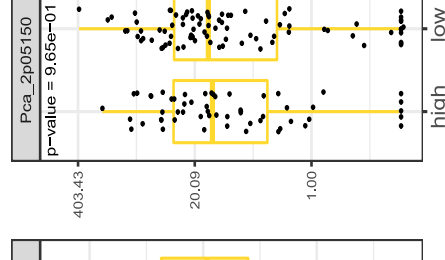
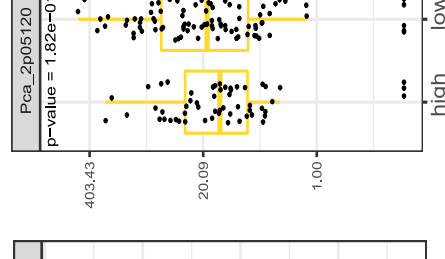
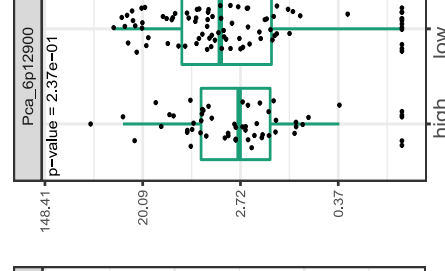
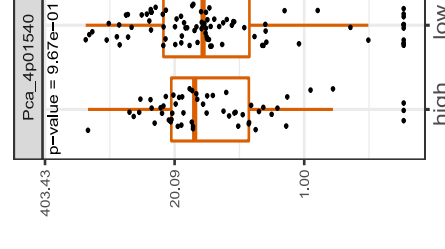
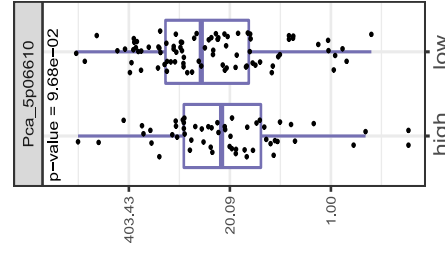
## Carbomoyl Phosphate Synthetase



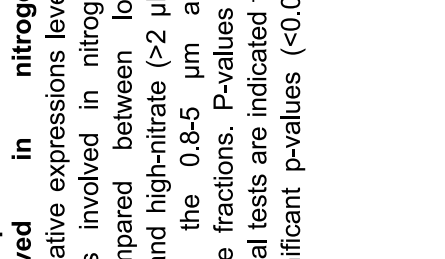
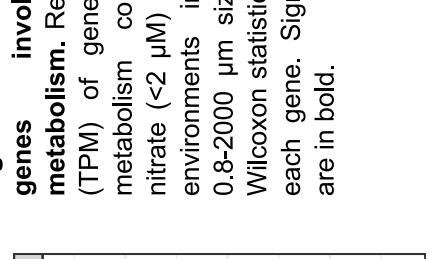
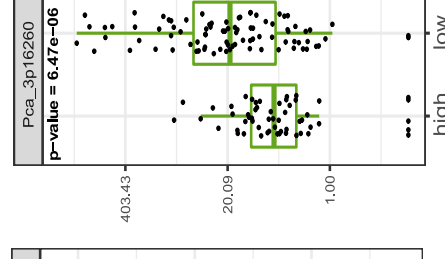
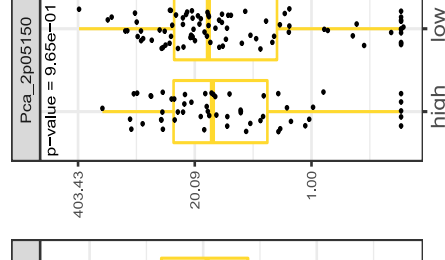
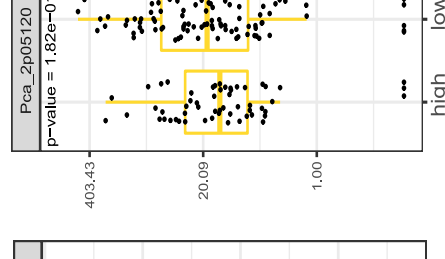
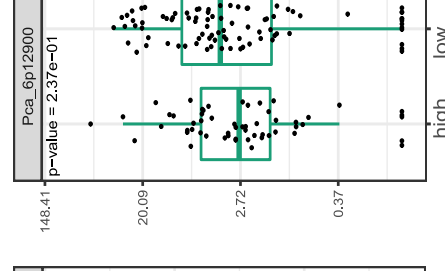
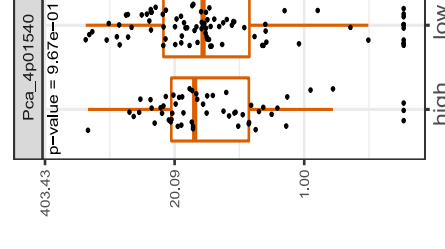
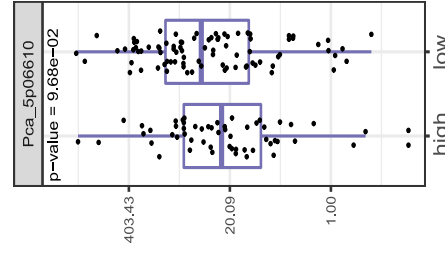
## Ornithine carbamoyltransferase



