Picoplancton de l'Océan Atlantique Nord-Ouest
Diversité et structure des communautés eucaryotes

Thèse

Cindy Dasilva

Doctorat interuniversitaire en Océanographie

Philosophiae doctor (Ph. D.)

Québec, Canada

Résumé

Le picoplancton inclut des organismes dont la taille varie entre 0.2 et 3 µm, comprenant des cyanobactéries, ainsi qu'une grande diversité d'eucaryotes dont l'écologie et la diversité sont peu connues. Ces deux composantes assurent une part importante de la fixation du carbone dans les océans en contribuant de manière significative à la biomasse et à la production primaire.

L'objectif de cette thèse a été d'appréhender l'importance relative des espèces picoeucaryotiques autotrophes et hétérotrophes. L'étude a porté sur des échantillons de surface et du maximum de chlorophylle sub-surface prélevés dans le golfe du Maine (Juillet 2010) et le plateau Néo-Écossais (Avril et Octobre 2009). La combinaison de plusieurs techniques moléculaires (banques de clones, pyroséquençage), de plusieurs approches (gène structural 18S ARNr, gène fonctionnel *psbA* codant pour la proteine D1 du photosystème II), ainsi que la comparaison de l'ADNr vs ARNr ont été appliqués.

Il apparait que les picoeucaryotes forment une communauté dynamique avec des assemblages distincts, entre les communautés issues des efflorescences printanières vs conditions automnales, ainsi qu'entre régions proches des côtes vs au large. Les différences de distribution, de composition taxonomique et de structure sont associées à des changements hydrographiques, où la température (qui influence fortement la stratification) semble jouer un rôle majeur. L'étude de la composition taxonomique des picoeucaryotes autotrophes souligne la diversité des haptophytes et l'abondance relative des Mamiellophyceae (Chlorophyta). La diversité au sein des flagellés hétérotrophes montre une variabilité dans le temps pour certains taxons (Choanoflagellates, MAST-1, -4, -7) alors que d'autres persistent. La comparaison ADNr/ARNr a permis d'obtenir une vision d'ensemble sur l'état des communautés et clarifie les potentialités de ces approches.

Abstract

Picoplankton includes planktonic microbes ranging from 0.2 to 3 μm in diameter. Among photosynthetic groups are cyanobacteria, and microbial eukaryotes whose ecology and diversity are poorly known. These picophytoplankton contribute significantly to oceanic carbon fixation and primary production.

The objective of this thesis was to investigate the relative importance of autotrophic plankton along with heterotrophic picoeukaryotes. The study focused on samples of surface and subsurface chlorophyll maximum collected in the Gulf of Maine (July 2010) and the Scotian Shelf (April and October 2009). The combination of molecular techniques clone libraries and pyrosequencing, targeting 18S rRNA genes and transcripts as well as the functional gene *psbA,* which codes for a protein D1 of photosystem II were used.

It was apparent that the picoeukaryotes are dynamic communities with distinct assemblages during the spring blooms compared to late summer - autumn conditions, and between regions near shore vs offshore. Differences in distribution, species composition and community structure were associated with hydrographic changes, where the temperature, which strongly influences the stratification, seemed to play a major role. Overall the study revealed the diversity of haptophytes and the high relative abundance of Mamiellophyceae (Chlorophyta). Heterotrophic flagellates varied over time for some taxa (Choanoflagellates, MAST-1, -4, -7) while others persisted. Comparisons between rRNA genes rRNA provided a new perspection on the status of communities and highlighted the potential of using both approachesto describe the state of the ecosystem.

Table des matières

Résumé iii

Abstract v

Liste des tableaux xi

Liste des figures xiii

Liste des abréviations et des sigles xvii

Remerciements xix

Avant-propos xxiii

Chapitre 1 - Introduction générale 1

1.1 Les protistes marins 1

1.2 La boucle microbienne 2

1.3 Composition et diversité du picoplancton 2

1.3.1 Diversité des picocyanobactéries 4

1.3.2 Diversité des picoeucaryotes 5

1.4 Écologie du picoplancton : caractéristique, distribution et abondance 11

1.5 Objectifs de recherche 16

1.6 Méthodologies 17

1.6.1 Séquençage : banques de clones vs pyroséquençage 17

1.6.2 Les gènes cibles : 18 S vs *psbA* 17

1.6.3 Les acides nucléiques : ADNr vs ARNr 19

1.6.4 Autres techniques supplémentaire 19

1.7 Sites d'étude 19

1.8 Plan de thèse 23

Chapitre 2 - Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean 27

Résumé 27

Abstract 28

2.1 Introduction 29

2.2 Methods 31

2.2.1 Study sites and general sampling 31

2.2.2 Laboratory protocols 33

2.2.3 Sequence and phylogenetic analyses 34

2.2.4 Data analysis and statistical tests 37

2.3 Results 37

2.3.1 Environmental variables 37

2.3.2 Clone libraries and community comparisons 40

2.3.3 Phytoplankton phylogenies 43

2.3.4 Communities, environment and taxa 45

2.4 Discussion 45

2.4.1 New records of taxa in a well studied system 45

2.4.2 Spring picophytoplankton 49

2.4.3 Autumn phytoplankton 50

2.4.4 Heterotrophic microbial eukaryotes 52

2.4.5 Environmental influences 55

2.5 Acknowledgements 56

Annex 2 57

Chapitre 3 - Tracking environmental selection of marine microbial eukaryotes using 18S rRNA genes and 18S rRNA in the Gulf of Maine ... 69

Résumé ... 69

Abstract ... 70

3.1 Introduction ... 71

3.2 Materials and methods ... 73

3.2.1 Study region and general sampling ... 73

3.2.2 Laboratory protocols ... 74

3.2.3 Quality control and taxonomic assignments ... 75

3.2.4 Statistical analyses ... 75

3.3 Results ... 76

3.3.1 Environmental parameters ... 76

3.3.2 Pyrosequencing results, quality control and diversity index ... 78

3.3.3 Biogeography and biodiversity of MME ... 80

3.3.4 Global differences between the DNA and RNA templates ... 84

3.4 Discussion ... 86

3.4.1 Communities from rRNA gene and rRNA ... 86

3.4.2 Biogeography and taxonomic composition of MME community ... 91

3.5 Acknowledgements ... 94

Annex 3 ... 95

Chapitre 4 – Structure and phylogenetic diversity of active photosynthetic picoplancton by photosystem-II psbA transcript in the Gulf of Maine ... 105

Résumé ... 105

Abstract ... 106

4.1 Introduction ... 107

4.2 Materials and methods 109
4.2.1 Study region and general sampling 109
4.2.2 Laboratory Protocols 110
4.2.3 Phylogenetic community structure analyses 111
4.2.4 Statistical analyses 112
4.3 Results 114
4.3.1 Environmental variables 114
4.3.2 Chlorophyll *a* concentration and picophytoplanktonic categories 116
4.3.3 Phylogeny and *psbA* diversity 118
4.4 Discussion 124
4.4.1 Picophytoplanktonic cells and Fv:Fm 124
4.4.2 Phylogeny and distribution of picocyanobacteria 124
4.4.3 Diversity of eukaryotes inferred by *psbA* 126
4.5 Acknowledgements 130
Annex 4 132
Chapitre 5 - Conclusions générales 139
5.1 Principales contributions de la thèse 139
5.2 Synthèses des principales variations taxonomiques dans le temps 141
5.3 Discussion sur la méthodologie 142
5.4 Perspectives 144
Bibliographie 147

Liste des tableaux

Table 2.1 : Environmental and cell counts informations for samples used for 18S rRNA gene clone libraries collected stations from the Halifax line and Browns Bank line in April and October 2009. Nutrients; total extracted chlorophyll *a* ; percent of Chl *a* <3µm in total Chl *a*; flow cytometry cell counts with nanophytoplankton (Nano), picoeukaryotes (PPE), *Synechococcus* (SYN), *Prochlorococcus* (PRO), and bacteria (Bact)..... 36

Table 3.1 : Water column characteristics of the six stations from Jordan Basin (JB), North East Channel (NEC) and North East Fan (NEF) with coordinates of each station (Latitude: Lat., and Longitude: Long.); date; and local time of sampling. Bottom depth is for bathymetry comparison between two regions; depth of subsurface chlorophyll maximum (SCM depth); depth of euphotic zone (Euph depth); depth of pycnocline (Pycno depth, see materials and methods); and euphotic zone/mixture zone ratio (Z_{eu}/Z_{mix}, see materials and methods)……..77

Table 3.2 : Chlorophyll *a* concentrations and physicochemical parameters of the samples selected for tag pyrosequencing from Jordan Basin (JB) and North East Channel (NEC) or Fan (NEF). Extracted total chlorophyll *a* (Tot. Chl *a*); temperature (Temp); salinity (Sal); and nutrient concentrations: ammonium (NH_4^+), nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^{4-})……..….81

Table 3.3 : Linear correlation between DNA-reads and RNA-reads (Significant $p < 0.05$). Standard error (St. Err.) was calculated only for taxa found in all stations; others are no determinate (nd)…92

Table 4.1 : Sampling locations, Chlorophyll *a* concentrations, and physicochemical parameters. Total (Tot Chl *a*) and small size fraction (Chl *a* ≤3µm) chlorophyll *a* concentration; Temperature (Temp); Salinity (Sal); Photosynthetically Available Radiation (PAR); nutrients: ammonium (NH_4^+), nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^{4-})……..….112

Table 4.2 : Picophytoplankton from flow cytometry data: photosynthetic picoeukaryotes (PPE), *Synechococcus* (SYN), and *Prochlorococcus* (PRO); and the contribution of PPE to total picophytoplankton (PPP) in term of cell concentrations and carbon biomass expressed as percentage…….116

Tableau 5.1 : Tableau résumant les données physiques, chimiques, et biologiques des trois pèriodes d'échantillonnages dans le golfe du Maine en Juillet 2010 (été), et sur le plateau Néo-Écossais en Avril 2009 (printemps) et Octobre 2009 (automne)…….144

Liste des figures

Figure 0.1: Cartes représentant les zones d'échantillonnages des deux programmes de recherches associés **a)** PMZA et **b)** Discovery Corridor..... xxiv

Figure 1.1: Schéma simplifié du réseau trophique marin regroupé selon la taille et selon le mode de nutrition.…… ……..3

Figure 1.2: Relations phylogénétiques entre les picocyanobactéries marines : *Synechococcus* et *Prochlorococcus*, basé sur des séquences du gène 16S ARNr.. 6

Figure 1.3: Picoeucaryotes marins observés par **(a, b)** microscopie optique, **(c, g)** microscopie à épifluorescence, et **(h, i)** microscopie électronique à balayage..... 8

Figure 1.4: Arbre phylogénétique des eucaryotes (tiré de Adl et al., 2012)..... 10

Figure 1.5: **a)** Relative contribution des haptophytes à la biomasse totale de chlorophylle *a* dans la couche euphotique des océans lors de l'année 2000; **b)** Arbre phylogénétique des haptophytes basé sur LSU ADNr; **c)** Zoom sur la famille des Calcihaptophycidae 12

Figure 1.6: Arbres phylogénétiques basés sur des séquences du gène 18S ARNr mettant en évidence les différents clades au sein des **a)** Chlorophyta; **b)** marines stramenopiles..... 13

Figure 1.7: Moyenne annuelle des abondances à la surface des eaux océaniques pour **a)** *Prochlorococcus*, et **b)** *Synechococcus*; et **c)** moyenne de la biomasse (µg C l^{-1}) en fonction de la profondeur pour les 3 groupes du picophytoplancton *Prochlorococcus* en bleu, *Synechococcus* en rouge, et picoeucaryotes en vert..... 15

Figure 1.8: Schéma représentant les limites de la diversité accessibles par les deux techniques de séquençage utilisées au cours de cette thèse. 18

Figure 1.9: **a)** Illustration simplifiée de la structure et du fonctionnement du photosystème II chez les eucaryotes. **b)** Schéma de la dégradation et de la re-synthèse de la protéine D1 suite à une photoinhibition et inactivation du photosystème II. 20

Figure 1.10: Carte représentant les deux sites d'échantillonnages de cette thèse situés dans l'Océan Atlantiques Nord-Ouest: le golfe du Maine et le plateau Néo-Écossais 22

Figure 1.11: **a)** Séries temporelles de 1980 à 2009 de la température et de la salinité d'une station fixe du plateau Néo-Écossais. **b)** Tendances temporelles de 1972 à 2009 des nutriments dans la partie centrale du plateau Néo-Écossais. **c)** Cycle annuel pluriannuel de l'abondance du phytoplancton et de la température issues de plusieurs stations du programme AZMP. 24

Figure 1.12: Schéma récapitulatif de la structure de la thèse. 25

Figure 2.1: Sampling sites along the Scotian Shelf transects. Transects were carried out in April and October 2009. Stations used for clone libraries are indicated by an open circle..... 32

Figure 2.2: Section plots along the Halifax and Browns Bank Lines, from April and October 2009. Temperature and salinity values from CTD cast along transects. Picoeukaryote concentrations (× 10^3 cells mL^{-1}) using FCM counts. Carbon biomass picoeukaryotes is the contribution of picoeukaryotes to picophytoplankton carbon expressed as percentage of total. 38

Figure 2.3: Global maximum likelihood (ML) phylogenetic tree using 704 nearly full length sequences..... 40

Figure 2.4: Cluster analysis based on phylogenetic lineages using weighted UniFrac (824 sequences) and matrix of relative abundance of taxa. 42

Figure 2.5: ML phylogenies of Chlorophyta rooted with *Cyanoptyche gleocystis* (AJ07275) and *Glaucocystis nostochinearum* (X70803). 44

Figure 2.6: ML phylogenies of **a)** Photosynthetic stramenopiles were rooted with *Gyrodinium helviticum* (AB120004). **b)** Haptophyta rooted with Cryptophyceae sp. CCMP2045 (GQ375264). **c)** Cryptophyta rooted with *Emiliania huxleyi* (HQ877901). 46

Figure 2.7: ML phylogeny of dinoflagellates rooted with *Tetrahymena pyriformis* (EF070254) and *Strombidium* sp.SBB99-1 (AY143565) 48

Figure 2.8: Canonical Correspondence Analysis (CCA) plot using relative abundance of taxa (%) and physical-chemical variables detected **a)** in April and **b)** in October run on CANOCO……. 54

Figure 3.1: CTD profiles of the upper 70m of the six stations: temperature (°C, red solid line), Salinity (black dashed line), in situ fluorescence (µg L^{-1}, green heavy hatched line), and *in situ* nitrate concentrations (µM, blue dotted line) calibrated from bottle sample data ……79

Figure 3.2: **a)** Relative abundance of reads according major groups. **b)** Number of OTUs shared between regions and templates (total OTUs 15 717). 83

Figure 3.3: Phylogenetic structure of eukaryotes communities based on OTUs clustered at ≥98% sequences identity. **a)** Corresponding Principal Coordinate Analysis (PCoA) using UniFrac weighted distance metric. **b)** Unweighted Unifrac cluster analysis associated with matrix of relative abundance of taxa which are mainly responsible of difference between stations and/or templates identified by SIMPER test..... 85

Figure 3.4: Canonical correspondence analysis (CCA) showing the relative abundance of marine microbial eukaryotes taxa (%) recovered by RNA in relation to environmental parameters…….87

Figure 3.5: a) Scatterplots of paired comparisons between number of DNA-reads and RNA-reads according to sampling stations and major groups 89
b) Comparison of DNA to RNA ratio by taxon. Taxa were clustered by similarity of the DNA to RNA ratio using cluster analysis (Bray Curtis) run on PAST software.……90

Figure 4.1: Si:N:P molar ratio of samples collected in surface and SCM during GoMA cruise, delimited by the Redfield ratios Si:N:P = 15:16:1. 115

Figure 4.2: Relationship between changes in F_V/F_m ratio measured from small size fraction and biological parameters related to picophytoplankton **a)** Small Chl *a* concentrations, **b)** Abundance of total picophytoplanktonic cells measured using flow cytometry (FCM), and **c)** Carbon biomass of picophytoplanktonic cells. 119

Figure 4.3: Canonical Correspondence Analysis (CCA) plot run on PAST using small Chl *a* concentrations, physical-chemical variables and abundance of picophytoplanktonic cells (*Synechococcus*, *Prochlorococcus*, and picoeukaryotes photosynthetic) measured by flow cytometry..... 121

Figure 4.4: UniFrac cluster analysis and Principal Coordinate Analysis using UniFrac weighted distance metric for picocyanobacteria. 123

Figure 4.5: ML phylogenetic trees of cyanobacteria based on psbA OTUs (≥99% similarity) from the Gulf of Maine; **a)** *Synechococcus*, and **b)** *Prochlorococcus.* 125

Figure 4.6: Unifrac cluster analysis (left) and Principal Coordinate Analysis (right) using UniFrac weighted distance metric for photosynthetic picoeukaryotes (PPE). 127

Figure 4.8: Phylogenetic tree (ML) of OTUs at 99% **A)** Chlorophyta (349 *psbA* sequences), **B)** Cryptophyta (43 *psbA* sequences), and **C)** Stramenopiles with classes annoted like as "a" is Dictyochophyceae, "b" is Chrysophyceae, "c" is Pelagophyceae, "d" is diatoms)…….130

Liste des abréviations et des sigles

18S : Petite sous-unité du ribosome des eucaryotes

ADN ou DNA : Acides désoxyribonucléiques

ADNc ou cDNA : Acides désoxyribonucléiques complémentaires

ARN ou RNA : Acides ribonucléiques

ARNr ou rRNA : Acides ribonucléiques ribosomaux

BLAST : Basic local alignment search tool

pb ou bp : paires de bases

C : Carbone

CCA : Canonical correspondence analysis

CCMP : Provasoli-Guillard National Center for Culture of Marine Phytoplankton (National Center for Marine Algae and Microbiota)

Chl *a* : Chlorophylle *a*

CLUMEQ : Consortium de recherche pour le calcul scientifique de haute performance

C:N : Carbon to nitrogen ratio

CTD : Conductivity-Temperature-Depth

DAPI : 4'6-diamidino-2-phenylindole

dNTP : Désoxyribonucléotide triphosphate

EDTA : Acide éthylènediaminetetraacetique

GF/F : Glass microfiber filter

GoMA : Golfe du Maine

HPLC : High performance liquid chromatography

IBIS : Institut de Biologie Intégrative et des Systèmes

JB : Jordan Basin

MID : Multiplex identifier

ML : Maximum-likelihood

NCBI : National Center for Biotechnology Information

NE: North East

NEC: North East Channel

NEF: North East Fan

NJ : Neighbour-joining

N:P : Nitrogen to phosphorus ratio

OTU : Operational taxonomic unit

PAR : Photosynthetically available radiation

PAST : PAleontological STatistics software

PC : Polycarbonate

PCA : Principal component analysis

PCoA : Principal Coordinate analysis

PCR : Polymerase chain reaction

PPE : Photosynthétique picoeucaryote

PPP : Picophytoplancton

SCM : Subsurface chlorophyll maximum

SF : Surface

SS : Scotian Shelf

Tris : Tris (hydroxymethyl) aminomethane

V4 : région hypervariable 4 de l'ARN ribosomal

Remerciements

Je souhaite tout d'abord remercier ma directrice de thèse Connie Lovejoy qui m'a offert l'opportunité d'effectuer cette thèse sur un projet qui m'a vraiment passionné. Je la remercie pour tous ce qu'elle m'a enseignée, son aide et son soutien au cours de ces années. Elle a su obtenir le meilleur de moi-même grâce à son sens critique, sa patience et surtout son humour. J'estime sincèrement avoir eu de la chance d'être supervisé par elle, tant sur le plan scientifique que personnel. Je souhaite aussi exprimer ma reconnaissance à mon co-directeur William K. W. Li (chercheur à l'Institut Océanographique de Bedford) qui a accepté de se joindre à ce projet. C'est un grand honneur pour moi d'avoir pu bénéficier de ses compétences, ses conseils et de sa disponibilité.

Ensuite, je voudrais remercier les membres de mon comité d'encadrement qui m'ont apportés leurs excellents conseils au cours de nos réunions: M. Jean-Éric Tremblay, M. Guillaume Massé et M. Christian Landry (tous professeurs à l'Université Laval). C'était une excellente motivation tout au long du parcours de recevoir ces critiques extérieures (toujours dans la bonne humeur). Merci aussi à Mme Julie Laroche (professeure à l'Université de Dalhousie) d'avoir accepté de faire partie de mon jury d'évaluation en tant qu'examinateur externe.

Des remerciements spéciaux reviennent à l'exceptionnelle équipe de l'Institut Océanographique de Bedford lors des missions sur le bateau CCGS Hudson. Leur générosité et leurs compétences a fait toute la différence. Un merci particulier à Emma et Karen pour ses longues heures de filtration (parfois avec tempête et/ou malade).

Mes remerciements aux membres des laboratoires Lovejoy et Vincent pour leur accueil. J'ai été chanceuse de travailler au sein d‟un groupe de recherche où se promouvait un climat de collaboration et d'entraide. Une pensée particulière à mes deux collègues de bureau Ramon Terrado et Mary Thaler. Ce fut un réel plaisir de vous avoir à mes cotés pendant ces années et je garderai d'excellents souvenirs de nos discussions et de votre soutien. Un énorme merci aux post doc pour leur disponibilité et enseignement André Comeau, Adam Monier, et Marie Lionard surtout en bioinformatique. Merci aussi à Marianne Potvin de m'avoir formé et guidé dans les techniques de biologie moléculaire

pour mes premiers pas au sein du labo. Sans oublier « Les filles » du labo Lovejoy qui ont partagés l'aventure du doctorat à travers nos pauses café, nos soirées filles, ou nos séances sportives : Vanit Mohit qui est plus qu'une collègue à présent, Bérangère Péquin pour tous nos papotages, Emmanuelle Médrinale mon binome de projet, Sophie Charvet, Annabelle Baya pour leur aide lorsque que j'étais en déplacement, et tous nos nouveaux membres…..

D'énormes bisous aux non scientifiques qui sont entrée dans ma vie et dont j'ai grande joie de connaitre. Vous venez des quatre coins du monde, vous êtes dans des disciplines très différentes et c'est ca que j'aime!. J'ai énormément appris de vous, merci pour votre soutien et votre écoute pour un domaine qui n'était pas le votre : Samuel, Alice, Cheick, Ronald, Momo, Nina, et spécialement Abdel (sans qui l'aventure n'aurai pas été la même. Je remercie le destin d'avoir croisé nos chemins). Une grande reconnaissance aussi à mes grandes « cop's » qui malgré la distance on toujours été d'un énorme soutien à travers nos longues discussions: Allen, Fanny, Coco, et Aurélie.

Un merci particulier à Marc Bouvy (professur à IRD de Montpellier), qui en m'accueillant pour mes deux stages de maitrise m'a donné la chance de découvrir une réelle passion pour la microbiologie marine et de trouver ma voie. Et même à distance, son soutien et ses conseils pertinents dans les périodes de doutes ont été très bénéfiques.

Enfin, (et il n'y a pas de mots pour dire à quel point je peux être reconnaissante), ma formidable famille : mon petit frère Antoine et ma petite sœur Aida que j'aime trop!!! Aida, vu ton soutien tu mériterais d'être co-auteur! Mes parents, Stéphane et Marie-Dominique qui ont beaucoup fait pour moi malgré certains caprices. La dernière phrase sera pour ma mère : je te remercie de tout mon cœur de m'avoir toujours soutenue et guidée sur le droit chemin!

« Rien ne t'atteindra si ce n'est ce que Dieu te destine »

Avant-propos

Cette thèse étudie la diversité phylogénétique et la structure des communautés picoplanctoniques marines. Elle est structurée en cinq chapitres incluant une introduction générale, trois chapitres de recherche qui constituent le corps de la thèse, et une conclusion générale.

Deux zones géographiques de l'océan Atlantique Nord Ouest sont étudiées : le plateau néo-écossais et le golfe du Maine. Les échantillons utilisés proviennent de trois missions océanographiques effectuées à bord du *CCGS* Hudson en avril et en octobre 2009 (plateau néo-écossais), puis en juillet 2010 (golfe du Maine). Ces missions ont été menées en collaboration avec deux programmes de recherche :

► PMZA (Programme de Monitorage de la Zone Atlantique) débuté en 1998 par le département Pêches et Océans Canada de l'Institut Océanographique de Bedford (Therriault *et al*., 1998). Le principal objectif est de récolter et d'analyser des données biologiques, chimiques et physiques afin de détecter et de suivre les variabilités saisonnières et interannuelles dans les eaux de l'Est canadien (**Figure 0.1a**).
http://www.bio.gc.ca/science/monitoring-monitorage/azmp-pmza-fra.php

► Le programme " Discovery Corridor, " lancé en 2004 conjointement par Pêches et Océans Canada, le Centre pour la Biodiversité Marine et le projet Gulf of Maine Census of Marine Life. Le principal objectif est de dresser l'inventaire des espèces marines qui habitent le corridor afin de savoir comment concilier la conservation de la biodiversité marine et l'utilisation durable des ressources de la mer (**Figure 0.1b**).
http://www.mar.dfo-mpo.gc.ca/f0011522

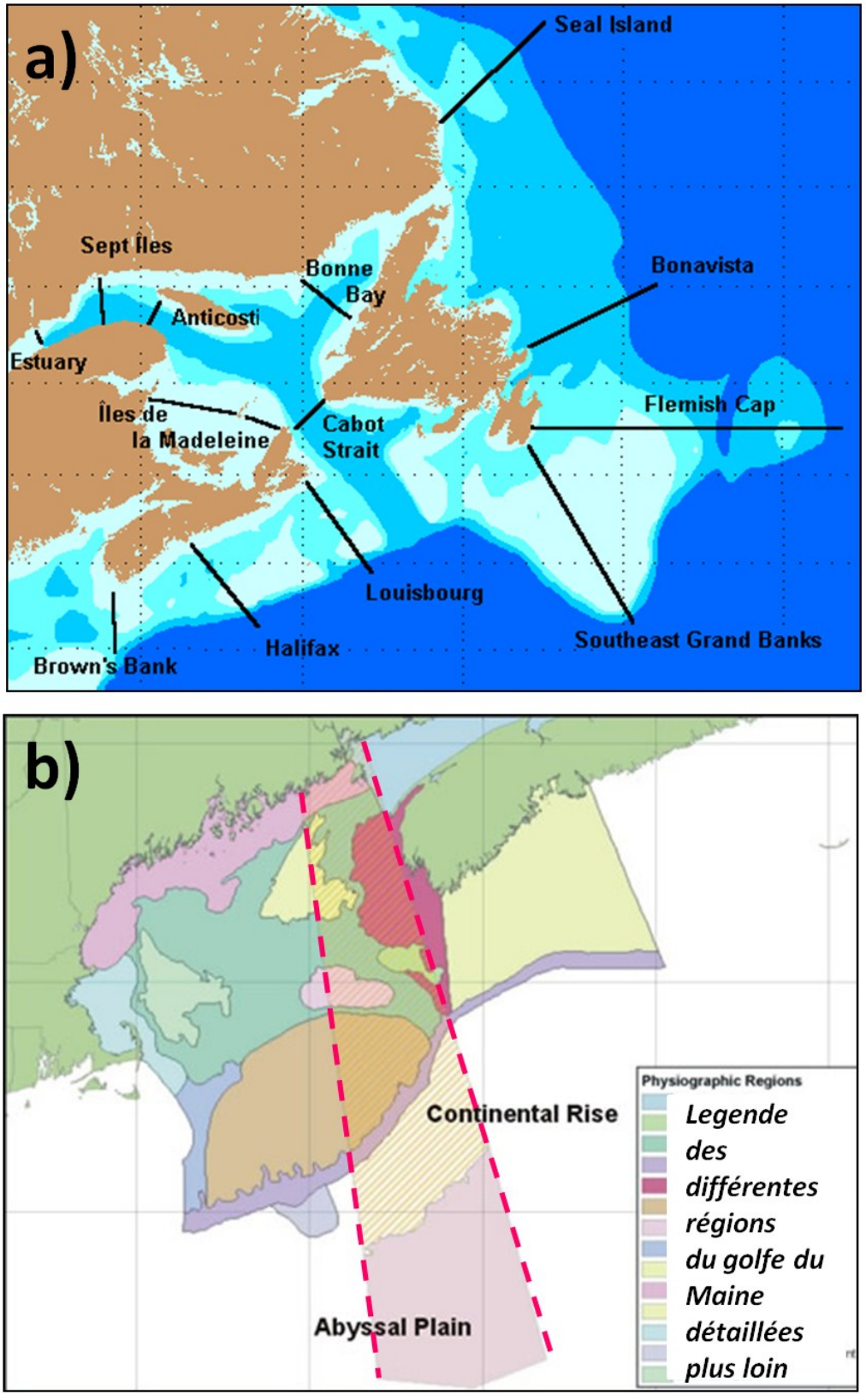


Figure 0.1: Cartes représentant les zones d'échantillonnages des deux programmes de recherches associés a) PMZA et b) Discovery Corridor.

Les principaux résultats de cette thèse ont donné lieu à diverses communications scientifiques lors de congrès nationaux et internationaux, ainsi que lors de séminaires :

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2012) *Gulf of Maine picoeukaryote communities unveiled by 18S rDNA and rRNA pyrosequencing, and comparisons with psbA transcripts*. International Society for Microbial Ecology (ISME), Copenhague, Danemark.

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2011) *Structure and diversity of the spring and autumn subsurface picophytoplankton community in the North Atlantic.* Association for the Sciences of Limnology and Oceanography (ASLO), San Juan, Puerto Rico.

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2011) *Phylogenetic diversity of picoeukaryotes in upper water column from Scotian Shelf.* Canadian Healthy Oceans Network (CHONe), Montréal, Canada.

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2011) *Small planktonic microbes: The smallest have the greatest role.* Canadian Healthy Oceans Network (CHONe), Ottawa, Canada.

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2011) *Variations saisonnière au sein de la communauté des picoeucaryotes du plateau Néo-Écossais.* Institut de Biologie Intégrative et des Systèmes (IBIS), Québec, Canada.

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2010) *Distribution et contribution des groupes du picophytoplancton dans l'Océan Atlantique Nord*. Colloque de biologie de l'Université Laval, Québec, Canada.

Cindy R. Dasilva, William K.W. Li and Connie Lovejoy (2010) *Phylogenetic diversity of picoeukaryotes from Scotian Shelf.* Canadian Society for Ecology and Evolution (CSEE), Québec, Canada.

Ces travaux ont fait l'objet d'affiches aux cours des réunions annuelles 2009, 2010, 2011, et 2012 du regroupement inter-institutionnel Québec-Océan.

De plus, des travaux réalisés à titre de coauteur ont abouti à une publication :

Terrado R., Medrinal E., **Dasilva C**. **R.**, Thaler M., Vincent W.F., and Lovejoy C. (2011) *Protist community composition during spring in an Arctic flaw lead polynya*. Polar Biology, 34 : 1901–1914.

Enfin, ce travail est une contribution au réseau CHONe (Canadian Healthy Oceans Network) financé par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG). Un soutien financier complémentaire a également été apporté par le Fonds de Recherche du Québec –nature et technologie au regroupement interinstitutionnel Québec-Océan et par les subventions CSRNG du Dr Connie Lovejoy.

Chapitre 1 - Introduction générale

1.1 Les protistes marins

L'importance quantitative et fonctionnelle des microorganismes dans les réseaux trophiques aquatiques (bactéries, archées et eucaryotes) a été soulignée dans les années 70 (Pomeroy, 1974). Les protistes marins forment l'ensemble des eucaryotes unicellulaires constituant une grande part de la diversité génétique des océans (Giovannoni and Stingl, 2005; Pedròs-Aliò, 2006). Ils sont à la base des réseaux trophiques pélagiques et de nombreuses études ont permis de mettre en évidence leur rôle essentiel dans l'équilibre de leur environnement notamment par leur rôle prépondérant dans le maintien des cycles biogéochimiques (Delong et Karl, 2005; Arrigo, 2005; Sogin *et al*., 2006; Kirchman, 2008). Le phytoplancton marin tient un rôle clé dans la production primaire, où en fixant le dioxyde de carbone par photosynthèse, il est un acteur essentiel du cycle du carbone (Calbet et Landry, 2004). Ce phytoplancton est connu pour réagir rapidement aux changements environnementaux dépendant des paramètres des masses d'eaux dans lesquelles il flotte librement (Durham *et al*., 2009). Ainsi, la structure des communautés phytoplanctoniques reflète leur adaptation à divers facteurs environnementaux. Sa croissance et sa distribution sont influencées par des facteurs nutritionnels (essentiellement les concentrations en sels azotés, phosphates, fer), des facteurs physiques (lumière, turbulence, stratification induite par la température ou la densité), et des facteurs biotiques (broutage ou lyze virale) (Bougis, 1974; Caron *et al*., 1999). Au niveau de sa diversité, de nombreuses espèces sont dites mixotrophes combinant alternativement l'autotrophie et l'hétérotrophie qui s'avère être avantageuse lorsque les nutriments et/ou la lumière sont limités (Hartmann *et al*., 2012; Flynn *et al*., 2012; Hartmann *et al*., 2013). Établir comment la variabilité temporelle et spatiale de l'environnement structure ces communautés est donc essentiel pour nous éclairer sur le fonctionnement et la dynamique des écosystèmes aquatiques.

1.2 La boucle microbienne

C'est en 1983 qu'Azam *et al*. sous-entend que les bactéries hétérotrophes, les flagellés et les ciliés participent au maintien d'un système extrêmement dynamique : la boucle microbienne (Fenchel, 2008). Ils exposent que 25% de la production primaire des écosystèmes aquatiques est utilisée par la boucle microbienne, participant au recyclage des éléments et à la reminéralisation de la matière organique et permettant ainsi le transfert de matière et d'énergie vers les maillons trophiques supérieurs. Le picoplancton aussi bien autotrophes qu'hétérotrophes intervient à plusieurs étapes de ce processus, et il apparait nécessaire de mieux caractériser et d'appréhender sa variation en terme de distribution, structure et composition taxonomique.

C'est à travers la photosynthèse (fixation du carbone inorganique et utilisation de l'énergie lumineuse) que le picoplancton autotrophe produit tous les constituants nécessaires à la cellule. La matière organique produite par ces petits producteurs primaires est alors rapidement transférée aux échelons supérieurs par broutage des consommateurs primaires, le picoplancton hétérotrophe ainsi que le nano- et micro- plancton (**Figure 1.1**). Une partie du carbone fixé est libéré sous forme de matière organique dissoute et est consommé majoritairement par des bactéries hétérotrophes et des archées (Caron, 1994; Sarmento et Gasol, 2012). Ces bactéries sont à leur tour broutées par des petits flagellés hétérotrophes (Sherr et Sherr, 2002) et des protozoaires ciliés du nanoplancton, lui-même consommé par le microzooplancton. Les virus aussi jouent un rôle important au sein de la boucle microbienne à travers un contrôle de la production primaire et bactérienne agissant sur la terminaison des efflorescences algales, l'affectation de la dynamique des populations ou encore la diversité des communautés planctoniques (Suttle *et al*., 1990; Wommack et Colwell, 2000).

1.3 Composition et diversité du picoplancton

Les protistes sont souvent classifiés en fonction de leur taille, dont les dimensions sont comprises entre 0,2 µm et 20 mm, allant du picoplancton au macroplancton (Sieburth *et al*. 1978) (**Figure 1.1**). Pendant longtemps, la diversité des cellules picoplanctoniques fut sous estimées étant difficilement détectable et dénombrable par les techniques de micro-

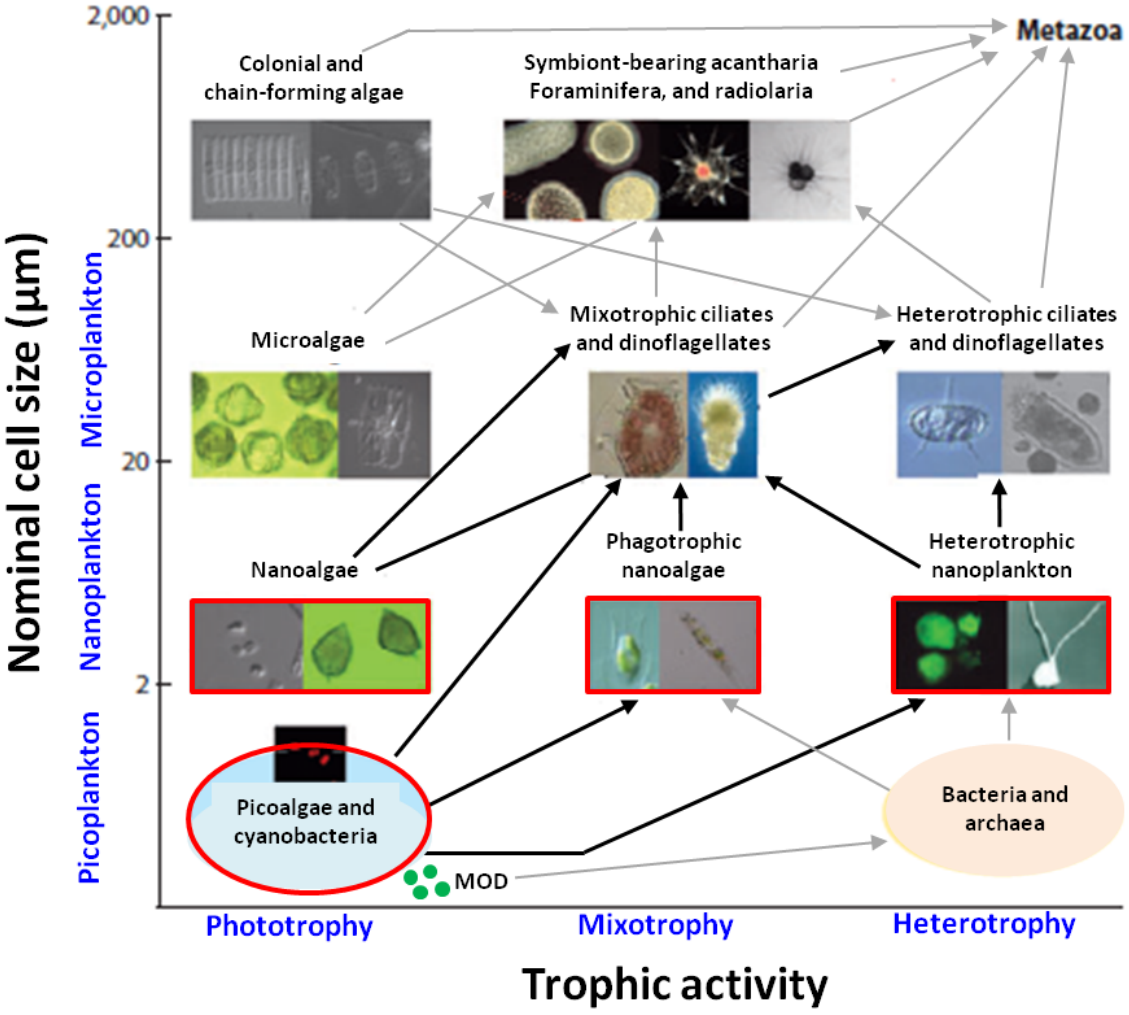


Figure 1.1: Schéma simplifié du réseau trophique marin regroupé selon la taille (axe Y) et selon le mode de nutrition (axe X). Les organismes ciblés au cours de la thèse sont encadrés en rouge. Les flèches représentent les interactions trophiques (flux de carbone au cours de la consommation de proies par les prédateurs) avec les flèches noires pour les organismes ciblés et les flèches grises pour les niveaux supérieurs (tiré et modifié de Caron *et al.*, 2012).

-scopie traditionnelles. Ces cellules sont définies comme ayant une taille allant de 0,2 à 2 µm, mais dans la pratique elles sont souvent déterminées comme des cellules dont la taille est inférieure à 3,0 µm, en raison de la taille des pores des filtres utilisés lors de la filtration (Charpy et Blanchot 1998; Vaulot *et al*., 2008). Au sein du picoplancton on retrouve deux composantes, l'une procaryote appelé bactérioplancton (dont seules les cyanobactéries seront discutées au cours de cette thèse), et l'autre eucaryote. Les picocyanobactéries ont été sujettes à des études plus approfondies en termes de taxonomie, physiologie et écologie, contrairement aux picoeucaryotes dont l'importante diversité fût suggérée tardivement en raison de leur manque de caractères morphologiques distinctifs. L'avènement de nouvelles technologies ont permis de pallier ces limitations et de les étudier directement dans leur environnement.

1.3.1 Diversité des picocyanobactéries

Les cyanobactéries sont généralement considérées comme les ancêtres des plastes selon la théorie endosymbiotique (Morden *et al*., 1992; Selosse *et al* 1997; Stiller et Hall, 1997). Au sein des picocyanobactéries, on dénombre deux genres *Prochlorococcus* et *Synechococcus*. Ils sont majoritairement des photoautotrophes, mais quelques études ont mis en évidence qu'ils avaient également la capacité d'absorber une grande variété de molécules simples comme l'ATP, l'urée, le glucose, le phosphore et l'azote organique ou inorganique (photohétérotrophie) (Zubkov *et al*., 2005; Malmstrom *et al*., 2005; Church *et al*., 2006).

La diversité de ces deux genres a été démontrée à travers diverses études (Partensky *et al*., 1993; Moore *et al*., 1995; Scanlan et West, 2002). Premièrement, en terme de morphologie : *Prochlorococcus* est le plus petit organisme photosynthétique connu à l'heure actuelle avec une taille variant de 0,5 à 0,7 µm (Campbell *et al*., 1994; Johnson *et al*., 2006; Partensky *et al*., 1999a), alors que *Synechococcus* est légèrement plus grand (environ 1 µm de diamètre). Deuxièmement, en terme de composition pigmentaire : il est facile de distinguer *Synechococcus* et *Prochlorococcus* en utilisant la cytométrie en flux où chaque genre présente une signature caractéristique (Olson *et al*., 1990; Vaulot *et al*., 1990). Les cellules du genre *Synechococcus* sont reconnaissables par leur fluorescence orange en raison de leurs phycobilines, contrairement aux cellules du genre

Prochlorococcus qui en manquent. Et troisièmement, en terme de diversité phylogénétique : *Prochlorococcus* présente plusieurs écotypes phylogénétiquement distincts dont certains sont adaptés aux fortes luminosités (HL) et d'autres sont adaptés aux plus faibles luminosités (LL) (Moore *et al*., 1998; Moore *et al*., 1999; Rocap *et al*., 2003) (**Figure 1.2**). A l'inverse, *Synechococcus* est génétiquement plus diversifié dénombrant trois sous-groupes, dont le sous-groupe 5.1 est dominant et au sein duquel on dénombre pas moins de dix clades génétiquement distincts ayant des adaptations spécifiques pour faire face aux gradients de nutriment et de lumière. (Fuller *et al*., 2003; Muhling *et al*., 2006) (**Figure 1.2**).

Finalement, à notre connaissance, actuellement, 40 génomes de *Synechococcus* et 14 génomes de *Prochlorococcus* ont été séquencés (centre nationale de séquencage Genoscope : http://www.genoscope.cns.fr/spip/). Ils ont permis d'avoir une estimation des répertoires de gènes communs, accessoires et uniques, et ainsi de mettre en évidence la présence d'îlots génomiques spécialisés dans l'adaptation à des niches écologiques particulières.

1.3.2 Diversité des picoeucaryotes

En raison de leur manquent de caractères morphologiques distinctifs (**Figure 1.3**), l'étude de la diversité phylogénétique des picoeucaryotes marins nécessite l'utilisation de diverses méthodes moléculaires. Leur grande diversité a été initialement mise en évidence par des études pionnières ciblant l'ARN ribosomal du gène 18S par des banques de clones environnementales (Diez *et al*., 2001; Lopez-Garcia *et al*., 2001; Moon-Van Der Staay *et al*., 2001). Divers taxa jusqu'alors inconnus ont été découverts, et la plupart d'entre eux restent encore à ce jour incultivés (Massana *et al*., 2002). Par la suite de nombreuses études ont confirmé cette tendance, démontrant que les picoeucaryotes sont constitués de groupes paraphylétiques répartis entre plusieurs phylums de l'arbre des eucaryotes (Adl *et al*., 2012) (**Figure 1.4**). On y retrouve des organismes autotrophes participant à la production primaire, et des organismes hétérotrophes qui sont des bactérivores ou des parasites (Guillou *et al*., 2008; Siano *et al*., 2011).