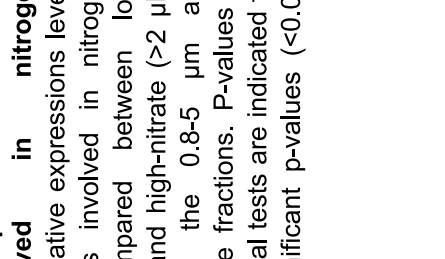
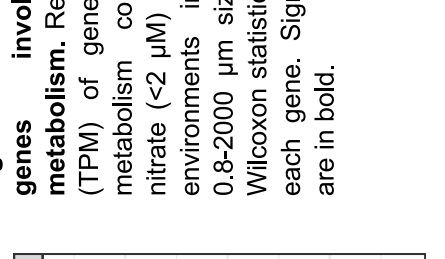
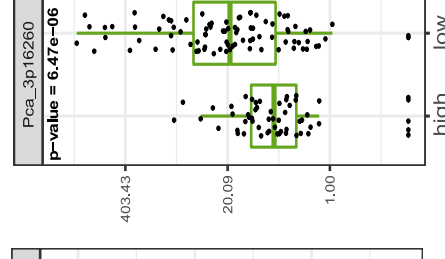
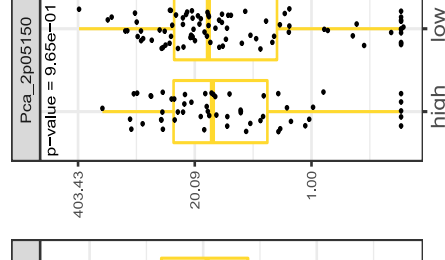
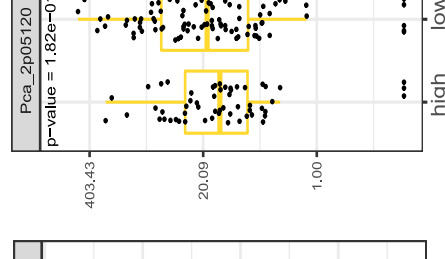
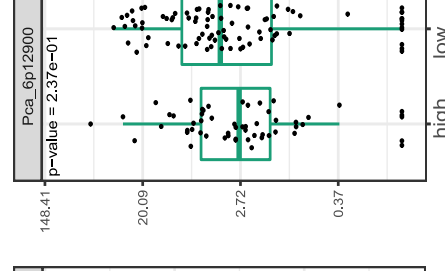
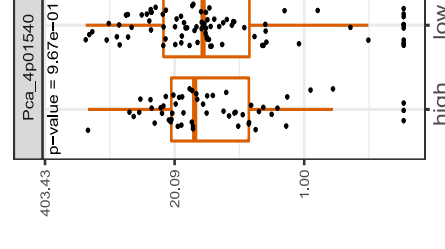
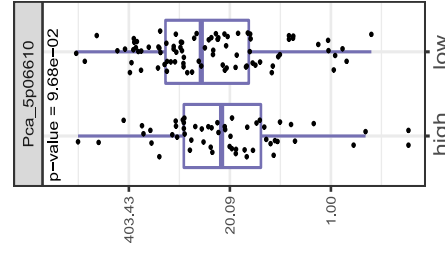
## Argininosuccinate synthase