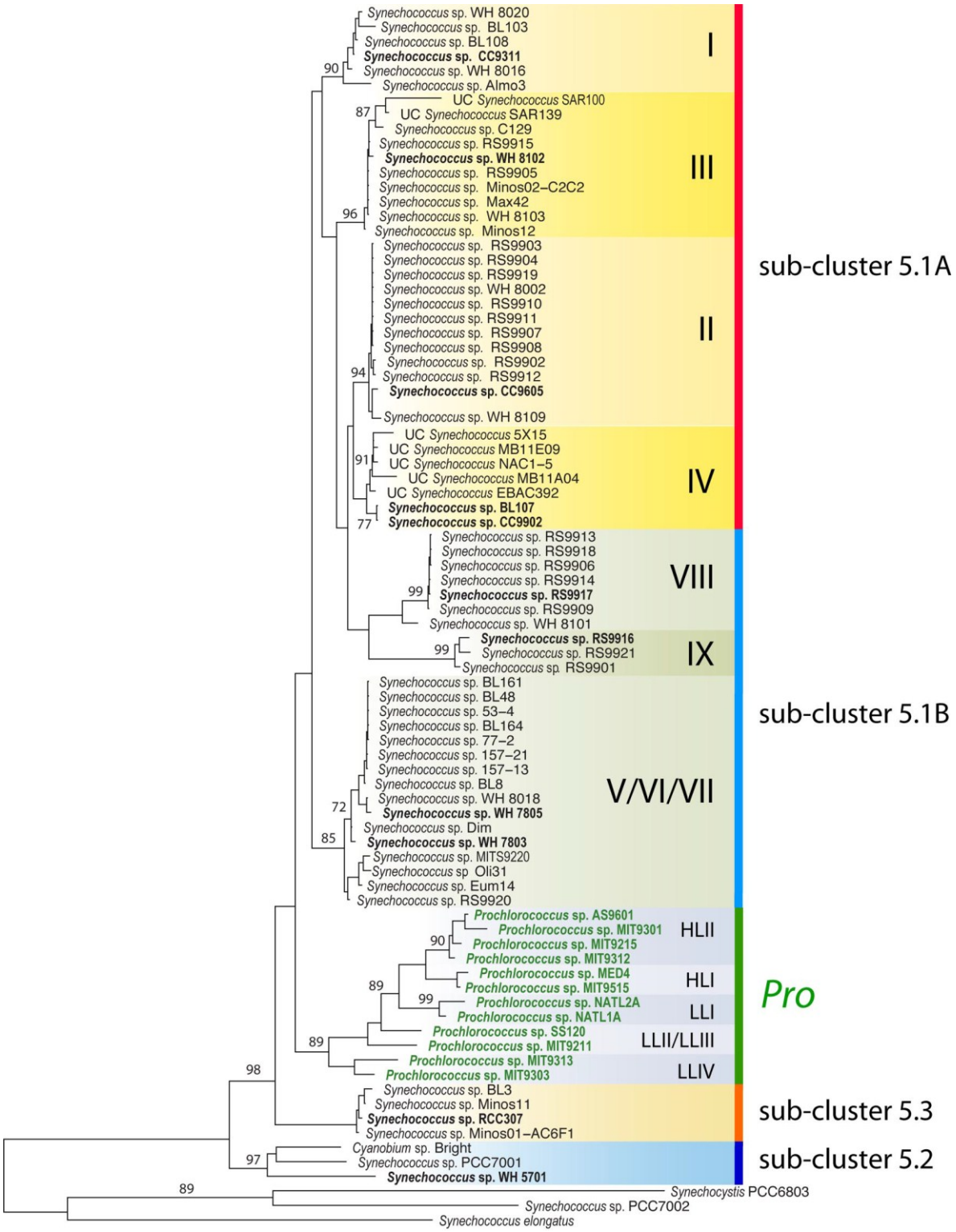


Figure 1.2: Relations phylogénétiques (Neighbour-Joining) entre les picocyanobactéries marines : *Synechococcus* et *Prochlorococcus*, basé sur des séquences du gène 16S ARNr. Seules les valeurs de bootstrap >70% sont annotées. Les souches dont le génome est séquencé sont en gras (tiré de Scanlan *et al*., 2009).

Les représentants photosynthétiques sont représentés par quatre principaux groupes (Vaulot *et al*., 2008; Massana, 2011). D'une part, les Chlorophyta qui selon la théorie endosymbiotique sont issus de l'endosymbiose primaire, et d'autre part les Haptophyta, les stramenopiles et les Cryptophyta issus de l'endosymbiose secondaire (Selosse *et al*., 1997).

Les haptophytes sont des cellules dont la taille varie entre 2 et 20 µm, et la plupart sont reconnaissables grâce à leur haptonème (appendice entre 2 flagelles). Ils forment un groupe taxonomique très important étant largement distribué dans les eaux océaniques (**Figure 1.5a**). Des études estiment que leur contribution au stock de chlorophylle est importante, variant de 25 à 50% dans la zone photique des océans (Liu *et al*., 2009; Lepère *et al*., 2009) (**Figure 1.5a**). Ils dominent aussi souvent la composante photosynthétique au sein de la fraction <3 µm constituant jusqu'à 25% de la biomasse carbonée (Cuvelier *et al*., 2010; Kirkham *et al*., 2011). Initialement, la prépondérance des haptophytes a été démontrée par la présence d'un pigment caractéristique le 19'-hexanoyloxyfucoxanthin (Andersen *et al*., 1996). Morphologiquement et phylogénétiquement, ils présentent une extrême diversité et un grand nombre de taxons reportés sont dits mixotrophes, expliquant probablement leur larges répartition et abondance (Hammer et Pitchford, 2005; Liu *et al*., 2009). Taxonomiquement, les haptophytes sont subdivisés en plusieurs ordres tels que les Prymnesiales, les Phaeocystales, les Pavlovales, et la famille des Calcihaptophycidae. Les Calcihaptophycidae semblent dominés au sein des haptophytes et sont caractérisés par un exosquelette de plaques de calcaires. On y retrouve plusieurs ordres bien connus comme les Coccolithales, les Syracosphaerales et les Isochrysidales (**Figure 1.5b, 1.5c**). Parmi les autres ordres, on retrouve des taxa écologiquement importants régulièrement reportés comme *Chrysochromulina* (Prymnesiales), *Prymnesium* (Prymnesiales), ou *Phaeocystis* (Phaeocystales).

Les chlorophytes sont majoritairement retrouvées dans les sites côtiers, alors que leur contribution au large semble plus faible (Not *et al*., 2004; Worden, 2006; Worden et Not, 2008). Ils sont souvent répertoriés dans les banques de clones, et l'ordre des Mamiellophyceae est majoritairement retrouvé avec des représentants cultivés comme *Micromonas*, *Bathycoccus*, et *Ostreococcus* (**Figure 1.6a**). Le génome de ces 3 genres a été séquencé, et des études phylogénétiques ont mis en évidence deux écotypes chez *Ostreococcus* et cinq écotypes chez *Micromonas* selon les conditions ambiantes de lumière

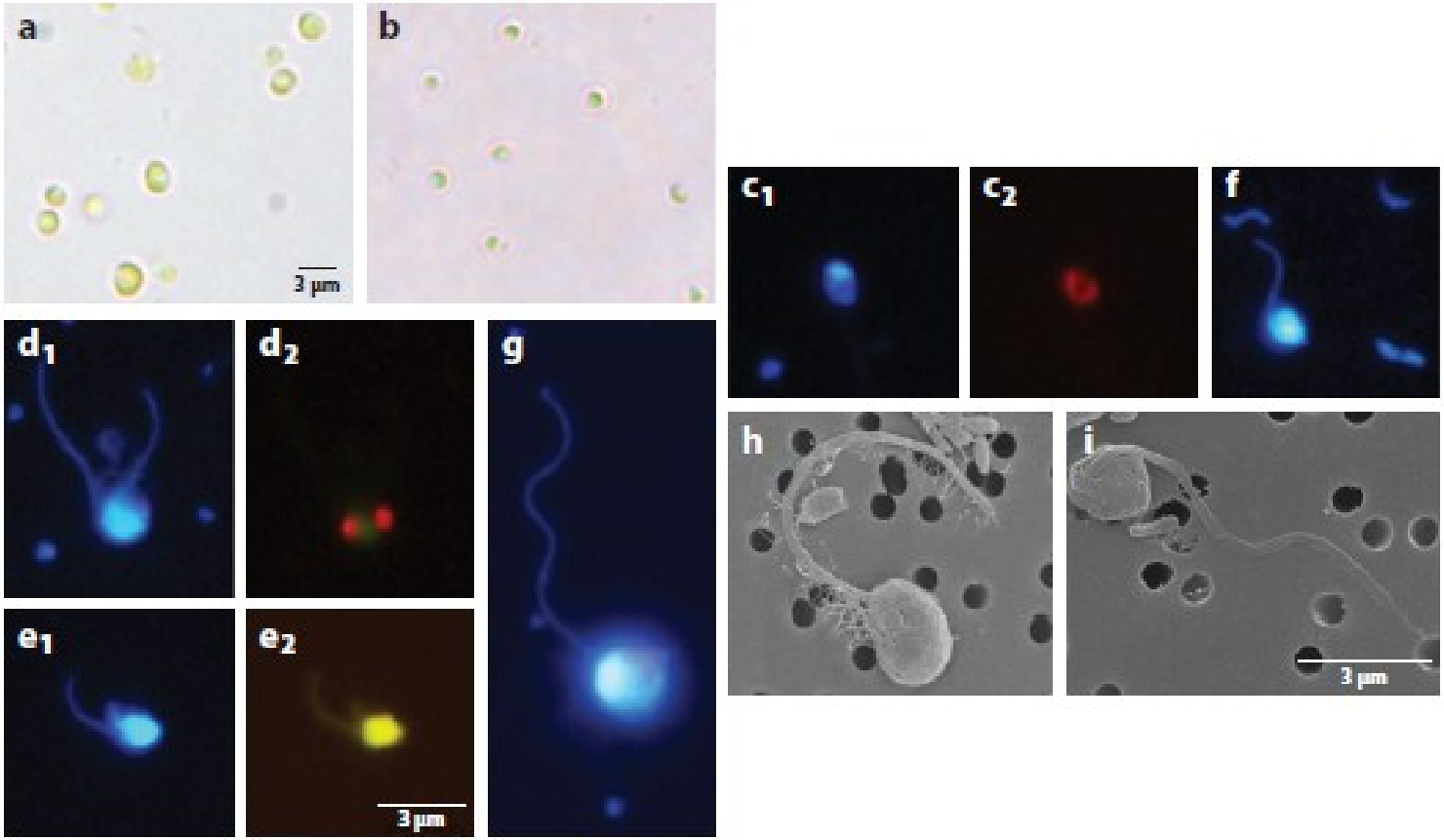


Figure 1.3: Picoeucaryotes marins observés par **(a, b)** microscopie optique, **(c, g)** microscopie à épifluorescence, et **(h, i)** microscopie électronique à balayage. Les images d'épifluorescence sont prises par excitation UV de l'ADN teinté de bleu par DAPI pour **(c1, e1, f, g)**, ou par excitation bleu de la chlorophylle autofluorescence rouge pour **(c2, e2)**. Les organismes sont **(a)** *Pelagomonas calceolata*, **(b)** *Micromonas pusilla*, **(c)** un prasinophyte phototrophe non identifié, **(d)** un haptophyte phototrophe non identifié, **(e)** un picoeucaryote hétérotrophe non identifiés, et **(f, i)** cellules picoeucaryotes non identifiées (tiré de Massana, 2011 suite aux résultats de Forn, Cabello, del Campo, et Fortu). Abréviation : DAPI, 4,6-diamidino-2-phenylindole.

ou de température (Rodriguez *et al*., 2005; Lovejoy *et al*., 2007; Worden *et al*., 2009). *Ostreococcus,* le plus petit eucaryote connu à ce jour avec une taille de 0,8 µm de diamètre, a un petit génome de 13 Mb et 8000 gènes (Courties *et al*., 1994, 1998). *Micromonas* possède un génome plus grand (20 Mb et 10000 gènes) ce qui lui permet d'avoir une plus grande flexibilité écologique (Parker *et al*., 2008; Worden *et al*., 2009). Récemment, *Bathycoccus* s'est révélé être intermédiaire avec un génome de 15Mb et 19 chromosomes (Moreau *et al*., 2012).

Parmi les straménopiles autotrophes, les études environnementales effectuées sur la fraction ≤ 3 µm identifient régulièrement des chrysophyceae, des dictyochophyceae, des pélagophyceae et des diatomées (Vaulot *et al*., 2008). Les diatomées sont facilement reconnaissables par leur frustule de silice. Les chrysophyceae et les pelagophyceae apparaissent comme une composante importante des picoeucaryotes photosynthétiques, alors que les petites diatomées et les dictyochophyceae semblent avoir une contribution moindre (Lepère *et al*., 2009; Not *et al*., 2009; Man-Aharonovich *et al*., 2010).

Les cryptophytes semblent relativement abondantes près des côtes et moins au large (Cuvelier *et al*., 2008; Not *et al*., 2008; Lepère *et al*., 2009). Ils sont facilement identifiables par microscopie ou cytométrie en flux par la présence du pigment phycobiline dans leurs chloroplastes. Cette caractéristique pigmentaire leur permet de se développer en profondeur dans des conditions de faible luminosité (Hammer *et al*., 2002).

L'avancée des techniques moléculaires a aussi grandement contribué à nous éclairer sur l'écologie et la diversité des petits hétérotrophes flagellés (Thaler and Lovejoy, 2012; Monier *et al*., 2013). Lors des études environnementales, on reporte régulièrement des parasites, des osmotrophes, et des bactérivores.

Parmi les stramenopiles hétérotrophes, les stramenopiles marins (MAST) sont souvent rapportés. Ils ont été tardivement recensés, retrouvés dans tous les océans et peuvent représenter jusqu'à 35% des hétérotrophes flagellés (Massana *et al*., 2006). Polyphylétiques, ils sont constitués de plusieurs clades phylogénétiquement distincts (Massana *et al*., 2009) (**Figure 1.6b**). Leur taille varie entre 2 et 8 µm, avec certains clades principalement retrouvés dans la fraction picoplanctonique comme les MAST-4, -7, et -11, alors que d'autres incluent des cellules pico- et nano- planctoniques comme les MAST-1, -3, et -8 (Massana *et al*., 2006, 2009; Piwosz et Pernthaler, 2010, Logares *et al*., 2012).

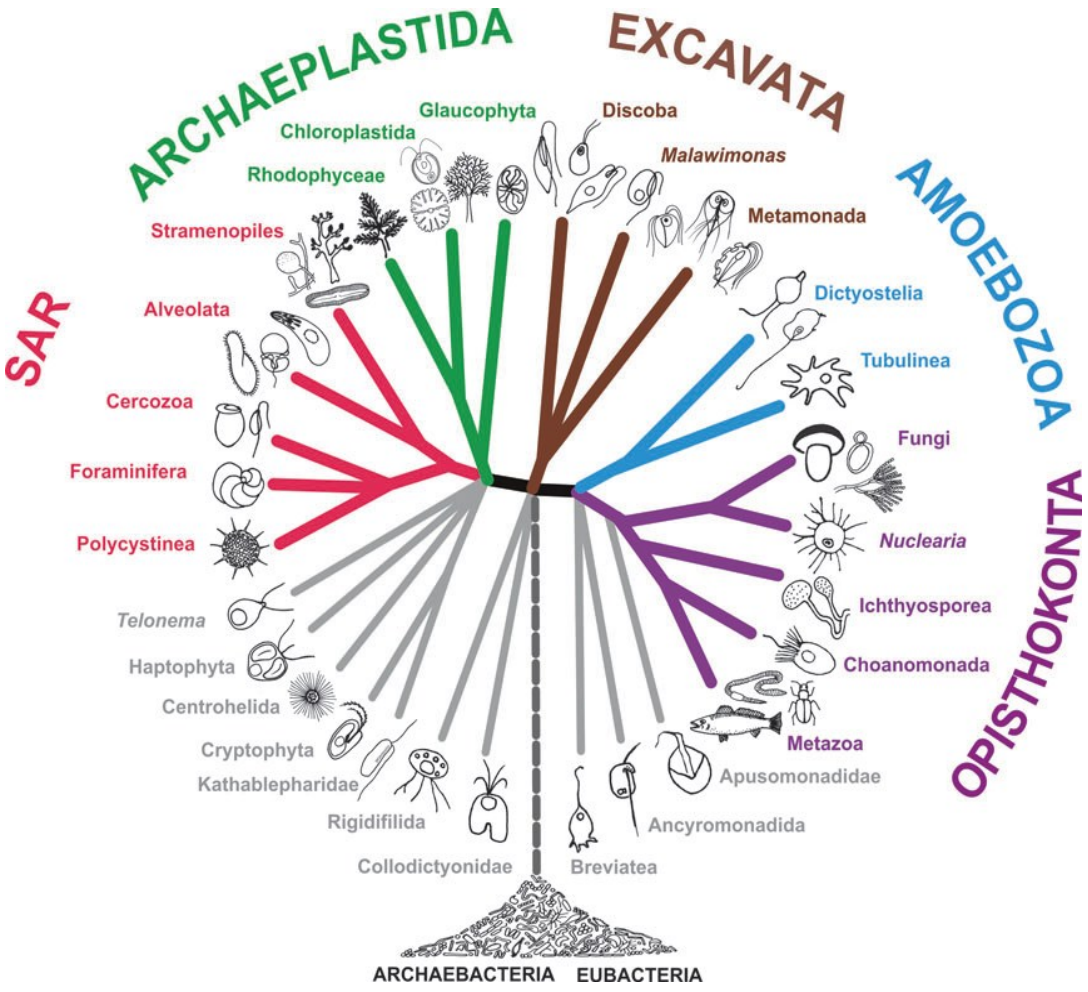


Figure 1.4: Arbre phylogénétique des eucaryotes (tiré de Adl *et al*., 2012).

La grande majorité de ces flagellés hétérotrophes sont des bactérivores (Massana *et al*., 2009), mais certains peuvent être algivores comme MAST-6 (Piwosz et Pernthaler, 2010). D'autres ont une préférence pour les cyanobactéries comme MAST-4 (Lin *et al*., 2012), et certains peuvent être des parasites (Gomez *et al*., 2011). MAST-1 semble être globalement répandu et seul MAST-3 a un représentant cultivé : *Solenicola setigera* (Gomez *et al*., 2011).

Le supergroupe des alvéolés marine (MALV) est aussi régulièrement retrouvé en abondance dans les études environnementales (Coats et Park, 2002; Guillou *et al*., 2008). Ce sont des parasites infectant différents hôtes et plusieurs sous-groupes ont été mis en évidence (groupe I à groupe V). Ils ont un spectre d'hôtes assez large incluant des dinoflagellés, ciliés, des radiolaires, ou des œufs de poissons (Guillou *et al*., 2008). Comme les virus, ces parasites réacheminent une proportion importante du carbone dans le réseau trophique, et interfèrent dans la compétition entre espèces infectant préférentiellement les espèces à forte croissance (Chambouvet *et al*., 2008).

D'autres flagellés hétérotrophes sont aussi régulièrement répertoriés dans les bandes de clones environnementaux portant sur la fraction <3 µm comme Picozoa, Telonemia, les choanoflagellés, ou les cercozoa (Rhizaria), et seront détaillés plus en détail dans les chapitres 2 et 3 de cette thèse (Monier *et al*., 2013; Thaler, 2013).

1.4 Écologie du picoplancton : caractéristique, distribution et abondance

A cause de leur très petite taille, il a été longtemps difficile d'identifier et de dénombrer les cellules picoplanctoniques. C'est dans les années 70 que l'on commence à suggérer la contribution significative des cellules picophytoplanctoniques à la production primaire océanique (Wauthy *et al*., 1967; Throndsen, 1976; Jeffrey, 1976). Leur taille réduite est une caractéristique importante dans le processus d'échange entre l'intérieur de la cellule et l'environnement extérieur. En effet, grâce à leur rapport surface/volume élevé ces petites cellules ont une acquisition plus efficace des nutriments, où leur volume va influencer leurs besoins et leur surface va influencer leur capacité d'échange.

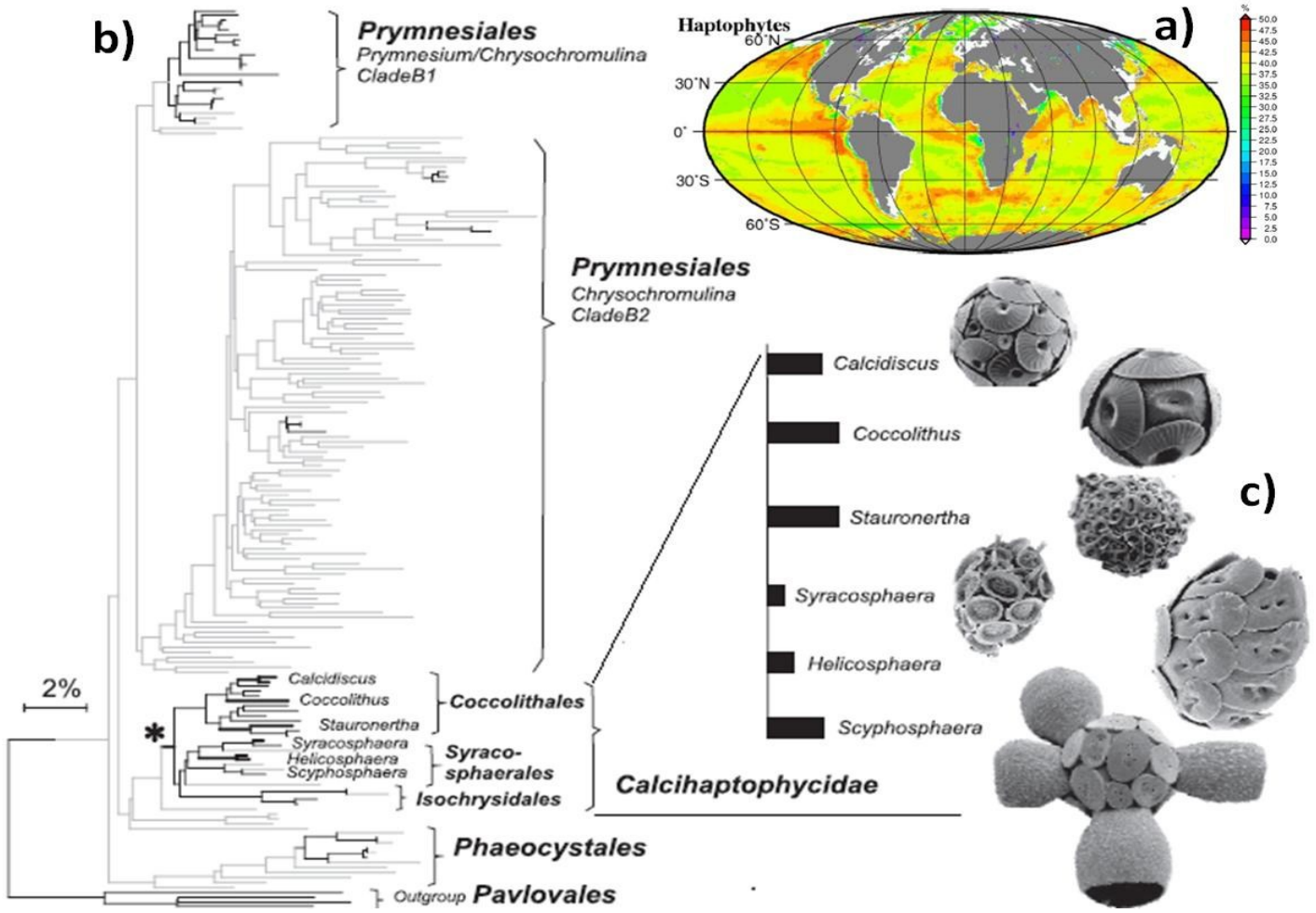


Figure 1.5: **a)** Relative contribution des haptophytes à la biomasse totale de chlorophylle *a*, basée sur les pigments accessoires dans la couche euphotique des océans lors de l'année 2000; **b)** Arbre phylogénétique des haptophytes basé sur LSU ADNr, incluant un large nombre de séquences environnementales (en gris) ou cultivées (en noir); **c)** Zoom sur la famille des Calcihaptophycidae (Tiré et modifié de Liu *et al*., 2009).

Ceci leur confère un avantage compétitif face à d'autres cellules planctoniques plus grandes, spécialement dans des conditions où les nutriments sont potentiellement limitants (Fogg, 1991; Raven, 1998; Veldhuisa *et al*., 2005). Leur petite taille a aussi une influence sur la durée de génération, les pertes dues au broutage, la sédimentation, et l'absorption lumineuse par unité de pigment (Raven, 1998). Comme les larges cellules, la croissance des cellules picoplanctoniques va subir des effets « bottom-up » via des contraintes environnementales (lumière, stratification, turbulences) et des contraintes face aux ressources (éléments nutritifs), ainsi que des effets « top-down » via la prédation ou l'infection virale.

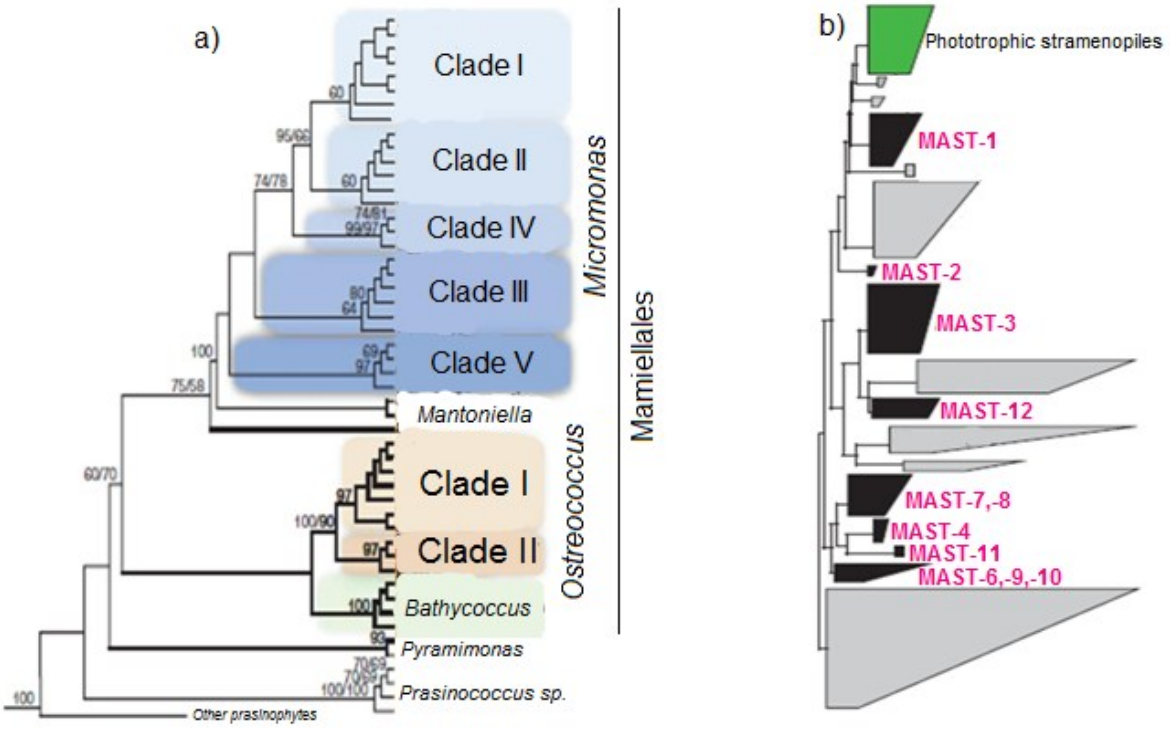


Figure 1.6: Arbres phylogénétiques basés sur des séquences du gène 18S ARNr mettant en évidence les différents clades au sein des **a)** Chlorophyta (Tiré et modifié de Worden *et al*., 2009); **b)** marines stramenopiles (tiré et modifié de Massana *et al*., 2006).

Le picoplancton est retrouvé dans de nombreux systèmes aquatiques où chaque groupe (cyanobactéries et eucaryotes) occupe une niche écologique différente. Les cyanobactéries dominent dans les eaux oligotrophes (Partensky *et al*., 1999b) où la contribution de *Prochlorococcus* et *Synechococcus* à la biomasse et production totale est importante (Agawin *et al*., 2000). *Prochlorococcus* est principalement retrouvé dans des eaux chaudes oligotrophes et reste limité dans une bande latitudinaire de 40°N-40°S (Bouman *et al*., 2006; Johnson *et al*., 2006; Martiny *et al*., 2009) (**Figure 1.7a**). Au sein de la colonne d'eau, *Prochlorococcus* peut être distribué à des profondeurs allant jusqu'à 200 mètres (**Figure 1.7c**). À l'inverse, *Synechococcus* est moins abondant, mais plus ubiquiste à l'échelle globale (Scanlan et West, 2002; Scanlan, 2003) (**Figure 1.7b**). *Synechococcus* domine dans des eaux mésotrophes à eutrophes, à des profondeurs dépassant rarement les 100 mètres (Partensky *et al*., 1999b; Worden et Not, 2008; Buitenhuis *et al*., 2012) (**Figure 1.7c**). Il y a très peu de mise en évidence de barrières de dispersion chez les picoeucaryotes qui sont ubiquistes et majoritairement distribués en zone photique (0-100 m) (**Figure 1.7c**), mais il semble que la température joue un rôle prépondérant sur leur abondance relative (Li *et al*., 2006; Rodríguez-Martínez *et al*., 2013).

Finalement, en termes d'abondance relative Buitenhuis *et al*. (2012) estime qu'au sein de la biomasse picophytoplanctonique mondiale, *Prochlorococcus* contribue à hauteur de 17 à 39%, *Synechococcus* entre 12 à 15%, et les picoeucaryotes entre 49 à 69%. Cependant, les concentrations typiques de *Prochlorococcus*, *Synechococcus* et des picoeucaryotes sont respectivement de l'ordre de 10^5, 10^4 et 10^3 cellules ml^{-1} (Campbell *et al*., 1994; Partensky *et al*. 1999b; Li, 2009). On constate ainsi que même si les picoeucaryotes ne sont pas toujours numériquement abondants, ils dominent souvent la biomasse phytoplanctonique totale et la production annuelle (Li, 1994, 1995; Worden *et al*., 2004). En effet, les picoeucaryotes qui sont légèrement plus grands que les picocyanobactéries, dominent souvent en termes de stockage de carbone et peuvent être les principaux agents de la fixation du carbone et de la respiration.

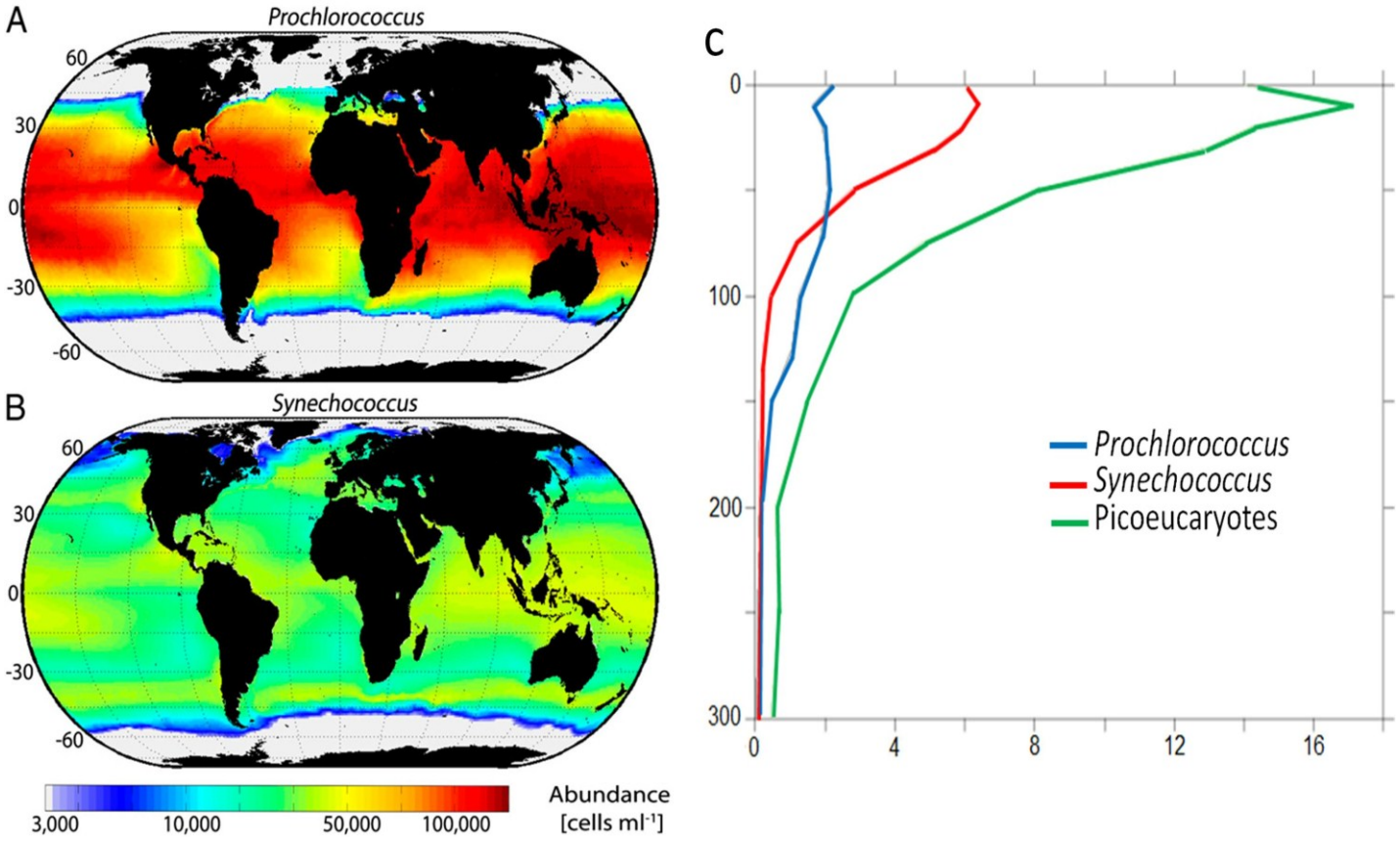


Figure 1.7: Moyenne annuelle des abondances à la surface des eaux océaniques pour **a)** *Prochlorococcus*, et **b)** *Synechococcus* (tiré de Flombaum *et al*., 2013); et **c)** moyenne de la biomasse (µg C l^{-1}, axe X) en fonction de la profondeur (axe Y) pour les 3 groupes du picophytoplancton *Prochlorococcus* en bleu, *Synechococcus* en rouge, et picoeucaryotes en vert (tiré de Buitenhuis *et al*., 2012).

1.5 Objectifs de recherche

Au regard de l'importance écologique du picoplancton au sein des écosystèmes aquatiques, il apparaît comme déterminant de pouvoir comprendre les facteurs qui régissent la répartition et la structure de ces communautés. Les connaissances sur les picoeucaryotes étant moins approfondies que pour les picocyanobactéries, cette thèse porte principalement sur les picoeucaryotes autotrophes et hétérotrophes.

Ainsi, cette thèse cherche à répondre à la question suivante : **« Comment les changements saisonniers et/ou spatiaux, dans le développement des structures hydrographiques du niveau supérieur de la colonne d'eau, influencent la croissance, la diversité et la structure des picoeucaryotes ? »**.

Pour cela, les trois objectifs généraux consistent à :

1. Estimer la diversité des picoeucaryotes et identifier les principaux taxons qui composent la communauté par la combinaison de plusieurs méthodes moléculaires.
2. Identifier les principaux changements taxonomiques dans la structure des communautés et décrire leurs répartitions dans le temps (avril, juillet, octobre) et l'espace (côte vs large; surface vs maximum de chlorophylle subsurface).
3. Déterminer les principaux facteurs physico-chimiques qui contrôlent l'abondance des taxons spécifiques et qui structurent l'ensemble de la communauté picoeucaryotes.

Ces trois objectifs reposent sur l'hypothèse principale que :
« *Loin d'être persistants dans le temps et l'espace, les picoeucaryotes aussi bien photosynthétiques qu'hétérotrophes constituent une communauté dynamique régulée par les paramètres environnementaux* ».

1.6 Méthodologies

Comme expliquée précédemment, l'avancée des techniques d'étude a beaucoup contribué à nos connaissances actuelles sur le picoplancton. Dans cette thèse, il a donc été important de combiner plusieurs techniques d'approches pour l'étude de la diversité phylogénétique des communautés picoeucaryotiques.

1.6.1 Séquençage : banques de clones vs pyroséquençage

L'outil principal utilisé au cours de cette thèse est le séquençage à travers deux techniques. Le premier dit « clonage/séquençage » à consistée à l'amplification des séquences cibles (via des amorces spécifiques) par PCR (Polymerase Chain Reaction), suivi d'un clonage par des banques de clones, et du séquençage basé sur la méthode de Sanger *et al*. (1977). Cette méthode est longue, coûteuse, et dite limitée en raison du nombre de séquences obtenues (~100 clones par banque de clone). Elle permet principalement d'accéder aux taxons abondants, et offre l'opportunité d'obtenir de longues séquences (>1000pb) qui fournissent une profondeur d'information au niveau phylogénétique (**Figure 1.8**). La seconde technique à aussi consistée à l'amplification des séquences cibles par PCR, mais suivi d'un séquençage à haut débit par la technologie Roche 454- GS-FLX Titanium (Sogin *et al*., 2006; Margulies *et al*. 2005). Cette méthode est plus rapide, relativement peu coûteuse, et permet d'explorer la diversité en profondeur grâce à un grand nombre de séquences obtenues (~5000 à 100 000 séquences/échantillons) identifiant ainsi les taxons de moindre abondance dits « rares ». Cependant les séquences obtenues sont courtes (400 et 500 pb) limitant l'information au niveau phylogénétique (**Figure 1.8**).

1.6.2 Les gènes cibles : 18 S vs *psbA*

Deux gènes ont été ciblés afin d'accéder au mieux à l'ensemble de la diversité phylogénétique des picoeucaryotes. Premièrement, la petite sous-unité ribosomale du gène 18S ARN (~1900 nucléotides) qui est un gène constitutif retrouvé chez tous les eucaryotes. Il est le constituant principal des ribosomes et est impliqué dans la traduction de l'ARN messager en protéines. Ce gène dit «universel » est régulièrement utilisé pour des études phylogénétiques en raison de ces régions hautement conservées et ces régions variables permettant de facilement distinguer les groupes ou sous-groupes taxonomiques. Cependant,

il ne permet pas de résoudre toute la complexité phylogénétique au sein des genres ou espèces (Woese *et al*., 1990; Diez *et al*., 2001; Lopez-Garcia *et al*., 2001). C'est pourquoi, dans un deuxième temps, un gène fonctionnel (*psbA*) a été ciblé permettant ainsi une comparaison des diversités obtenues. Le gène *psbA* est présent dans les chloroplastes de tous les organismes qui effectuent la photosynthèse oxygénique et permet ainsi de cibler uniquement les organismes photosynthétiques participant potentiellement à la production primaire (Zeidner *et al*, 2003; Man-Aharonovich *et al*., 2010). Le gène *psbA* code pour la protéine D1 du photosystème II (PSII). Le PSII est un ensemble formé par des protéines et des pigments interviennent dans les mécanismes de la photosynthèse en absorbant les photons de la lumière (**Figure 1.9a**). L'intensité lumineuse peut inactiver le PSII à cause des dommages engendrés sur la protéine D1. Ainsi cette protéine est constamment dégradée et resynthétisée (**Figure 1.9b**) (Aro *et al*., 1993a, 1993b; Golden, 1994; Mulo *et al*., 2012; Nickelsen et Rengstl, 2013).

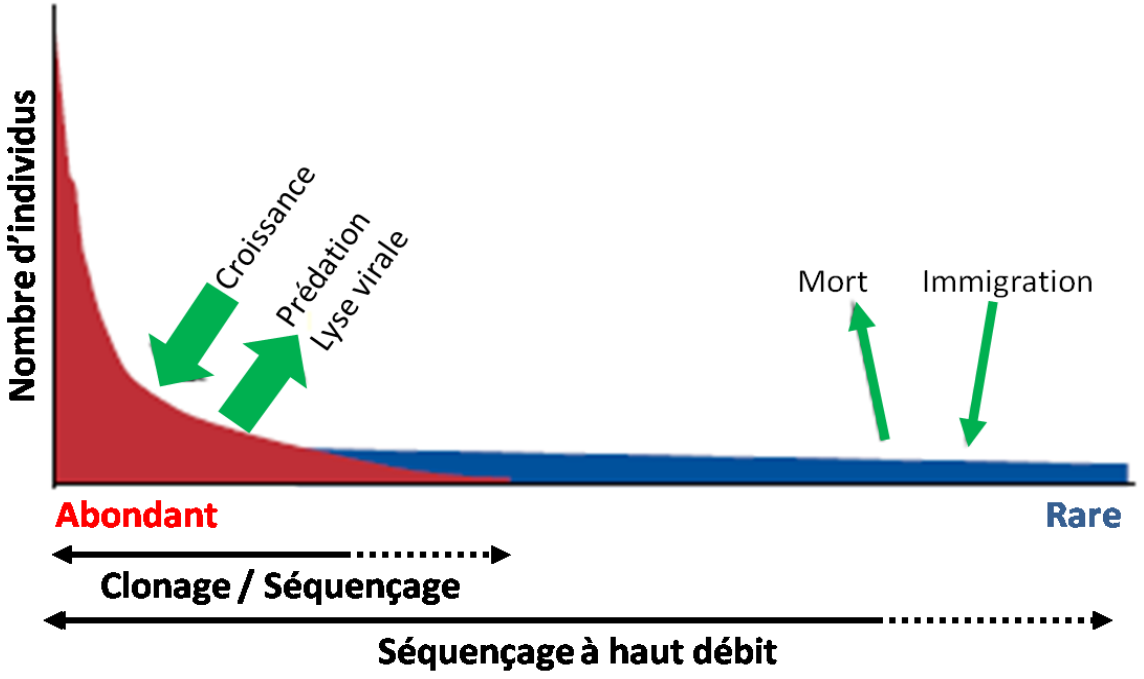


Figure 1.8: Schéma représentant les limites de la diversité accessibles par les deux techniques de séquençage utilisées au cours de cette thèse. En rouge les taxons abondants, et en bleu les taxons de plus faible abondance. Les taxons abondants et rares sont contrôlés (dans ce schéma) par des processus biotiques en vert (tiré et modifié de Pedros Alios, 2006).

1.6.3 Les acides nucléiques : ADNr vs ARNr

En addition, ADNr (pour le gène 18S) ainsi que ARNr (pour le gène 18S et *psbA*) ont été ciblés et amplifiés. ADN (acide désoxyribonucléique) constitue le support de l'information génétique. Sa grande stabilité est la principale caractéristique qui favorise son utilisation. Cependant, cela induit la détection de microorganismes qui peuvent être inactifs ou morts (Stoeck *et al*., 2007). Pour pallier à ce problème, l'ARN (acide ribonucléique) peut-être utilisé une fois converti en ADNc (ADN complémentaire). ARN est présent dans les ribosomes qui sont eux-même retrouvés en grande quantité au sein des cellules dites « métaboliquement actives » pour la production des protéines nécessaires à la croissance et/ou à la division cellulaire. L'ARN est rapidement dégradé et ainsi il est souvent caractérisé comme étant un représentant des microorganismes potentiellement actifs (Stoeck *et al*., 2007; Blazewicz *et al*., 2013).

1.6.4 Autres techniques supplémentaire

Finalement et brièvement, en complément à ces analyses moléculaires, la biomasse phytoplanctonique est estimée par la quantification de la chlorophylle *a* (Mitchell *et al*., 2002). La cytométrie en flux est utilisée pour la quantification et l'identification des bactéries, du nanoplancton, et des trois populations picophytoplanctoniques (picoeucaryotes photosynthétiques, *Synechococcus*, *Prochlorococcus*) (Li and Dickie, 2001). ,Puis, le rapport Fv/Fm mesurant l'efficience photochimique du PSII est effectué par l'utilisation du Phyto-PAM (Pulse Amplitude Modulation) afin d'estimer la capacité des cellules a convertir l'énergie lumineuse recue sous forme chimique (Houliez *et al*., 2013) (**Annexe 4**).

1.7 Sites d'étude

Le plateau néo-écossais (face aux côtes de la Nouvelle-Écosse) et le golfe du Maine (à la pointe sud de la Nouvelle-Écosse) dans l'Océan Atlantique Nord-Ouest ont été échantillonnés (**Figure 1.10**). Ces deux régions forment un écosystème très dynamique et très productif, abritant de nombreuses espèces marines (Mills, 1989). Elles sont traversées par deux principaux courants, le Labrador qui un courant froid provenant de l'Océan

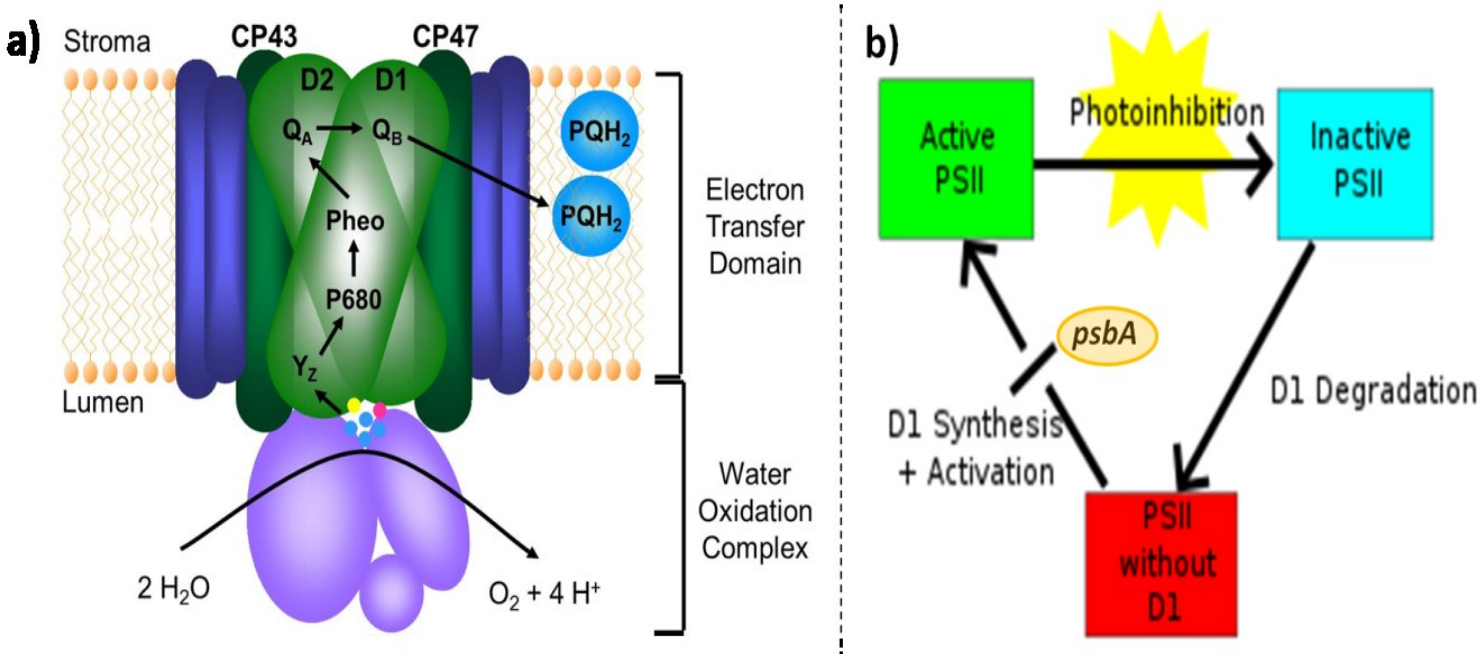


Figure 1.9: **a)** Illustration simplifiée de la structure et du fonctionnement du photosystème II chez les eucaryotes : Les pigments situés dans l'antenne, transfèrent l'énergie lumineuse jusqu'au centre réactionnel constitué d'une paire de molécules de chlorophylle *a* P680 (absorbant la lumière de longueur d'onde 680 nm). L'énergie accumulée désactive par voie photochimque le P_{680} qui céde un électron à la phéophytine. Il y a ensuite séparation des charges (P_{680}^+ - $pheo^-$). L'électron cédé à la phéophytine est ensuite transféré via le centre Fe-S, jusqu'aux quinones QA et QB qui stockent les électrons reçus un par un et les transfèrent sur l'accepteur suivant présent dans la membrane à l'extérieur du PSII : la plastoquinone. La chlorophylle (P_{680}^+) récupère alors un électron provenant de l'oxydation de l'eau. Les électrons nécessaires à la réduction de la chlorophylle P_{680} sont fournis par oxydation de l'eau qui produit les protons (H^+) pour la réduction du NADP, selon la réaction $H_2O \rightarrow 2H^+ + O^- + 2e^-$. **b)** Schéma de la dégradation et de la re-synthèse de la protéine D1 suite à une photoinhibition et inactivation du photosystème II (Aro *et al*., 1993b).