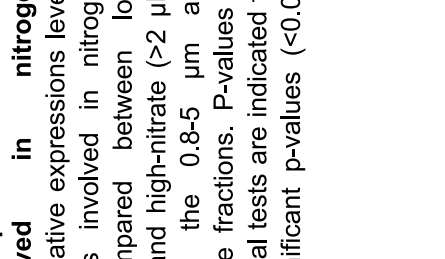
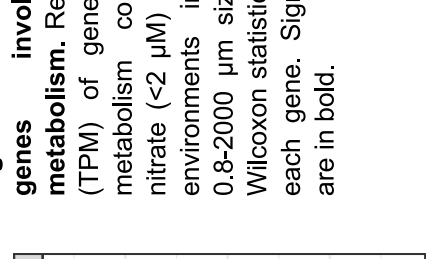
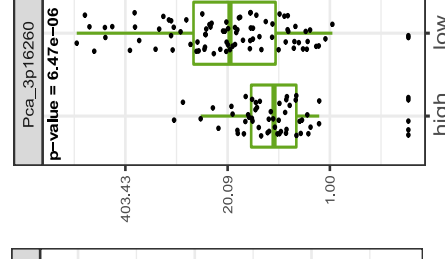
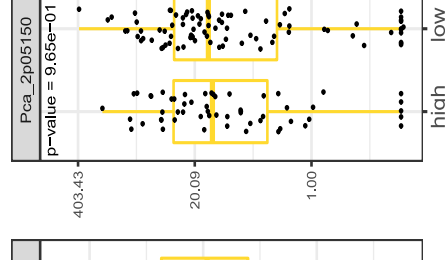
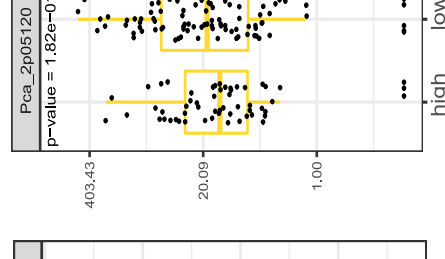
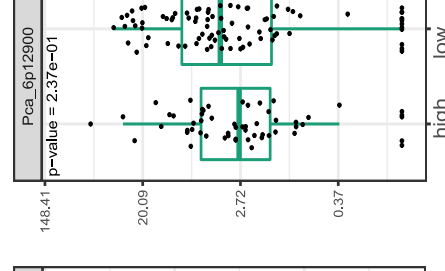
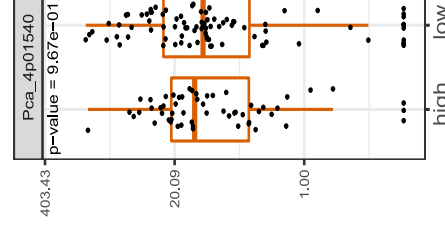
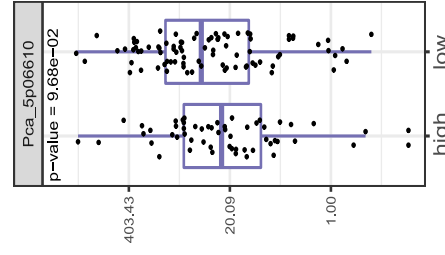
## Argininosuccinate lyase



## Urease



## Cyanate lyase



**Figure S10: Expression levels of genes involved in nitrogen metabolism.** Relative expressions levels (TPM) of genes involved in nitrogen metabolism compared between low-nitrate (<2  $\mu$ M) and high-nitrate (>2  $\mu$ M) environments in the 0.8-5  $\mu$ m and 0.8-2000  $\mu$ m size fractions. P-values of Wilcoxon statistical tests are indicated for each gene. Significant p-values (<0.01) are in bold.