Arctique et le Gulf Stream qui est un courant chaud qui prend naissance dans le golfe du Mexique (Loder *et al*., 1998). La rencontre des différentes masses d'eau engendre ainsi différentes conditions environnementales rencontrées sur des distances géographiques relativement courtes.

De nombreux programmes de surveillance ont été menés sur le long terme dans ces deux régions qui visent à collecter et analyser l'information physique, chimique et biologique (voir partie « Avant-propos »). Ainsi à travers le programme AZMP, on apprend que sur le plateau néo-écossais, les conditions de température et de salinité en 2009 (dates de nos échantillonnages pour le chapitre 2) étaient proches de la normale (**Figure 1.11a**) (Johnson *et al*., 2012). Les conditions chimiques étant très variables dans l'espace et le temps cela rend difficile la détection de changements à court terme. Cependant, sur le long terme on observe une diminution dans le temps des concentrations sous la surface en silicates, en phosphates et en nitrates, ces dernières ayant diminué plus rapidement (**Figure 1.11b**) (Yeats *et al*., 2010). Au niveau biologique, les changements à long terme du plancton sur le plateau néo-écossais et le golfe du Maine ont été principalement acquis grâce au programme *Continuous Plankton Recorder* surveys. Les résultats publiés concernent le zooplancton, et d'autres études portent sur la phénologie des efflorescences phytoplanctoniques printanières et automnales (Greenan *et al*., 2008; Song *et al*., 2011), le cycle annuel de la concentration des cellules phytoplanctoniques qui est positivement corrélé au cycle annuel de température (Li, 1998; Li *et al*., 2006), ou encore sur la composition des espèces phytoplanctoniques où on observe une dominance des diatomées au printemps et des dinoflagellés surtout en automne (Leterme et Pingree, 2007; Head et Pépin, 2010). Sur le plateau néo-écossais, la biomasse phytoplanctonique semble suivre un cycle lié à la température, avec un maximum en automne (**Figure 1.11c**). Les études portant sur le picophytoplancton ont principalement été effectuées par cytométrie en flux et révèlent une dominance de *Synechoccocus* surtout en automne et la présence des picoeucaryotes tout au long de l'année (Li, 1998, 2009; Li and Dickie 2001). Ainsi, l'écologie du microplancton (de 20 à 200 μm) et dans une moindre mesure du large nanoplancton (de 5 à 20 μm) est bien décrite dans ces deux régions, cependant moins est connu sur le picoplancton (de 0.2 à 2 μm) et le petit nanoplancton (de 2 à 5 μm) (Li *et al*., 2011).

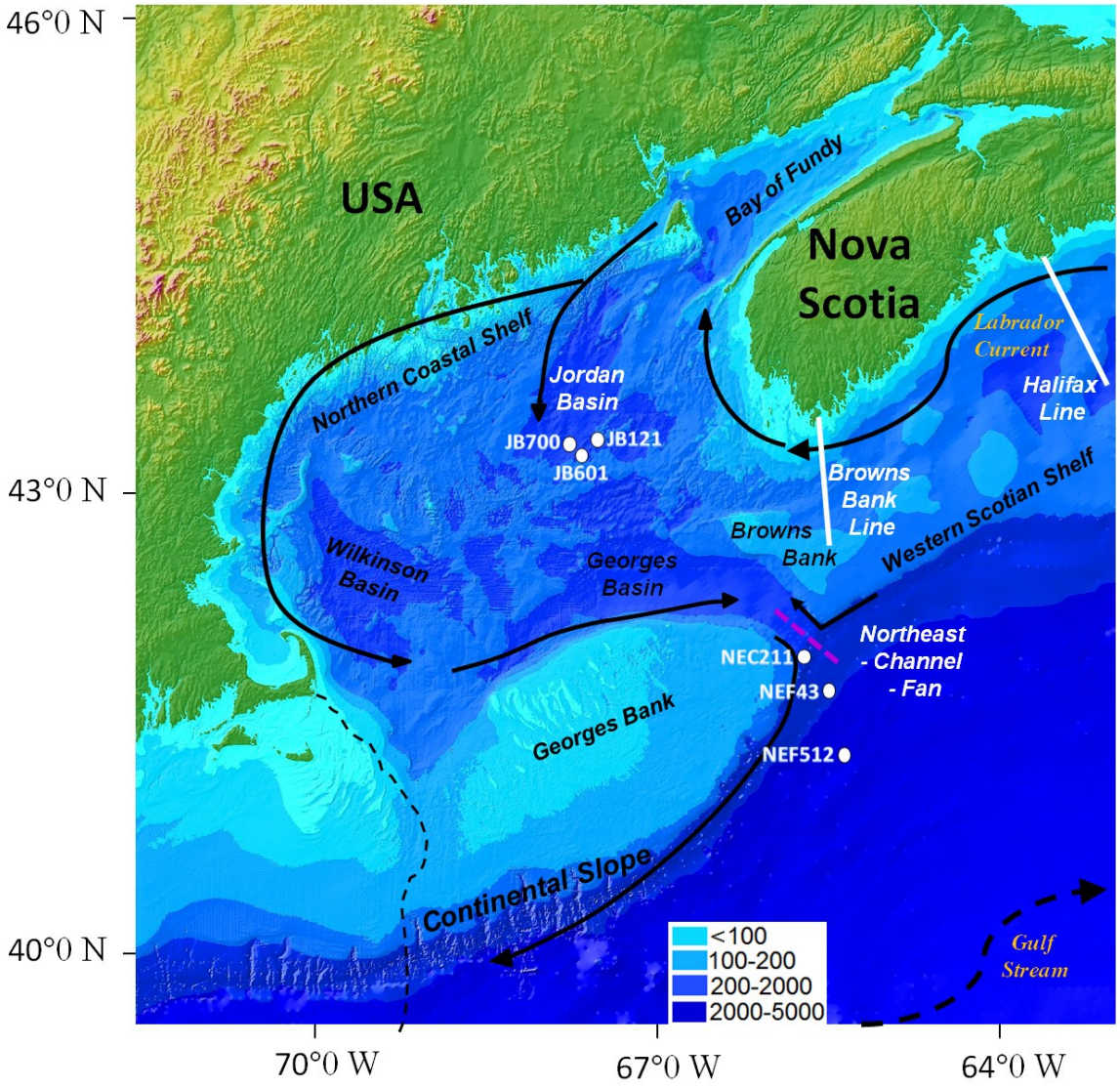


Figure 1.10: Carte représentant les deux sites d'échantillonnages de cette thèse situés dans l'Océan Atlantiques Nord-Ouest: le golfe du Maine (3 stations dans Jordan Basin et 3 stations dans le North East Channel et Fan), et le plateau Néo-Écossais (2 transects : Halifax Line et Browns Bank Line).

1.8 Plan de thèse

Cette thèse est présentée sous forme d'articles scientifiques correspondants aux chapitres 2, 3 et 4.

Chapitre 2: Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean.

(Publié dans : ***Journal Plankton Research, decembre 2013, doi: 10.1093/plankt/fbt123)***

Cet article a pour objectif d'avoir un premier aperçu de la diversité spatiale et temporelle des picoeucaryotes qui reste peut connue dans les régions d'études. Ainsi des échantillons sont prélevés sur le plateau néo-écossais à des profondeurs et des périodes d'intérêt écologique : en surface et au maximum de chlorophylle, en avril et octobre 2009. La diversité phylogénétique est étudiée par clonage et séquençage Sanger de la petite sous-unité ribosomale ARN du gène 18S. Les majeurs changements d'abondance et de taxonomie entre les deux saisons/profondeurs sont déterminés, et les principaux facteurs environnementaux qui conduisent ces divergences identifiés. De plus, cette étude permet d'obtenir des séquences « pleines tailles » qui pourront servir de références pour des études phylogénétiques futures (**Figure 1.12**).

Chapitre 3: Tracking environmental selection of marine microbial eukaryotes using 18S rRNA genes and 18S rRNA in the gulf of Maine.

(Destiné à être soumis à : ***ISME journal*****)**

Le chapitre 2 a offert un aperçu des principaux changements taxonomiques observés en avril et octobre 2009. Cependant, l'étude reportait les taxons dominants et n'offrait aucune information sur l'état des cellules identifiées ciblant ADNr (mortes, vivantes, dormance…). Dans ce chapitre 3, la diversité des picoeucaryotes est étudiée .à partir de même gène (18S), mais optimiser grâce à l'utilisation du séquençage à haut débit. De plus, afin de caractériser la communauté dite « potentiellement active » qui dépend des facteurs environnementaux, ARN (ADNc) est ciblé et est comparé à l'ADN. Ici, seule la diversité dans l'espace est étudiée, par échantillonnage de la surface et du maximum de chlorophylle de deux sites aux propriétés hydrogéographiques très différentes du golfe du Maine en juillet 2010. Ainsi, en comparant la structure des communautés et la diversité phylogénétique inférée par l'ADN

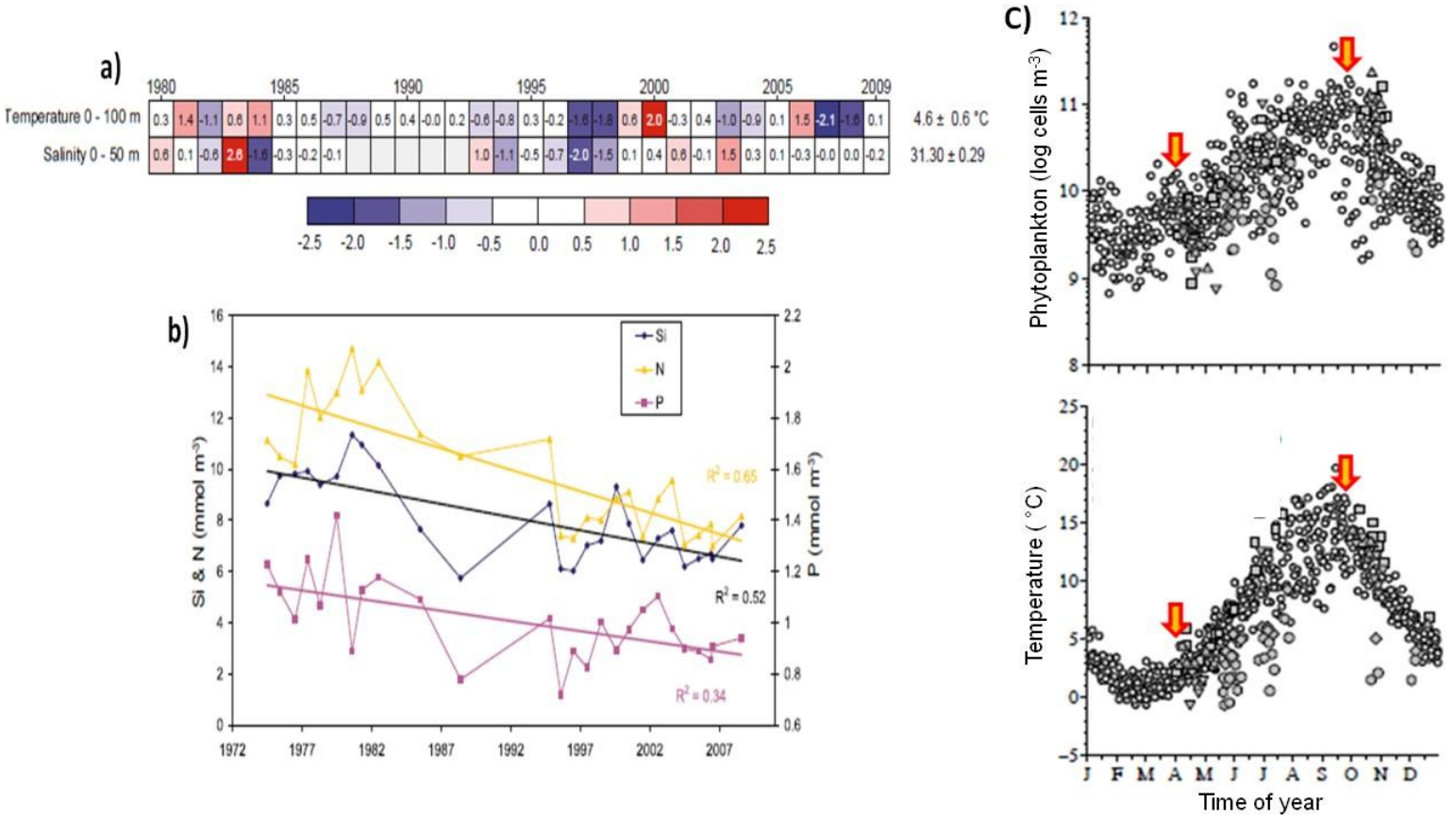


Figure 1.11: **a)** Séries temporelles de 1980 à 2009 de la température et de la salinité d'une station fixe du plateau Néo-Écossais (Halifax 2). Les cellules indiquent : une donnée manquante (gris), valeur entre 0.5 écart-type de la moyenne à long terme (blanc), conditions au-dessus de la normale (rouge), ou en-dessous de la normale (bleu). Les chiffres à l'intérieur des cellules sont les différences par rapport à la moyenne à long terme divisées par l'écart-type. Les moyennes et écarts-types sont présentés à droite de la figure. **b)** Tendances temporelles de 1972 à 2009 des silicates, phosphates, et nitrates (60-100 m de profondeur) dans la partie centrale du plateau Néo-Écossais (Tiré du Yeats *et al*., 2010). **c)** Cycle annuel pluriannuel de l'abondance du phytoplancton et de la température issues de plusieurs stations du programme AZMP, où les flèches représentent nos deux mois d'échantillonnages en 2009 dans la même région (Li *et al*., 2006).

et l'ARN, le lien entre l'activité métabolique des microorganismes et un environnement donné permettra de mieux estimer les paramètres qui définissent les stratégies de vie des picoeucaryotes (**Figure 1.12**).

Chapitre 4: Mapping diversity of picophytoplankton revealed by photosystem-II *psbA* transcripts from the gulf of Maine.

Contrairement aux chapitres 2 et 3 qui visent la diversité des picoeucaryotes autotrophes et hétérotrophes, ici l'attention s'est portée sur la diversité phylogénétique des cellules picophytoplanctoniques participant à la production primaire. L'objectif est de déterminer taxonomiquement le plus grand nombre de taxons photosynthétiques jusqu'à l'espèce. Ainsi, seuls les picoeucaryotes photosynthétiques sont étudiés conjointement aux picocyanobactéries (*Synechococcus* et *Prochlorococcus*). Les mêmes échantillons du chapitre 3 ont été utilisés, mais la méthode est modifiée. La diversité est estimée en utilisant le clonage et le séquençage Sanger des ARN transcrits du gène *psbA* (voir partie 1.5 : méthodologies). Une comparaison entre la structure phylogénétique des picocyanobactéries (connues pour avoir des écotypes) et des picoeucaryotes donnera un éclaircissement sur la flexibilité d'adaptation des picoeucaryotes photosynthétiques (**Figure 1.12**).

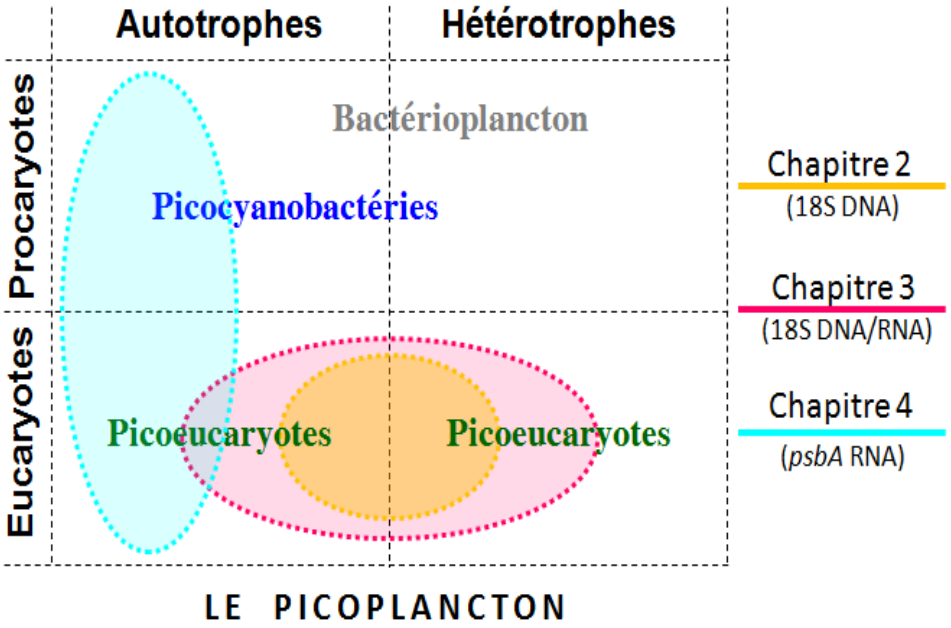


Figure 1.12: Schéma récapitulatif de la structure de la thèse.

Chapitre 2 - Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean

Résumé

Le phytoplancton sur le plateau néo-écossais (SS) à été intensivement étudié par la microscopie et la cytométrie en flux (FCM). Ces outils ont montrés que les diatomées dominent les floraisons printanières, alors que les communautés d'été et d'automne sont dominées par la cyanobactérie *Synechococcus*. De plus, les études FCM ont montrées que les picoeucaryotes (non diatomées) sont présents tout au long de l'année. Cependant, la diversité et le caractère saisonnier de ces petites cellules dont les espèces sont souvent difficiles à identifier, ne sont pas bien connus. Ici, nous avons étudié l'identité taxonomique des communautés picoeucaryotes en utilisant le clonage et le séquençage du gène de l'ARN 18S. Les échantillons sont prélevés en avril et octobre, le long de deux transects allant de la côte au large, en surface et au maximum de chlorophylle subsurface. Nos résultats mettent en évidence de nouveaux taxons non rapportés précédemment dans ce système bien étudié. Comme par exemple, *Phaeocystis* spp. retrouvés en avril, ou encore le complexe Bolido-Parmales identifié aussi en avril au sein des maximums d'abondance de picoeucaryotes déterminé par FCM. Les communautés d'octobre étaient phylogénétiquement diverses et les chlorophytes étaient les plus répandues. Parmi les taxons hétérotrophes, les choanoflagellés ainsi que deux clades incultivés de straménopiles marins (MAST) : MAST-7 et MAST- 1 étaient caractéristiques d'avril. En revanche, MAST -4 était plus répandu en Octobre. Il y avait aussi des différences entre la diversité phylogénétique estimée prés des côtes et au large, où les communautés d'avril les plus au large étaient plus semblables aux communautés d'octobre, reflétant en avril une communauté post-efflorescence au large.

Abstract

Phytoplankton on the Scotian Shelf (SS) are well studied using microscopy and flow cytometry (FCM). These tools provide a picture of a diatom dominated spring bloom, and summer-autumn communities dominated by the cyanobacterium *Synechococcus*. Flow cytometry (FCM) studies have shown that non-diatom microbial eukaryotes are present throughout the year, but the diversity and seasonality of these often smaller difficult to identify species is not well known. We investigated the taxonomic identity of surface and sub-chlorophyll maxima communities, using cloning and sequencing of the 18S rRNA gene, in April and October, along two inshore to offshore transects. Here we report new taxonomic records and evidence of novel taxa in this well studied system. For example, during April, previously unreported *Phaeocystis* spp. were recovered and picophytoplankton 'patches' indentified from FCM were marked by the Bolido-Parmales complex. The October communities were phylogenetically diverse with chlorophytes more prevalent. Among heterotrophic taxa, choanoflagellates were characteristic of April along with two clades of uncultivated marine stramenopiles (MAST) clades; MAST-7 and MAST-1. In contrast, MAST-4 was more prevalent in October. There were also onshore versus offshore differences, with the April offshore communities more similar to October communities, reflecting nutrient draw-down and a post-bloom community offshore in April.

2.1 Introduction

The Scotian Shelf (SS) within the Northwest Atlantic Shelves Province (Longhurst, 1998) represents a highly productive region of the Northwest Atlantic Ocean with seasonal shelf-scale circulation. This area is under the influence of the cold southward inflow of surface Labrador Sea Water (LSW) and the warm northward flow of the Gulf Stream (Loder et al., 1998). Since 1998, the Atlantic Zone Monitoring Program (AZMP) has explored climate and ecosystem changes in this region of the Northwest Atlantic (Therriault et al., 1998). Compared to many regions, much is known about the relationship between physical and biological factors (Sherman et al., 1996), the phenology of phytoplankton blooms (Greenan et al., 2008; Song et al., 2011), the annual cycle of phytoplankton cell concentrations (Li, 1998; Li et al., 2006) and the species composition of microphytoplankton (Leterme and Pingree, 2007; Head and Pepin, 2010).

Based on microscopy, the SS spring bloom is dominated by large phytoplankton, especially diatoms. The bloom begins when the nutrient-rich water column is thermally stabilized, supporting high levels of biomass and production in the euphotic zone. This is followed by upper euphotic zone nutrient depletion and much lower summer phytoplankton concentrations that are typically made up of mostly unidentified flagellates maintained by regenerated nutrients (Levasseur et al., 1984; Lochte et al., 1993; Longhurst, 1995; Mousseau et al., 1996). In addition, flow cytometry (FCM) studies have highlighted the importance of *Synechococcus* in this region, with maximum concentrations occurring near the autumnal equinox (Li, 1998; Li and Dickie, 2001). FCM has also revealed the year round occurrence of nanophytoplankton, nominally 2-20 μm, and photosynthetic picoeukaryotes (PPE) defined as 0.8 – 2.0 μm, on the SS (Li, 2009).