### Supplementary Note 1: Duplicated regions in *P. calceolata* genome

We observed two highly similar duplicated regions (>99% identity) in the *Pelagomonas calceolata* genome. We measured the vertical coverage along the genome to understand if these duplications are real or assembly artefacts. The vertical read coverage of duplicated regions is similar to that of other genomic regions suggesting that they are both present in all cells of this *P. calceolata* culture (Figure 1 and Figure S1d). In addition, 150 Kb at one extremity of contig 4 present a higher vertical coverage suggesting that this region is also duplicated in the *P. calceolata* genome but collapsed in the assembly. Interestingly, duplicate regions at the end of sequences have already been observed in the chlorophyte *Ostreococcus tauri* which has been maintained for several years in culture<sup>1</sup>. These observations suggest that long-time cultures could affect not only the sequence of a given organism's genome but also its structure.

### Supplementary Note 2: Description of the Hi-C results

We aligned 75.9% of Hi-C raw reads on the *P. calceolata* genome sequence and nearly 92% of them were considered as Hi-C contacts, showing that the Hi-C preparation was successful. Specific features were observed on the contact map (Figure S2). The red diagonal indicates a large number of contacts within each chromosome arm. In contrast, the number of contacts between arms of the same chromosome is barely above the background. The second striking feature is the unique spot of contacts between all low-GC regions of each chromosome. These contacts confirm the position of the centromeres in each chromosome. Indeed, centromeres co-localised in the nucleus architecture<sup>2</sup> and the Hi-C method is able to reveal this proximity<sup>3,4</sup>. Interestingly, weak contacts were observed between all centromeres and the end of chromosome 3. This is potentially because this region also encompasses the repeats detected in the centromeres (Figure 1). In addition, two duplicated regions in the extremities of contigs 1, 3, 5 and 6 present a high number of contact. This is probably an artefact due to the random alignment of these reads in the duplicated regions and do not reflect a specific spatial proximity in the nucleus for these regions. Finally, the beginning of contig 4 present a higher contact background compare to other genomic regions. This is because this region is probably duplicated in the genome but merged in the assembly. Overall, the Hi-C analysis confirms the assembly of *P. calceolata* genome based on ONT long reads.

### Supplementary Note 3: Introns and introner elements

The presence of introner elements (IE) in *P. calceolata* was investigated by overlapping the position of the introns in transcriptome alignments with the positions of repeat families detected *ab initio* with RepeatScout<sup>5</sup>. Intron overlapping repeats over more than 90% of their length were identified as

putative IEs. The *P. calceolata* genome contains 956 putative IEs. The logo representation of the putative IE sequences reveals the presence of GT and GC donor sites at the 5' end, AG acceptor site at the 3' end, and conserved TIR in the flanking regions (Figure S3b and Table S3).

#### Supplementary Note 4: GC content distribution along *P. calceolata* chromosomes

The chromosome level assembly of the *P. calceolata* genome obtained with ONT and Hi-C sequencing revealed unique genomic features that are not usually studied in short-read assembled genomes. The most striking observation is the presence of a unique GC trough (52%) in each putative chromosome contrasting with the high GC content (63%) of other genomic regions of *P. calceolata* suggesting that these low-GC regions encompass centromeres. In eukaryotes, centromeres have a large variety of structures and characteristics. They are composed of many repeated sequences or contain genes, they are determined genetically or epigenetically and their size can vary from 125 bases (in *S. cerevisiae*) to several Mb (in mammals)<sup>6</sup>. Short regional centromeres (1-5 Kb) are generally low-GC compared to the genome and the sequence is unique at each centromere. Among stramenopiles, centromeres were characterized in the diatom *P. tricornutum* where a short low-GC sequence (>500 bp) is enough to define a centromere<sup>7</sup>. Large regional centromeres (>10 Kb) are generally gene-free, contain repeated elements and are not transcribed<sup>6</sup>. *P. calceolata* putative centromeres derive from this general pattern with large low-GC regions containing genes and few repeated elements. Interestingly, the red alga *Cyanidioschyzon* as well as several yeast species seem to have similar centromere structures with large, low-GC centromeres containing genes<sup>8,9</sup>.