PPE are phylogenetically diverse (Worden et al., 2004; Vaulot et al., 2008), usually more abundant at latitudes greater than 30° (Buitenhuis et al, 2012), and because they are slightly larger than *Synechococcus,* they can dominate in terms of standing carbon stocks (Li, 1994; Worden et al., 2004). Many PPE and photosynthetic nanoflagellates are also mixotrophic (carry out photosynthesis and graze on other microbes). Other pico- and nano-size microbial eukaryotes are heterotrophic and since they lack chlorophyll are not routinely counted using FCM based on auto-fluorescence from chlorophyll. In general, pico- or nanoflagellates are rarely identified in routine microscopy surveys since correct

taxonomic assignation often requires examining live swimming cells or electron microscopy (Bérard-Therriault et al., 1999). Recently, Li and colleagues (Li et al., 2011) collated the taxonomic records of plankton throughout the wider Gulf of Maine and SS region, and reported that while 386 diatoms and 151 dinoflagellates species were recorded, there were only 72 taxa belonging to other marine microbial eukaryotes, with few records of heterotrophs.

Over the last two decades molecular techniques have highlighted the diversity of marine microbial eukaryotes (Vaulot et al., 2008; Caron et al., 2012), including cryptic species (Ellegaard et al., 2008) implying an incomplete record of taxa even in well studied regions. In addition, environmental surveys have revealed uncultivated lineages at many taxonomic levels (Lopez-Garcia et al., 2001) and there has been some success linking enviromental sequences to trophic function (Massana et al., 2004; Yoon et al., 2011; Guillou et al., 2008). Overall, this has resulted in new perspectives on carbon, nutrient and energy cycles in marine systems (Steele et al., 2011). However, despite the contribution of the SS system to our understanding of North Atlantic phytoplankton dynamics there have been no extensive molecular surveys of marine microbial eukaryotes carried out to date (Li et al., 2011).

The aim of this study was to apply cloning and Sanger sequencing of the 18S rRNA gene as a tool to indentify members of the phototrophic and mixotrophic communities enumerated by FCM as well as heterotrophic microbial eukaryotes that have been little studied to date. We sampled along the Halifax Line (HL) and the Browns Bank Line (BBL), which are two AZMP transects from the Nova Scotian coast to the continental slope (Fig. 1). Samples were collected in April, when chlorophyll concentrations along the coast were consistent with a spring phytoplankton bloom, and in October 2009 when chlorophyll and nutrient concentrations were low in the upper mixed layer. We firstly applied flow cytometry (FCM) to document the relative contribution of photosynthetic picoeukaryotes (PPE), *Synechococcus* and *Prochlorococcus* to picophytoplankton biomass, and to identify stations and depths to be targets for cloning and sequencing. The goal was to select samples with the potential to have different biological characteristics within the euphotic zone and uncover cryptic microbial eukaryotic diversity using molecular sequencing technology. We opted for traditional cloning and Sanger sequencing to increase the number of reference

sequences from this region that, for example, could be used for future studies using high throughput sequencing. The longer sequences also have the advantage of providing a robust basis for phylogenic analysis of key planktonic groups. The targeted samples (surface and the subsurface chlorophyll maximum) from sites with different oceanographic characteristics also provided an opportunity to investigate environmental factors that promote the occurrence of specific marine microbial eukaryote taxa or communities.

2.2 Methods

2.2.1 Study sites and general sampling

Data and water samples were collected along the Scotian Shelf in April and October 2009 in conjunction with the Atlantic Zone Monitoring Program (AZMP) of the Canadian Department of Fisheries and Oceans (Johnson et al., 2012) aboard the *CCGS Hudson*. Additional regional and annual oceanographic context for the local and seasonal aspects of this study are given in 'The Atlantic Zone Monitoring Program report' for 2009 (Johnson et al., 2012). Conductivity Temperature and Depth (CTD) profiles were collected from 7 stations along the Browns Bank Line (BBL) and 7 stations along the Halifax Line (HL) (**Figure 2.1**). In addition to the CTD data (SBE-911 CTD; Sea-Bird Inc., Bellevue, WA, USA), oxygen (Seabird Model SBE-43 dissolved oxygen sensor), *in situ* chlorophyll fluorescence (Chelsea Aquatracka MK III), and relative nitrate from an *in situ* ultraviolet spectrophotometer (ISUS; Satlantic) were acquired on the downcast. Discrete water samples were collected on the up-casts directly from 12-L Niskin bottles mounted on the Rosette. Samples for nutrients and FCM were collected every 10 m from the surface to 100 m, and at the subsurface chlorophyll maximum (SCM) determined by the chlorophyll florescence trace on the downcast. Nutrient samples were collected directly from the Niskin bottles and immediately frozen at -20°C. Water samples (2 mL) for FCM analyses were fixed in 1% paraformaldehyde for 15 min at room temperature, then quick frozen in liquid nitrogen, and stored at -80˚C in cryogenic vials (Marie et al., 2005). Samples for extracted chlorophyll *a* (Chl *a*) and DNA were collected from 3 m (surface) and the SCM. Samples for total Chl *a* were filtered directly onto GF/F filters (Whatman, GE Healthcare). The Chl *a* small fraction was pre-filtered through 3 μm pore-size polycarbonate (PC) filters (AMD

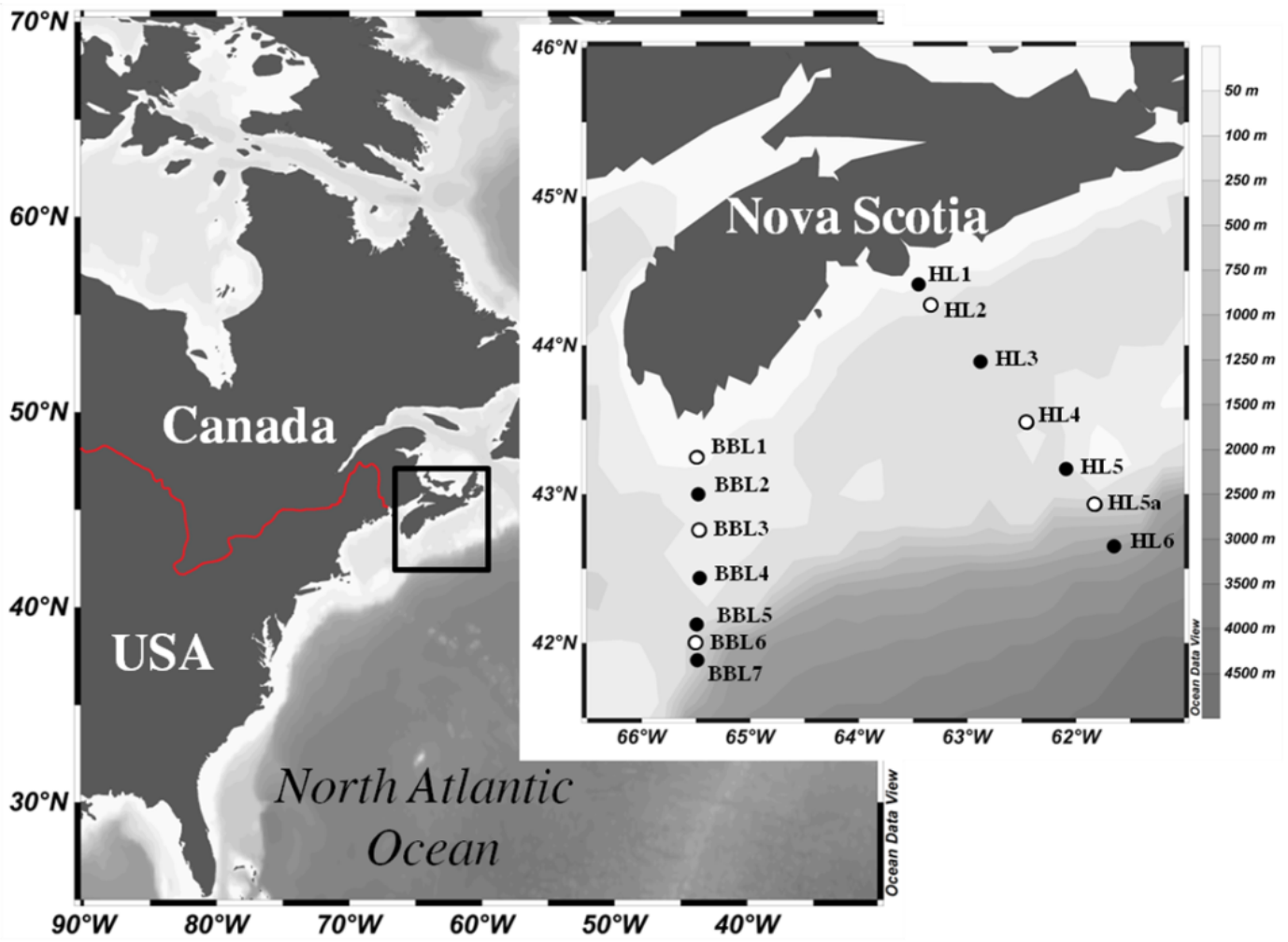


Figure 2.1: Sampling sites along the Scotian Shelf transects; Halifax Line (HL) from HL1 (44.41°N, -63.45°W) to HL6 (42.93°N, -61.82°W) and Browns Bank Line (BBL) from BBL1 (43.25°N, -65.47°W) to BBL7 (41.88°N, -65.49°W). Transects were carried out in April and October 2009. Stations used for clone libraries are indicated by an open circle.

Manufacturing) and then onto GF/F filters. The filters were stored at -80˚C until analysis. Samples for DNA were collected from Niskin bottles into clean carboys and 6 L was sequentially filtered through a 50 µm nylon mesh, a 47 mm diameter 3 µm pore size PC filter, and finally through a 0.2 µm pore size Sterivex unit (Millipore Canada Ltd, Mississauga, ON, Canada). Buffer was added to the Sterivex units (1.8 mL of 40 mM EDTA, 50 mM Tris pH 8.3, and 0.75 M sucrose) and samples were stored at -80˚C until nucleic acid extraction.

2.2.2 Laboratory protocols

Concentrations of nutrients: nitrate and nitrite (referred to as NO_3^-), phosphate (PO_4^{3-}) and silicate (SiO_4^{4-}) were analyzed by colorimetric techniques using a segmented-flow autoanalyzer (Technicon AutoAnalyzer II) at the Bedford Institute of Oceanography (Mitchell et al., 2002).

Filters for total Chl *a* (in April and October) were extracted in 90% acetone (Mitchell et al., 2002) determined by fluorescence on a Turner Design Model 10 fluorometer, before and after acidification. The small size fraction Chl *a* (≤3µm) in April samples were extracted in 95% ethanol at 70˚C for 5 min (Nusch, 1980) and concentrations determined with a spectrofluorometer (Varian Cary Eclipse). In October, small Chl *a* concentrations were determined by high performance liquid chromatography (HPLC) (Roy et al., 1996) only from the SCM samples.

FCM samples were run at Bedford Institute of Oceanography on a FACSort Flow cytometer (Becton Dickinson) as described in Li and Dickie (Li and Dickie, 2001). Archaea and heterotrophic bacteria cell counts were combined since they cannot be distinguished and are referred to as bacteria. Four groups of autotrophs (cells with autofluorescence) were distinguished: *Prochlorococcus*, *Synechococcus*, photosynthetic picoeukaryotes (PPE), nominally 0.8 to 2.0 µm and nanophytoplankton (2-20 µm). We compared PPE to picocyanobacteria in terms of average carbon content by converting cell abundances to estimated carbon biomass (C) for the three picophytoplankton categories using the suggested averaged conversion factors from Buitenhuis and colleagues (Buitenhuis et al., 2012). The values used were: 36 fg C $cell^{-1}$ for *Prochlorococcus*, 255 fg C $cell^{-1}$ for *Synechococcus*, and 2590 fg C $cell^{-1}$ for PPE.

DNA was extracted from cells retained in the Sterivex unit (0.2 – 3 μm size fraction) by a salt-extraction method (Aljanabi and Martinez, 1997), which was modified to include lysozyme, proteinase K and SDS steps (Harding et al., 2011). Six stations in April and four stations in October were selected for clone library construction. Surface and SCM libraries were from the two ends of transects carried out in April and October, with additional April libraries from a PPE 'patch' detected by FCM along both of the transects. Clone libraries were constructed targeting the small subunit 18S ribosomal RNA gene. The eukaryotic specific forward primer NSF 4/18 (5'-CTG GTT GAT YCT GCC AGT-3') (Hendriks et al., 1989) and reverse primer Euk R (5'-TGA TCC TTC TGC AGG TTC ACC TAC-3') (Medlin et al., 1988) were used to amplify nearly full-length eukaryote 18S rRNA genes using the iCycler™ Thermal Cycler (Bio-Rad Laboratories, Inc., CA, USA) following standard protocols (Potvin and Lovejoy, 2009). The inserts were amplified by PCR with vector-targeted primers M13F and M13R. Resulting PCR products, 96 per clone library, were then sequenced in both directions using the original forward and reverse primers at Service de séquençage et génotypage du Centre Hospitalier de l'Université Laval (CHUL) using an ABI 3730xl system (Applied BioSystems), which included a purification step.

2.2.3 Sequence and phylogenetic analyses

Reads <600 nucleotides (nt) were discarded. The remaining forward and reverse reads were assembled using SeqMan™II software version 5.03 (DNASTAR, Inc) with an overlap of ca. 200 nt and then checked visually using BioEdit version 7.1.3.0 (Hall, 1999). Chimeras were detected and removed using UCHIME (Edgar et al., 2011). Sequences were aligned using the Silva eukaryotic alignment as a template (Pruesse et al., 2007) and manually inspected. Initial taxonomic assignments at the level of ≥98% of similarity were carried out using Mothur v1.21.1 (Schloss et al., 2009) against our in-house curated reference database of marine eukaryotes based on NCBI taxonomy (Comeau et al., 2011). Sequences corresponding to multicellular metazoa accounted for 11.2% of the sequences from the 20 clone libraries (104 sequences) were excluded from further analysis. Overall we retained 824 sequences with a range of 950 to 1700 nt.

In a first step, to highlighted overall diversity and differences between months, a global Maximum Likelihood (ML) tree based on the GTR model was constructed using 704 sequences ≥1535nt (94% of the retained sequences from above) in our dataset using MEGA version 5.1 (Tamura et al., 2011). In a second step, to investigate the taxonomic composition of eukaryotes communities and their similarities, we conducted phylogenetic diversity analyses using weighted UniFrac distances metrics (Lozupone and Knight, 2005; Lozupone et al., 2006) for clustering samples based on phylogenetic lineages, to account for the relative proportion of lineages in the different samples. UniFrac distances were calculated based on the 824 sequences and a phylogenetic tree constructed based on the Mothur/Silva alignment using FastTree v2.1 (Price et al., 2010) in 'accurate mode' (-mlacc 2-slownni) with the general time reversible (GTR) model. Finally, to facilitate comparisons with previous records, we constructed reference phylogenetic trees for (1) the photosynthetic eukaryotes within the Chlorophyta, stramenopiles, Haptophyta, Cryptophyta, and (2) dinoflagellates using the longer sequences. The closest match and published reference sequences were found following a basic local alignment search tool (BLAST) (Altschul et al., 1997) against the nr/ NCBI database. The sequences were aligned using MAFFT software version 6 with default parameters (Katoh and Toh, 2010), and alignments were visually inspected. For photosynthetic eukaryotes, we generated Maximum Likelihood trees based on sequence divergence estimated with Tamura 3-parameter nucleotide evolution model (Tamura, 1992). Node support was assessed via 100 bootstrap replicates. All positions with less than 95% site coverage, meaning sequences with fewer than 5% alignment gaps, missing data, or ambiguous bases at any position, were removed. All evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). All steps were similar for the dinoflagellate phylogenetic tree construction except that the Tamura-Nei nucleotide evolution model was used (Nei and Kumar, 2000). Non-redundant sequences from this study are available from GenBank under accession (KC488334 – KC488620 and KF761280 – KF761291), and clones with the same sequence are noted in supplementary Table SI.

Table 2.1 : Environmental and cell counts informations for samples used for 18S rRNA gene clone libraries collected stations from the Halifax line (HL) and Browns Bank line (BBL) in April and October 2009. Nutrients (NO^{3-}, PO_4^{3-}, SiO_4^{4-}); total extracted chlorophyll *a* (total Chl *a*); percent of Chl *a* <3µm in total Chl *a*; flow cytometry cell counts with nanophytoplankton (Nano), picoeukaryotes (PPE), *Synechococcus* (*SYN*), *Prochlorococcus* (*PRO*), and bacteria (Bact); no data (nd).

Date	Station	Depth (m)	NO_3^- (µM)	PO_4^{3-} (µM)	SiO_4^{4-} (µM)	Total Chl*a* (µg Chl*a* L^{-1})	Chl*a* <3µm (% total Chl*a*)	Nano ($\times 10^3$ cells mL^{-1})	PPE ($\times 10^3$ cells mL^{-1})	*SYN* ($\times 10^3$ cells mL^{-1})	*PRO* ($\times 10^3$ cells mL^{-1})	Bact ($\times 10^6$ cells mL^{-1})
10 Apr	HL2	2	1.9	0.8	2.7	8.4	nd	1.5	3.8	0.6	0	0.63
		20	3.0	0.9	3.0	8.6	nd	0.9	1.8	0.4	0	0.45
11 Apr	HL4	2	0.8	0.6	0.8	7.3	8.2	2.5	4.6	1.2	0	0.92
		10	0.7	0.5	0.7	7.6	5.3	2.0	3.9	0.9	0	0.91
11 Apr	HL5a	2	0.8	0.5	0.6	0.7	71.4	1.0	1.0	0.2	0	0.44
		50	3.9	0.8	2.0	2.2	9.1	0.5	0.8	0.5	0	0.46
16 Apr	BBL1	3	0.6	0.4	0.2	6.2	12.9	0.4	0.6	0.2	0	0.37
		20	3.1	0.8	2.1	12.5	16.8	0.8	1.2	0.4	0	0.47
16 Apr	BBL3	3	0.6	0.5	0.4	12.6	3.2	0.4	6.5	0.5	0	0.56
		10	0.6	0.5	1.1	16.3	0.6	0.8	1.2	0.3	0	0.50
14 Apr	BBL6	2	0.6	0.5	0.4	10.8	3.7	2.4	1.1	0.5	0	0.56
		20	0.9	0.5	0.9	11.0	2.7	2.6	2.1	0.5	0	0.62
05 Oct	HL2	2	0.1	0.1	1.2	0.9	nd	3.4	7.8	121.6	0	1.08
		30	0.3	0.1	1.3	0.4	75.0	1.6	4.2	48.4	0	1.14
06 Oct	HL5a	3	0.1	0.2	1.2	0.5	nd	3.4	6.7	97.9	8.7	0.82
		40	3.3	0.5	2.9	0.3	33.3	1.1	4.8	33.4	18.5	0.80
10 Oct	BBL1	3	0.2	0.3	1.6	1.0	nd	2.8	16.9	133.1	0	1.14
		20	0.7	0.4	2.1	1.0	40.0	2.2	13.5	100.6	0	1.11
08 Oct	BBL6	3	1.2	0.4	2.7	1.2	nd	4.7	72.8	76.9	0	1.81
		20	4.6	0.6	4.1	0.4	25.0	2.0	21.5	82.3	0	1.74

2.2.4 Data analysis and statistical tests

Oceanographic section data from CTD profiles, FCM analysis and contribution of picoeukaryotes to picophytoplankton carbon biomass were plotted in Ocean Data View version 4.3.10 (Schlitzer, 2002). Species richness using Abundance-based Coverage Estimator (ACE), Chao1, and Shannon diversity indexes were calculated for April and October using Mothur (Clarke, 1993). Matrix and SIMPER tests were performed with the Paleontological Statistics Software Package (PAST) version 2.12 (Hammer et al., 2001) based on relative abundance of taxa. Canonical Correspondence Analysis (CCA) was carried out using the Canoco for Windows package version 4.5 (Ter Braak and Smilauer, 2002). For each CCA (April and October), we used all taxa recovered in each month but only taxa with contributions identified by the SIMPER test with ≥1% contribution are displayed.

2.3 Results

2.3.1 Environmental variables

In April the water column was salinity stratified at ca. 50 m (**Figure 2.2**). The April surface waters were cold ranging from 0.8 to 4.5°C with the warmer temperatures offshore. Surface salinity also increased offshore seaward ranging from 31.2 to 33.1. In contrast, the October upper ca. 50 m was thermally stratified (**Figure 2.2**). Surface water temperatures ranged from 9.1 to 17.2°C, and salinity from 30.3 to 33.3. As in April, both temperature and salinity tended to increase offshore.

Nutrient concentrations in the surface and subsurface chlorophyll maximum (SCM) were generally greater in April compared to October. At the stations sampled for clone library construction, April nutrient concentrations ranged from 0.6 to 3.9 μM for nitrate, 0.4 to 0.9 μM for phosphate, and 0.2 to 3.0 μM for silicate. All tended to be lower in the surface (**Table 2.1**) compared to the SCM. In October, nutrient concentrations ranged from 0.1 to 4.6 μM for nitrate, 0.1 to 0.6 μM for phosphate, and 1.2 to 4.1 μM for silicate (**Table 2.1**). Bacteria concentrations (**Table 2.1**) were greater in October (0.8×10^6 to 1.8×10^6 cells mL^{-1}) than in April (0.4×10^6 to 0.9×10^6 cells mL^{-1}).

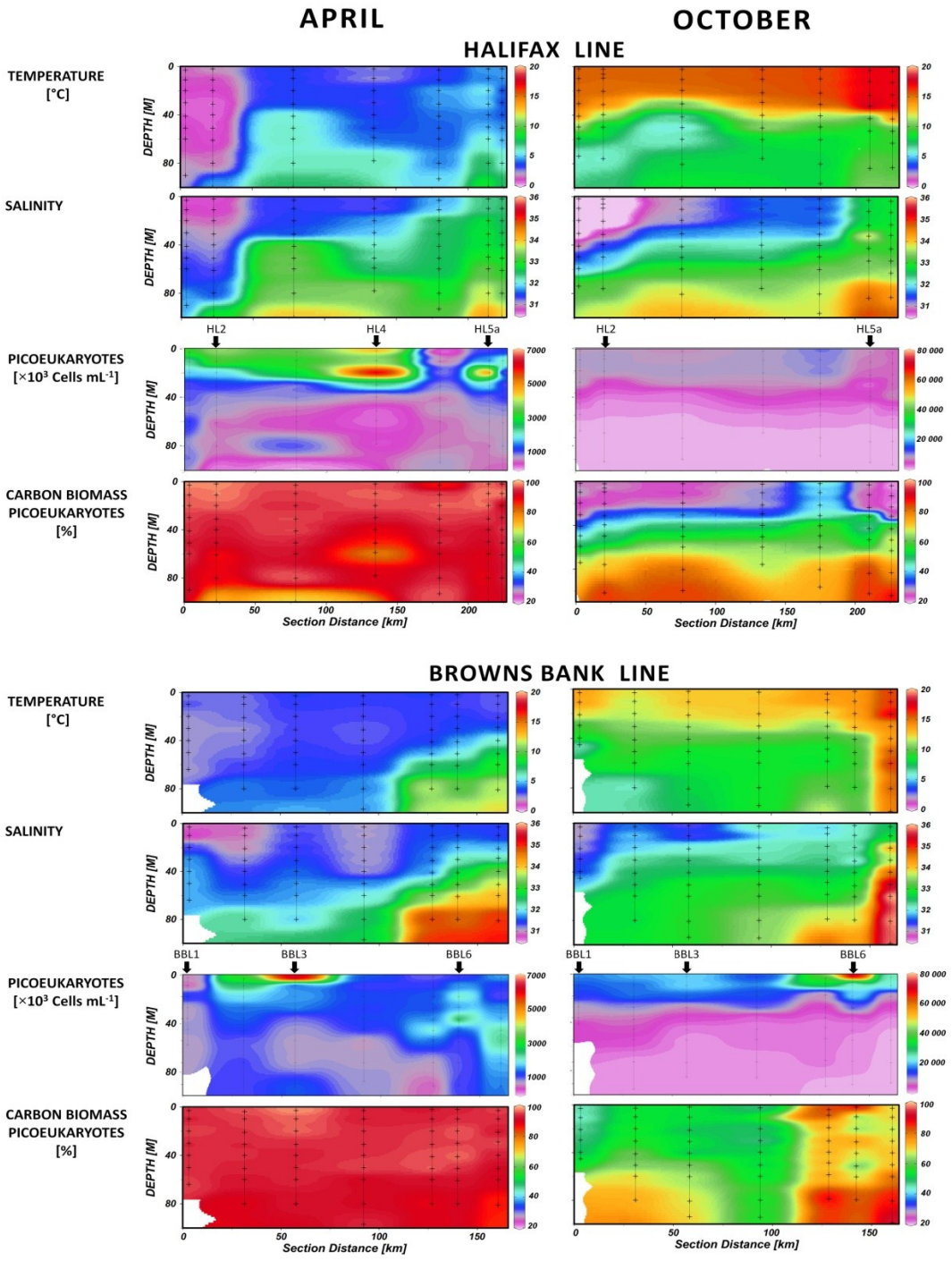


Figure 2.2: Section plots along the Halifax (the upper 8 panels) and Browns Bank Lines (the lower 8 panels), from April (left) and October (right) 2009. Temperature and salinity values from CTD cast along transects. Picoeukaryote concentrations (× 10^3 cells mL^{-1}) using FCM counts from discrete bottle samples indicated by a cross. Arrows indicates stations used for molecular analysis. Carbon biomass picoeukaryotes is the contribution of picoeukaryotes to picophytoplankton carbon expressed as percentage of total. For each panel the y-axis is the depth in meters and the x-axis is the distance in kilometers. Colors bars on the right indicate values for the variables.

Total Chl *a* concentrations were on average greater in April than October. In April, at the stations sampled for clone libraries, total Chl *a* ranged from 0.7 to 12.6 µg Chl *a* L^{-1} in surface and 2.2 to 16.3 µg Chl *a* L^{-1} at the SCM (**Table 2.1**), the lowest values were for the offshore station HL5a. The April small fraction contribution varied between 1 to 17%, except for the offshore HL5a surface station with very low Chl *a* concentrations where the small fraction accounted for 70% of the total (**Table 2.1**). In October, total Chl *a* concentrations were substantially lower everywhere and were more similar at the two depths, with values from 0.3 to 1.2 µg Chl *a* L^{-1} (**Table 2.1**). October ≤ 3 µm fraction Chl *a* was only measured at SCM, where it represented between 25 to 75% of the total (**Table 2.1**). The lowest concentrations were offshore, especially at station HL5a (**Table 2.1**).

FCM analysis indicated that the April small Chl *a* fraction would have been photosynthetic picoeukaryotes (PPE) (**Figure 2.2, Table 2.1**). In the upper ca. 50 m samples along both transects, the PPE represented 60 to 90% of the total picophytoplankton cell counts, ranging 0.6 to 6.5 × 10^3 cells mL^{-1} with highest concentrations at stations HL4 and BBL3 called 'patch' stations (**Figure 2.2**). *Synechococcus* concentrations were on average 3 fold lower than PPE, ranging 0.2 to 1.2 × 10^3 cells mL^{-1} (**Table 2.1**). *Prochlorococcus* was not detected in April. In October, the picocyanobacteria were much more abundant, in the upper ca. 50 m *Synechococcus* dominated with concentrations between 33 to 133 × 10^3 cells mL^{-1}, with PPE from 4 to 73 × 10^3 cells mL^{-1}; the maximum PPE concentrations were offshore at station BBL6. *Prochlorococcus* was detected at both offshore stations with concentrations 9 × 10^3 cells mL^{-1} in the surface and 18 ×10^3 cells mL^{-1} in the SCM at station HL5a. Nanophytoplankton concentrations were relatively similar in April and October. In the upper ca. 50 m; ranging 0.4 to 4.7 × 10^3 cells mL^{-1} along both transects (**Table 2.1**).

April PPE estimated carbon biomass (C) accounted for 89 to 99% of picophytoplankton C in the upper 50 m, and decreased below until 100 m with a minimum of 76% (**Figure 2.2**). In October, even when picophytoplankton cell counts were dominated by Synechococcus, PPE still accounted for 35 to 76% of the estimated picophytoplankton C in the upper 50 m along the Halifax Line, and at 100 m accounted for 60 to 91% of C. Along the Brown's Bank Line, values varied between 56 to 84% inshore and 70 to 95% of the along the continental slope and at depth (**Figure 2.2**).

2.3.2 Clone libraries and community comparisons

Following the FCM results, with an aim to capture a range of eukaryote communities especially within the small fraction, we constructed 20 clone libraries from surface and SCM samples collected along both transects: 12 in April and 8 in October.

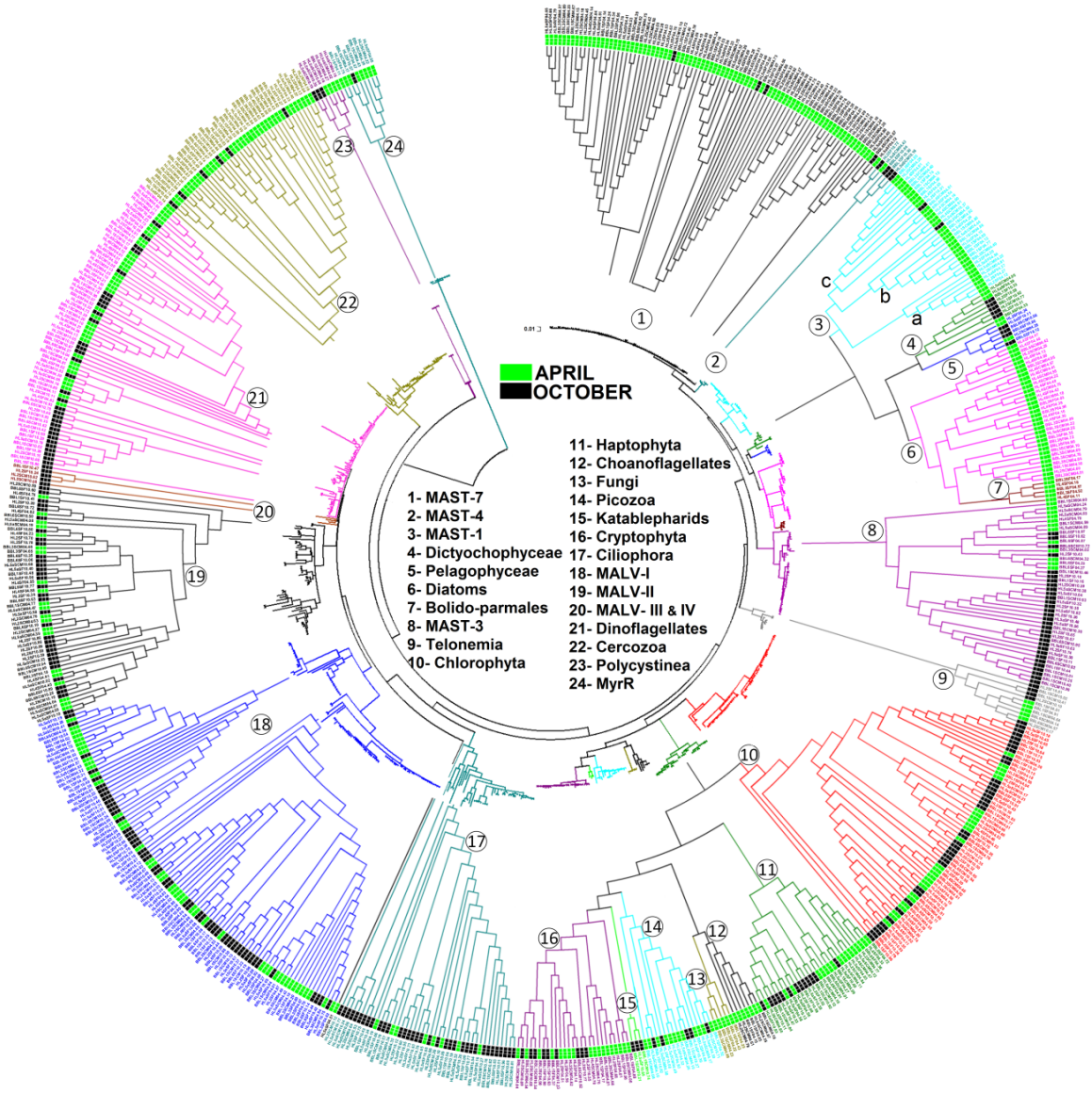


Figure 2.3: Global maximum likelihood (ML) phylogenetic tree using 704 nearly full length sequences. There were a total of 1195 positions in the final dataset. Clades are indicated by the number corresponding to the legend at the center of the tree. Actual ML distances are indicated at base of the circle, while individual groups have been expanded to show diversity and color coded by major group. (MyrR in the legend corresponds to *Myrionecta rubra*, a ciliate).

In total screening for quality, chimeras and excluding metazoan, 497 sequences in April and 327 sequenced in October were ≥950 nt and were classified into major eukaryotic lineages (**Annex 2_figure A1**). After grouping sequences at a level of 98% similarity (Romari and Vaulot, 2004) using Mothur, standard diversity estimates indicated that the communities were more diverse in October compared to April where similarly in April Chao1 was 360 in April compared to 418 in October, ACE diversity was 488 in April compared to 739 in October, while the Shannon index was 4.68 in April compared to 4.78 in October. In April, stramenopiles were dominant, except for offshore SCMs at stations HL5a and BBL6. October libraries had much greater proportions of alveolate sequences at all stations (**Annex 2_figure A1**).

Among the 824 sequences, we used 704 that were ≥ 1535 nt to provide a more accurate phylogeny for a Maximum Likelihood (ML) tree, which also highlighted overall diversity and differences between months (**Figure 2.3**). The most striking difference between months was within the stramenopiles, with most diatoms, Bolido-Parmales, MAST-1, and MAST-7 taxa unique to April. Cercozoa (Rhizaria) were also more common in April. Other groups such as dinoflagellates, Telonemia, MAST-3, and the haptophytes differed between the two months at the subclades and lower taxonomic levels (**Figure 2.3**). Ciliates and Polycystinea were mostly recovered from the October clone libraries (**Figure 2.3**).

Eukaryote communities differed significantly between April and October (**Figure 2.4**; weighted UniFrac significance test; $p \leq 0.01$) (Lozupone and Knight, 2005). Interestingly, the April HL5a SCM clustered nearer the October communities. In addition, the surface patches from April (Stations: HL4 and BBL3), identified by the FCM data, clustered apart from the other spring stations. Relative sequence abundance data (drawn from the sequences ≥ 950 nt in the 20 clone libraries) were used to identify which of the prevalent taxa influenced community clustering. We first tested explanatory power within the level of phylum, if differences were significant we then tested for specific effects at finer and finer taxonomic levels just to the level of genera. The most significant taxonomic levels that identified taxa responsible for the community clustering are shown in bold in **Figure 2.4**. Diatoms, Cryptomonadales, choanoflagellates, Bolido-Parmales and Fungi were exclusive to April, with the Bolido-Parmales only from the patches libraries.

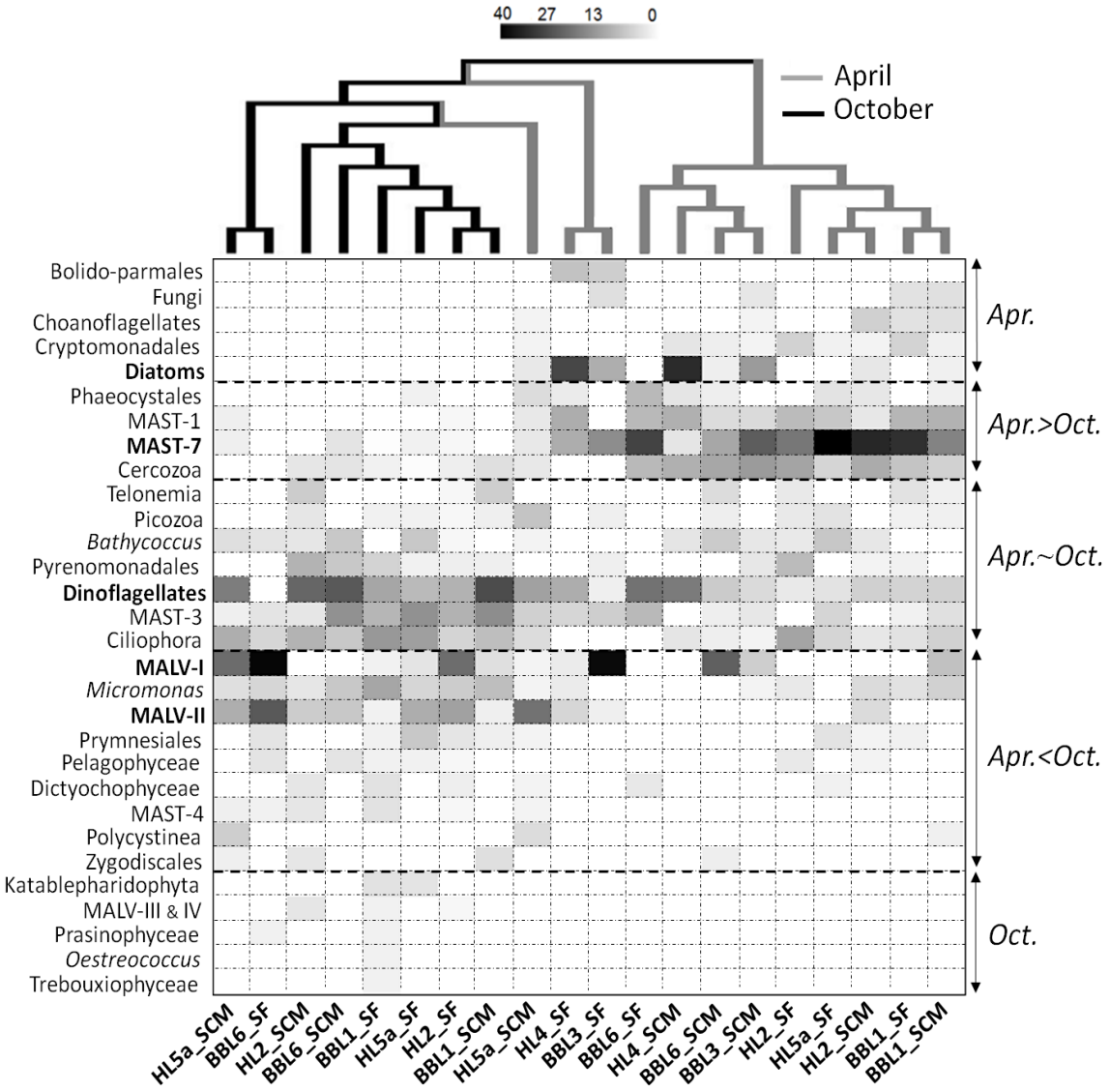


Figure 2.4: Cluster analysis based on phylogenetic lineages using weighted UniFrac (824 sequences) and matrix of relative abundance of taxa. Taxa in bold are those mainly responsible of difference between April and October, and were identified by SIMPER test.

The 8 sequences affiliated to fungi matched the uncultured eukaryote clone DSGM-64 (AB275064) with 99% of similarity (**Annex 2_Table A1**). Other taxa that were more prevalent in April compared to October included Phaeocystales, MAST-7, MAST-1 and Cercozoa. Taxa that were better represented in October, but also found in April, included marine alveolate (MALV) clades I and II, Pymnesiales, *Micromonas,* dictyochophytes,

pelagophytes, MAST-4, Polycystinea and Zygodiscales. There were several taxa unique to October, these included katablepharids (specifically *Katablepharis remigera* formerly *Leucocryptos*) and *Ostreococcus*. Few of the October specific sequences were found in more than two libraries (**Figure 2.4**).

2.3.3 Phytoplankton phylogenies

The majority of Chlorophyta sequences were affiliated with Mamiellophyceae, within three of the five proposed *Micromonas* clades (I, III and V) based on the 18S rRNA gene (Worden et al., 2009). *Micromonas* and *Bathycoccus* occurred both in April and October (**Figure 2.5**). *Ostreococcus* and representatives of other chlorophytes were found only in October (**Figure 2.5**). Among photosynthetic stramenopiles (Heterokonta), diatom sequences were assigned to the genera *Minidiscus, Thalassiosira, Chaetoceros, Fragilariopsis* and *Pseudo-nitzschia*, all of which were exclusively recovered in April. Also in April were *Bolidomonas* and *Triparma* (Bolido–Parmales) from patch stations (**Figure 2.6a**). Other taxa, were mainly but not exclusive to October libraries, had matches to Dictyochophyceae; *Florenciella, Pseudochattonella, Dictyocha,* and *Pedinella*, and Pelagophyceae (*Aureococcus*) (**Figure 2.6a**). Among Haptophyta sequences, those matching three separate species of *Phaeocystis* (Phaeocystales) were recovered. While April haptophytes were mostly *Phaeocystis*, *Chrysochromulina* (Prymnesiales) were more common in October (**Figure 2.6b**). One sequence clustered within the Zygodiscales, as its best match was to *Braarudosphaera* (**Figure 2.6b**). Within the Cryptophyta, sequences matching the genera *Rhodomonas* and *Hemiselmis* were detected mainly in April (**Figure 2.6c**), while those in a clade that includes *Teleaulax* and *Geminigera* were recovered in both seasons (**Figure 2.6c**). Dinoflagellates were taxonomically diverse; most within the poorly resolved Gymnodiniales Peridiniales, Prorocentrales (GPP) group (Taylor et al., 2008). G*ymnodinium* sensu stricto (ss) (Gomez et al., 2009) were predominant in the October libraries with other clades less specific to either month. A large proportion of the April sequences were closest to the heterotrophic *Gymnodinium* sp. reported in Saldarriaga and colleagues (Saldarriaga et al., 2004) (**Figure 2.7**).

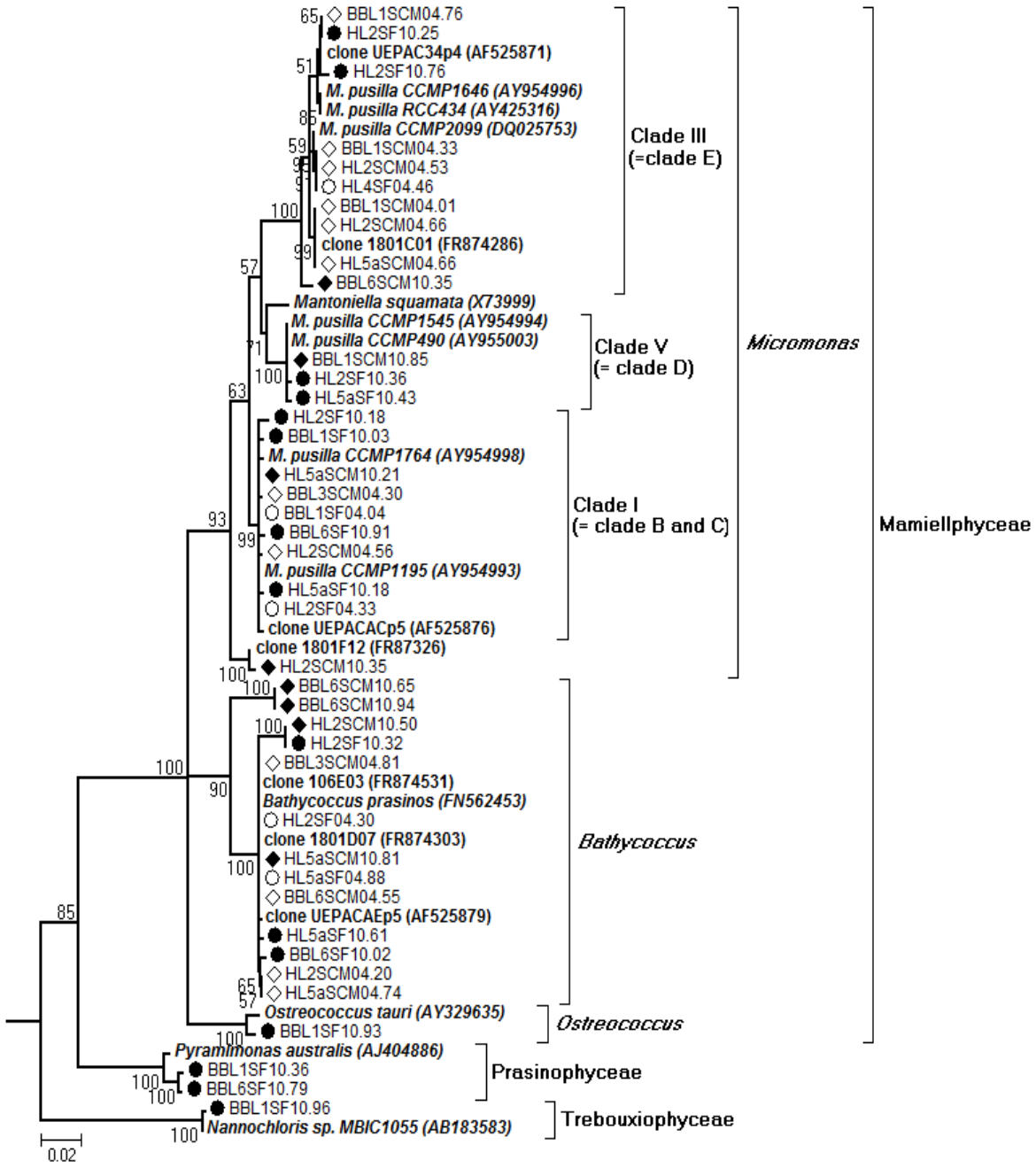


Figure 2.5: ML phylogenies of Chlorophyta rooted with *Cyanoptyche gleocystis* (AJ07275) and *Glaucocystis nostochinearum* (X70803) not shown. There were a total of 1621 positions in the final dataset and the analysis involved 77 nucleotide sequences. Bootstrap values are shown at nodes with >50% support (100 replicates). April clone libraries are indicated by open symbols and October by closed symbols; circles indicate surface sample and diamonds indicate the subsurface chlorophyll maximum (SCM). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

2.3.4 Communities, environment and taxa

Canonical Correspondence Analysis (CCA) was performed separately for April and October based on relative abundance of taxa drawn from Fig. 4. The suite of environmental parameters (temperature, salinity, nutrients, Chl *a* and bacterial concentrations) explained ca. 51% of variability in April (**Figure 2.8a**) and ca. 53% in October (**Figure 2.8b**). In April taxonomic composition of coastal and patch stations trended with total Chl *a*. Bacteria concentrations and taxonomic composition for the patch stations were associated with each other. The April offshore stations clustered apart from inshore stations, and physical factors including higher temperature and salinity were more important factors associated with the clustering. In contrast to April, where there was no separation between surface and SCM origin samples, the October communities from the two depths differed. Taxa from the surface stations followed salinity, while SCM communities showed little association with any environmental variable tested.