The gene structure within low-GC regions differs from that of other genomic regions (Figure 1 and Table S5). The slight decrease of gene density observed in Figure 1 is explained by a longer gene size (average of 2.8 Kb compared to 1.7 Kb in other regions). In addition, the 453 genes in low-GC regions contain more introns (2.2 introns per gene compared to 0.4) and the introns are shorter (119 bp compared to 214 bp). We also noticed a higher proportion of bases belonging to intergenic regions, 18% in low-GC regions and 11% in other chromosomal regions (Table S5). Interestingly, three repeats of more than 500 bases are present in several low-GC regions (R\_13 in contigs 2, 3 and 6; R\_25 in all contigs except contig 1 and R\_80 in contigs 1 to 4; see Figure 1). No homology was found between these sequences and known repeat elements. Gene function analysis of low-GC regions reveals an enrichment of genes involved in specific cellular mechanisms (Table S6). Thirteen genes are involved in DNA replication including the Anaphase-promoting complex subunit 4, the sister chromatid cohesion protein Dcc1 and 3 Mini-chromosome maintenance genes. Twenty-five genes are involved in microtubule synthesis and microtubule-binding motor proteins (9 genes carrying dynein domains, 6 genes carrying kinesin motor domains and 4 tubulin genes). These genes indicate that the low-GC

regions contain many genes required for *P. calceolata* cellular division. Finally, 18 genes are involved in transcription including 3 genes encoding RNA Pol II rpb2 subunits and 7 genes encoding transcription factors suggesting an important role of these chromosomal regions for the regulation of gene expression.

#### Supplementary Note 5: Centromeres, recombination and meiosis in *P. calceolata*

Among 23 gene functions specifically involved in meiosis, 18 homologs are present in the *P. calceolata* genome (Table S7). We did not identify homologs of ZIP, HOP1 and RED1 genes in the *P. calceolata* genome. These genes are known to be involved in homologous pairing of chromatids and construction of the synaptonemal complex in animals, plants and fungi but are not essential to perform meiosis. Indeed these 3 genes are absent in several phyla capable of meiosis like diatoms and ciliates<sup>10,11</sup>. Interestingly, MSH4 and MSH5 genes are absent from the *P. calceolata* genome. These genes necessary to perform the interfering (class I) recombination pathway through Double Holliday Junctions are present in most eukaryotic lineages. The large low-GC regions could be a consequence of the absence of the MSH4/5 genes. Indeed, the main hypothesis to explain these low-GC patterns in centromeres is the importance of GC-biased gene conversions (gBGC) during recombination and the inhibition of this recombination near centromeres<sup>12</sup>. gBGCs increase the GC content of recombining DNA over evolutionary time inducing GC content variations within and between genomes<sup>13</sup>. The kinetochore formation at centromeric regions inhibits recombination and double strand break formation during meiosis resulting in rare gBGC in these regions<sup>14</sup>. Centromeric and peri-centromeric regions may therefore have a lower GC content. In yeasts (*Yarrowia lipolytica*, *Candida lusitanae*, and *Pichia stipites*), a correlation has been observed between the importance of the GC trough near the centromeres (>10%) and the absence of MSH4/MSH5 genes<sup>9</sup>. It is therefore possible that the absence of this recombination pathway in *P. calceolata* induces more frequently double-strand break repair by synthesis-dependent strand annealing and a more rapid gBGCs across the genome except at the centromeres where double-strand breaks are inhibited. This recombination inhibition may have important consequences on the evolution of the *P. calceolata* genome. Genes within low-GC regions are significantly longer and contain more introns than genes in other genomic regions. Because intron gain and loss are closely related to double-strand break repair and homologous recombination, we suggest that centromere genes retain more introns because double-strand breaks are reduced<sup>15</sup>. Variant analysis in *P. calceolata* populations could be targeted specifically in future studies to infer an estimation of recombination rate and more generally characterize the evolutionary processes controlling these large centromere regions.

In addition, we used the mRNA extracted and sequenced during exponential growth of *P. calceolata* to estimate the relative expression level of meiosis genes. Four out of 18 genes were not detected including the double-strand DNA break initiator SPO11 (Table S7). This result suggests that *P. calceolata* do not perform meiosis in exponential growth.

### Supplementary References

1. Blanc-Mathieu, R. *et al.* An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics* **15**, 1103 (2014).
2. Mizuguchi, T. *et al.* Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* **516**, 432–435 (2014).
3. Varoquaux, N. *et al.* Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids Res* **43**, 5331–5339 (2015).
4. Muller, H., Gil, J. & Drinnenberg, I. A. The Impact of Centromeres on Spatial Genome Architecture. *Trends in Genetics* **35**, 565–578 (2019).
5. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**, i351-358 (2005).
6. Talbert, P. B. & Henikoff, S. What makes a centromere? *Experimental Cell Research* **389**, 111895 (2020).
7. Diner, R. E. *et al.* Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *PNAS* **114**, E6015–E6024 (2017).
8. Kanesaki, Y., Imamura, S., Matsuzaki, M. & Tanaka, K. Identification of centromere regions in chromosomes of a unicellular red alga, *Cyanidioschyzon merolae*. *FEBS Letters* **589**, 1219–1224 (2015).
9. Lynch, D. B., Logue, M. E., Butler, G. & Wolfe, K. H. Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biol Evol* **2**, 572–583 (2010).

10. Chi, J., Mahé, F., Loidl, J., Logsdon, J. & Dunthorn, M. Meiosis gene inventory of four ciliates reveals the prevalence of a synaptonemal complex-independent crossover pathway. *Mol Biol Evol* **31**, 660–672 (2014).
11. Patil, S. *et al.* Identification of the meiotic toolkit in diatoms and exploration of meiosis-specific SPO11 and RAD51 homologs in the sexual species *Pseudo-nitzschia multistriata* and *Seminavis robusta*. *BMC Genomics* **16**, 930 (2015).
12. Nambiar, M. & Smith, G. R. Repression of harmful meiotic recombination in centromeric regions. *Semin Cell Dev Biol* **54**, 188–197 (2016).
13. Pessia, E. *et al.* Evidence for Widespread GC-biased Gene Conversion in Eukaryotes. *Genome Biol Evol* **4**, 675–682 (2012).
14. Vincenten, N. *et al.* The kinetochore prevents centromere-proximal crossover recombination during meiosis. *eLife* **4**, e10850 (2015).
15. Farlow, A., Meduri, E. & Schlötterer, C. DNA double-strand break repair and the evolution of intron density. *Trends Genet* **27**, 1–6 (2011).

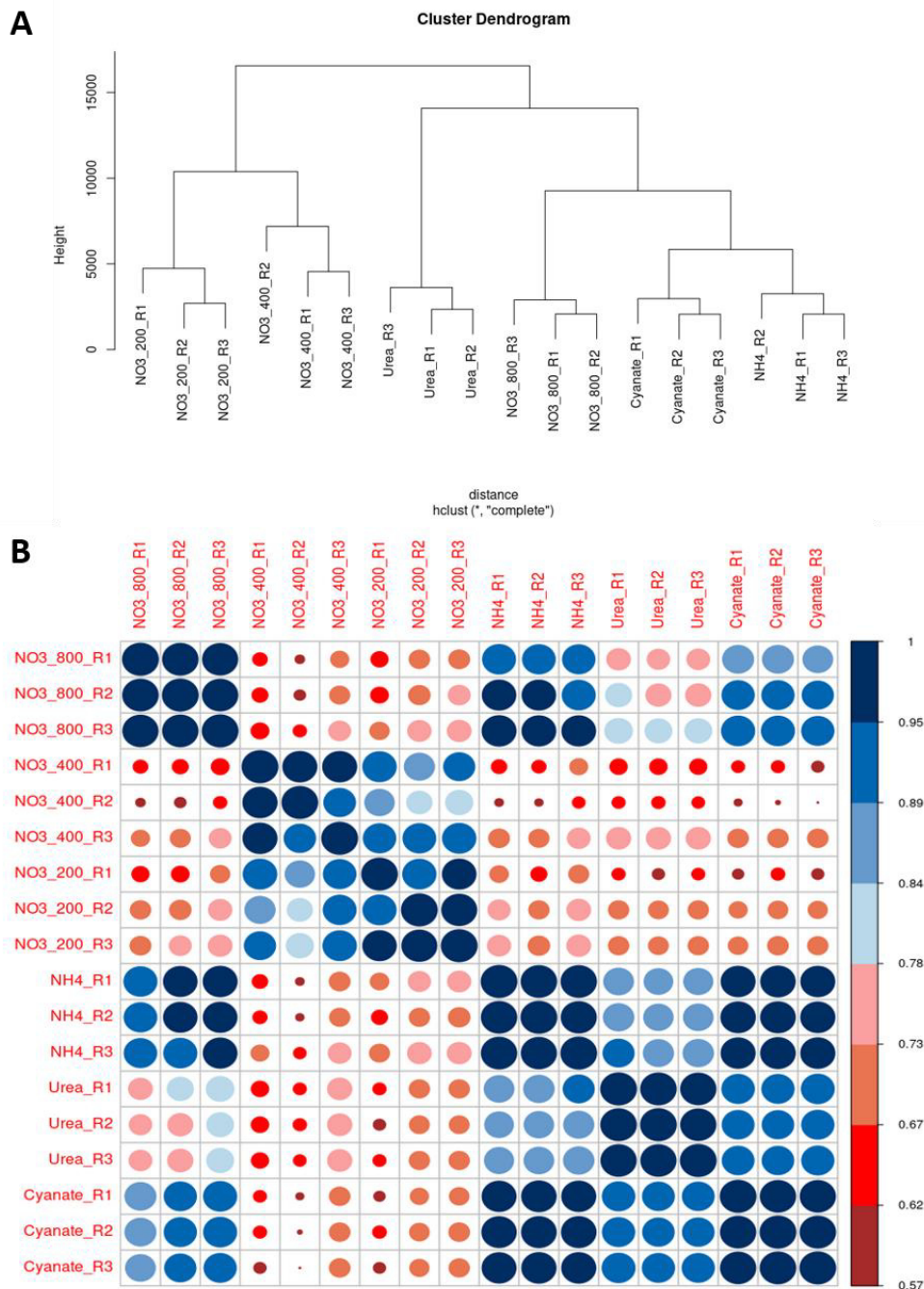


## 2 - Annexes du chapitre 2

### Supplementary Tables

Supplementary tables were send to the members of the dissertation comitee.

### Supplementary Figure



Supplementary Figure 1. Global gene expression patterns between samples. A) Hierarchical clustering of normalized gene expression levels (TPM) of all samples. B) Pearson's correlations of normalised transcriptomic profiles (TPM) between each pair of samples.



### 3 - Annexes du chapitre 3

#### Tables supplémentaires

Les tables supplémentaires ont été envoyées aux membres du jury de thèse.

#### Figures supplémentaires

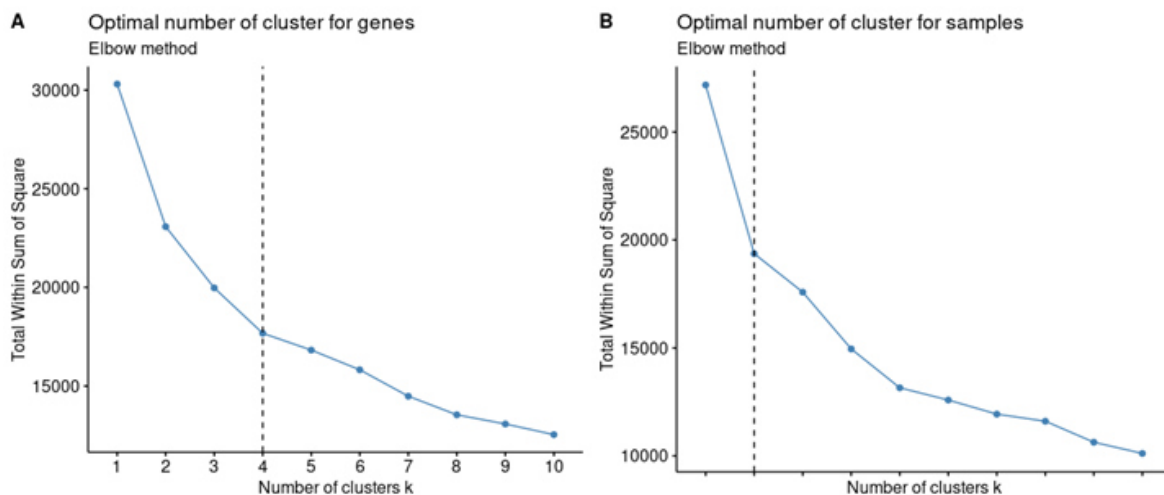


Figure supplémentaire 1. Calcul du nombre optimal de clusters pour la heatmap (figure 9) pour A) les gènes et B) les échantillons, en utilisant la méthode de calcul des k-moyennes Elbow.

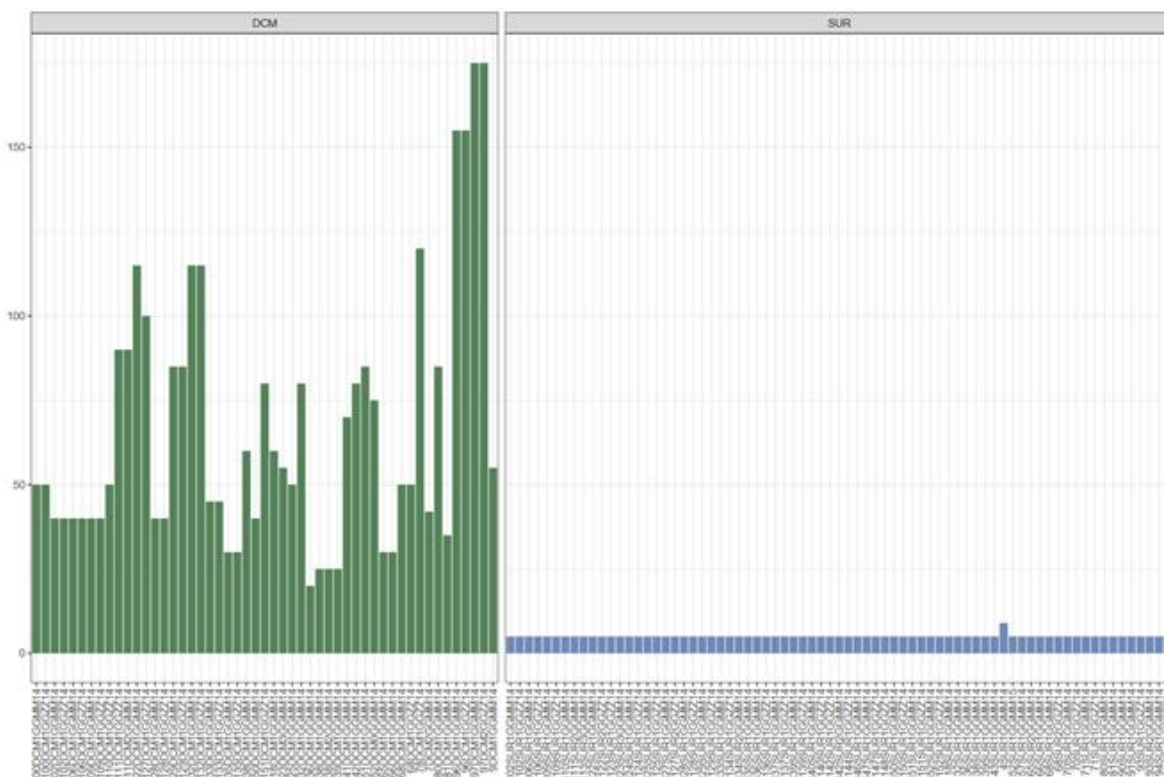


Figure supplémentaire 2. Profondeur d'échantillonnage en mètre pour les échantillons de DCM (vert) et de surface (bleu) en mètre.