2.4 Discussion

2.4.1 New records of taxa in a well studied system

The Scotian Shelf (SS) is a relatively well-studied temperate continental shelf with a long history of observational data, but molecular surveys of the microbial eukaryotes are lacking. The use of molecular surveys has highlighted the emerging diversity of marine microbial eukaryotes (Del Campo and Massana, 2011) as well as a cryptic diversity within described species (Stoupin et al., 2012). While we expected to recover new records of small heterotrophic taxa, which are largely uncultivated and rarely reported in microscopy surveys, it was somewhat surprising that our clone libraries revealed phytoplankton taxa that had not been reported previously. For example, we retrieved sequences with best matches to genera in the Dictyochophyceae including sequences related to *Aureococccus*, and *Pseudochattonella,* both considered to be harmful algae (Gobler and Sunda, 2012; Skjelbred et al., 2011). The libraries constructed in April included sequences of diatom species that have been previously reported, such as *Fragilariopsis cylindus*, *Thalassiosira*

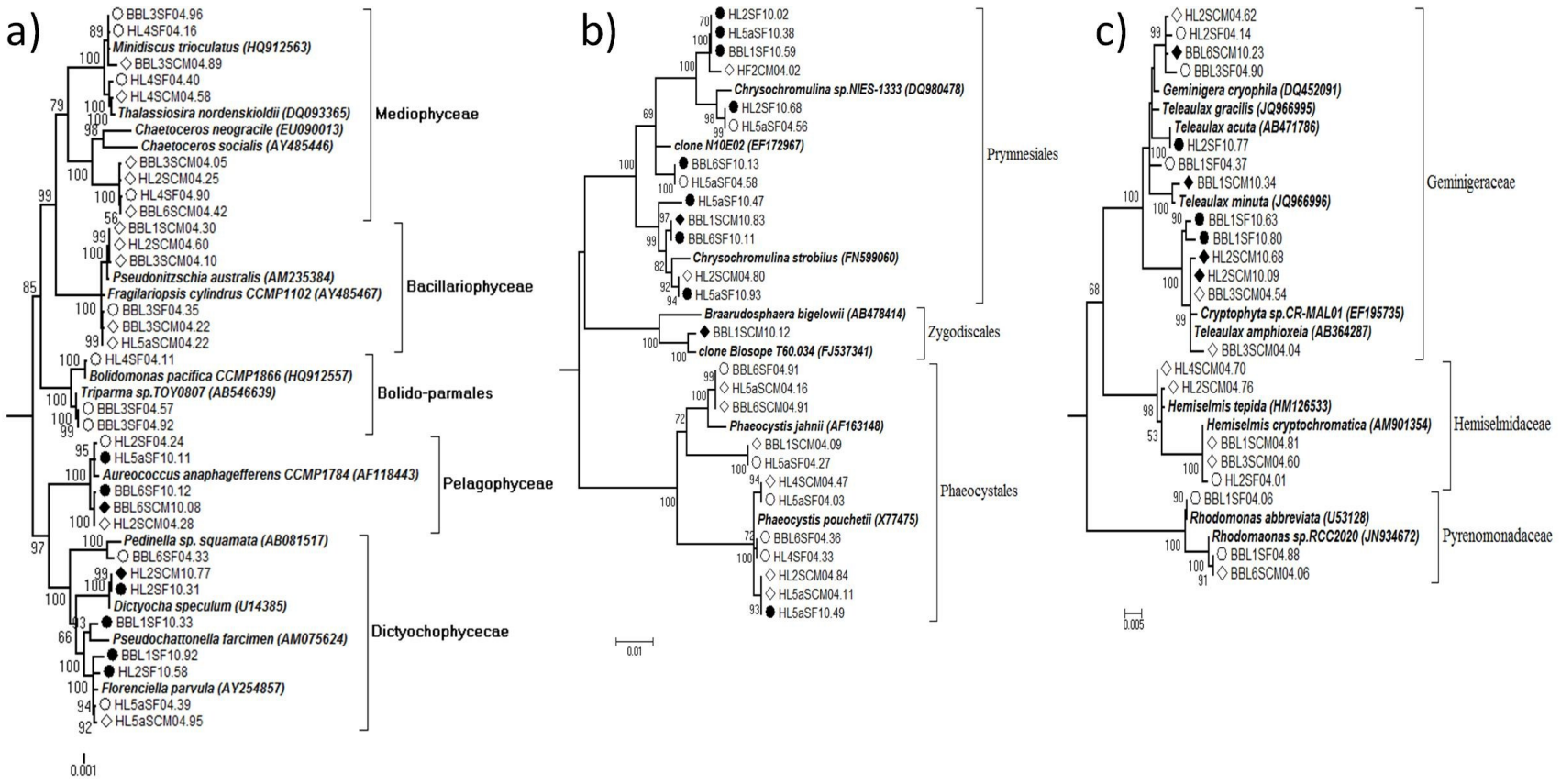


Figure 2.6: ML phylogenies of **a)** Photosynthetic stramenopiles were rooted with *Gyrodinium helviticum* (AB120004) not shown. There were a total of 1581 positions in the final dataset and the analysis involved 44 nucleotide sequences, **b)** Haptophyta rooted with Cryptophyceae *sp.* CCMP2045 (GQ375264), not shown. There were a total of 1539 positions and the analysis involved 34 nucleotide sequences, **c)** Cryptophyta rooted with *Emiliania huxleyi* (HQ877901) not shown. There were a total of 1566 positions and the analysis involved 37 nucleotide sequences. All bootstrap values are shown at nodes retaining >50% support (100 replicates). April clone libraries are indicated by open symbols and October by closed symbols; circles indicate surface sample and diamonds indicated the subsurface chlorophyll maximum. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site.

aestivalis and *Minidiscus trioculatus* (Mather et al., 2010), and sequences that could represent new records such as a sequence closest to *Pseudo-nitzschia australis* (**Figure 2.6a**). One cluster of four *Chaetoceros* sequences was clearly within the subgenera *Hylochaetae* but was < 95% similar to described or environmental *Chaetoceros* sequences in Genbank. These sequences could be due to poor representation of morphologically described species in public database or represent a novel genera level group. Small *Chaetoceros* are abundant in spring bloom waters and simple morphology could mask high diversity (Booth et al., 2002; Degerlund et al, 2012). Cryptophytes are also commonly reported from the coastal Atlantic (Breton et al., 2000; Gibb et al., 2001) but few actual species have been identifed from the more northern region with only three genera representing four species reported (Li et al., 2011). In contrast, we recovered 10 species level clusters from the two months, including those within previously unreported genera *Teleaulax* and *Geminigera* (**Figure 2.6c**). This and other studies suggest that cryptophytes may be routinely under reported in microscopy studies, for example Metfies and colleagues (Metfies et al., 2010) working in the German Bight, investigated clade level differences in cryptophytes using a phylochip, reporting previously underappreciated seasonal and interannual changes in abundance and taxonomic affinities. Similarly, within the non-coccolithophorid haptophytes we also found previously unreported diversity even within *Phaeocystis,* which were mostly found in the spring samples, with at least two other likely species in addition to *P. pouchetii* (**Figure 2.6b**). Many *Phaeocystis* species including *P. jahnii* occur as small single cells may have escaped detection in earlier studies that employed only light microscopy techniques. Other photosynthetic eukaryotes, such Florenciellales (Edvardsen et al., 2007) and Bolido-Parmales (Ichinomiya et al., 2011), are considered true picophytoplancton and often detected using molecular techniques but rarely by microscopy, and the potential ecological importance of these small cells has only been recognized over the last decade (Vaulot et al., 2008). Given the close phylogenetic relatedness of cultivated bolidophytes, which are picoflagellates, and the one cultivated Parmales species (Ichinomiya et al 2011) it is currently not possible to know the morphology of the Bolido-Parmales found in our transects. The taxonomy of Parmales species is based on the arrangement and form of silica plates and flagella are not reported.

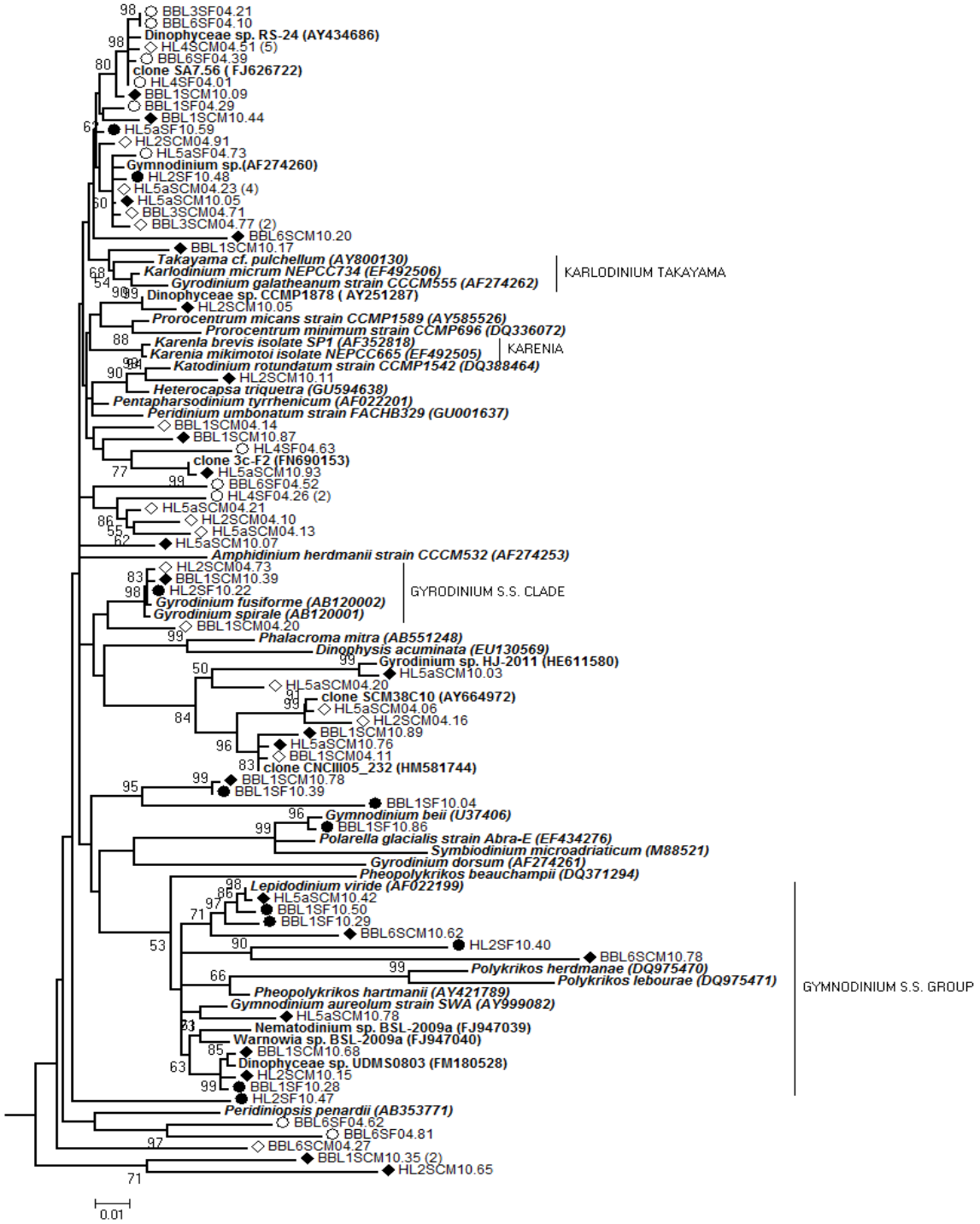


Figure 2.7: ML phylogeny of dinoflagellates rooted with *Tetrahymena pyriformis* (EF070254) and *Strombidium* sp.SBB99-1 (AY143565) not shown. There were a total of 1637 positions in the final dataset and the analysis used 121 sequences. Bootstrap values are shown at nodes retaining >50% support (100 replicates). April clone libraries are indicated by open symbols and October by closed symbols; circles indicate surface sample and diamonds indicated the subsurface chlorophyll maximum. Tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

These small siliceous cells are easily detected using scanning electron microscopy. Although these were not noted in a recent review of the Scotia Shelf and Gulf of Maine region (Li et al., 2011), three Parmales species have been reported from the Gulf of St. Lawrence (Bérade-Therriault et al., 1999), with *Tetraparma pelagic* and *Triparma strigata* found in both spring and autumn samples and *Triparma laevis* recovered in March only.

2.4.2 Spring picophytoplankton

Other likely true picophytoplankton recovered in April samples were two *Micromonas* clades and *Bathycoccus*. *Micromonas pusilla* although currently classified as a single species is phylogenetically diverse (Slapeta et al., 2006; Worden et al., 2006). We recovered two separate clades in spring, clades I and III (**Figure 2.5**). In the SS, Micromonas pusilla clade III included ribotypes matching the isolate CCMP 2099 which is considered an Arctic ecotype (Lovejoy et al., 2007; Balzano et al., 2012). This ribotype was recovered at the coastal stations and the Halifax patch station, where waters are strongly influenced by the Labrador Current, which originates in the Arctic. Since this strain is normally found in waters ranging from -1.7 to 4°C in the Arctic and does not grow above 10 °C (Lovejoy et al., 2007), the occurrence in the SS suggests transport by Polar waters.

In April, surface Chl *a* concentrations at the most offshore station along the Halifax transect (HL5a, ~ 200 km of the coast) was much lower than elsewhere. The high percentage of small fraction Chl *a* and higher PPE cell counts in HL5a and low nutrient levels indicated that the bloom had declined, at least in the surface waters. The fact that the community grouped closer to the October communities in the cluster analysis (**Figure 2.4**) suggests that the community was moving towards a summer consortium. Nutrient concentrations were also low at the HL4 patch stations suggesting that the bloom had started offshore in warmer Atlantic waters and moved toward the coast, HL4 representing a state where nutrients were depleted but Chl *a* levels had not yet declined or was in equilibrium with supply. The predominance of smaller cells in these lower nutrient waters are consistent their capacity to dominate the phytoplankton biomass under conditions when nutrients or light are limiting (Raven, 1998; Barton et al., 2010; Follows and Dutkiewicz, 2011). The placement of the surface patch communities outside of the other April communities would be consistent with a transitional state between the spring bloom and the

more summer community manifested offshore at HL5a. The lower silicate concentrations in the offshore surface samples is consistent with silicate depletion favoring non diatom microbial eukaryote recovered from the HL5a, with smaller diatoms still persisting in the patch BBL3 and HL4 surface clones libraries (**Figure 2.2**).

2.4.3 Autumn phytoplankton

In contrast to spring bloom, the dynamics of the late summer and fall phytoplankton prior to an autumnal bloom is much less studied (Mousseau et al., 1996; Green and Pershing, 2007; Friedland et al., 2008). Nutrients were depleted at all stations in October, leading to low total Chl *a* levels and a high proportion of picophytoplankton detected via FCM. The photosynthetic community was numerically dominated by picocyanobacteria; mostly *Synechococcus*, which is usually abundant in summer and fall on the SS (Table 1) (Li et al., 2006), and unusually *Prochlorococcus,* which is rarely reported at latitudes >35° (Rocap et al., 2003; Li, 2009; Malmstrom et al., 2010). The occurance of *Procholorocuccus* is consistent with the unusally high offshore temperatures in October and suggest an influence of the Gulf Stream which shifts north in autumn (Kelly and Gille, 1990; Frankignoul et al., 2001).

October picophytoplankton included *Micromonas* and *Bathycoccus,* which were reported previously in the SS (Townsend et al., 2006), and *Ostreococcus*, which was not reported previously. *Ostreococcus* is also favored by warmer waters and along with *Micromonas* and *Bathycoccus* are usually thought to be more coastal, where nutrient concentrations tend to be higher (Not et al., 2008; Not et al., 2012). While *Ostreococcus* were inshore, October *Micromonas* and *Bathycoccus* were also recovered offshore. *Bathycoccus* was originally described from deeper Mediterranean samples and is common in the northern Atlantic Ocean (Eikrem and Throndsen, 1990), and despite little difference in the 18S rRNA gene there are probably two distinct ecotypes (Monier et al., 2012; Vaulot et al., 2012). However at the 18S rRNA gene level little variation has been reported previously and two sequences from the offshore Browns Bank line were unusual (**Figure 2.5**) and with closest matches only 97% similar to other Mamiellophyceae and is outside of known *Bathycoccus* diversity and may be novel.

Despite the paradigm of low nutrients favoring picocyanobacteria, eukaryotic picophytoplankton often account for a higher proportion of C biomass even during periods when *Synechococcus* concentrations are high (Li and Dickie, 2001; Li et al., 2011) including here (**Figure 2.2**). Generally, picocyanobacteria are reported to be favored under low nutrient conditions compared to PPE due to the smaller size and higher surface to volume ratio (Murphy and Haugen, 1985; Shapiro and Guillard, 1986). It was also previously thought that cyanobacteria can utilize ammonium or urea as the sole nitrogen source for their growth (Waterbury et al,. 1986) and eukaryotes were better at taking up and using nitrate. However, both ammonium transporters an active urea cycle are reported in the clade V *Micromonas* CCMP 490 and in many other eukaryotic marine microbes (Worden et al., 2009). *Micromonas* in the CCMP 490 clade, which was only recovered in the October libraries, would have relied on recycled nutrients (ammonium and orther organic N sources) offshore. In addition to being able to compete for recycled nutrients, the persistence of photosynthetic picoeukaryotes could be aided by their slightly larger size and ability to escape grazing by bacterivors such as MAST-4, which reportedly prefers cyanobacteria (Lin et al., 2012) and was present in the same October samples. This persistence of PPE could also be explained in part by mixotrophy, and the same species would be able to persist under a broader range of nutrient and light regimes compared to strict phototrophs or heterotrophs (Hartmann et al., 2012; Hartmann et al., 2013).

Among dinoflagellates, the *Gymnodinium* sensu stricto clade (Gomez et al., 2009) was recovered only in October (**Figure 2.7**). The known species in this clade are ultrastructurally diverse, several of the sequences from the October SS libraries were closest to *Lepidodinium spp.*, which have green algal derived chloroplasts and other sequences grouped with the peridinin containing dinoflagellate *Gymnodinium aueolum*. These species are mixotrophic and feed on nanoflagellates using a peduncle (Hansen, 2001; Hansen et al., 2007; Jeong et al., 2010). A recent report that mixotrophs, graze at higher rates under warmer conditions (Wilken et al., 2013), would suggest that these species could be highly dependent on grazing in October when temperatures were higher and inorganic nutrients at low concentrations. All other dinoflagellates taxa were found in both October and April, and belonged to diverse clades. Interestingly, the majority of these dinoflagellates are also reported to be mixotrophic (Hansen et al., 2011), and their presence

in both April and October suggests greater tolerance to lower temperatures in April compared to the *Gymnodinium sensu stricto* clade.

2.4.4 Heterotrophic microbial eukaryotes

Microscopy studies where phototrophic and heterotrophic flagellates have been distinguished often report that heterotrophic flagellates account for a significant proportion of the microbial eukaryotic biomass (Sherr and Sherr, 2002; Sherr and Sherr, 2009). As well, the majority of 18S rRNA gene surveys of marine microbial eukaryotes are dominated by putative heterotrophs (Vaulot et al., 2008). In April along with choanoflagellates (King et al., 2009) were Cercozoa highlighting a dominant role for bacterivores during spring, at a time of high release of dissolved organic material (Piwosz and Pernthaler, 2010; Anderson et al., 2013). However, the majority of April taxa assigned to heterotrophic groups belonged to two clades of uncultivated marine stramenopiles (Massana et al., 2004); MAST-7 and MAST-1. Other heterotrophs, notably Telonemia and Picozoa were recovered almost equally in April and October (**Figure 2.3**). Telonemia consists of multiple clades (Shalchian-Tabrizi et al., 2007) and is widely distributed with records both from microscopy and clone libraries (Brate et al., 2010). Monier and colleagues (Monier et al., 2013) found that particular ribotypes of *Telonemia* likely bloomed under specific but unresolved conditions. We found one dominant clade in the April samples, with more diversity in October, but there was not strong segregation of different clades by season. Picozoa, previously known as 'picobiliphytes' (Seenivasan et al., 2013) lack a plastid and may feed on colloids. Picozoa are commonly recovered in 18S rRNA gene surveys and single cell genomes from off the coast of Maine (Yoon et al., 2011) suggest that some of them may also form close associations with bacteria and viruses. Fungi are rarely reported from microscopy studies, but tend to be persistently recovered at low frequencies using molecular techniques (Lovejoy, unpublished data). They are likely important at times in oceanic carbon and energy cycles, but detailed studies are lacking (Wang et al., 2012). The fungal sequences from April were 99% similar to an environmental clone recovered from a deep sea methane seep (Takishita et al., 2007). Our phylogenetic analysis puts this clone at the base of the Chytridiomycota or Zygomycoata (data not shown). Chytridiomycota are widely reported to be parasites of diatoms in freshwater systems (Rasconi et al., 2013;

Canter and Lund, 1953) leading to the speculation that they fungi here played a similar ecological role, and were associated with the decline of the spring diatom bloom.

While spring diatoms and *Phaeocystis* contributed to the separation of April and October as expected, much of the difference between the two months was attributed to phylogenetically diverse marine stramenopiles and putatively parasitic alveolates. These groups are 'invisible' using FCM protocols based on chlorophyll fluorescence and so have likely been underestimated. Interestingly we recovered more species level taxa of microbial eukaryotes in October compared to April. In addition, while the total proportion of sequences from heterotrophs was similar between the two months (65% in April and 58% in October), there were marked differences in the relative abundance of different clades. Among heterotrophic stramenopiles were representatives of four MAST clades. The majority of MAST-4 and MAST-3 sequences were recovered in October, while MAST-7 and MAST-1 were relatively rare in October compared to April. In April MAST-1 included three previously described subclades (1a, 1b, and 1c) (Massana et al., 2006) and were in all the libraries (**Figure 2.4**). Fluorescence *in situ* hybridization data from elsewhere suggests that MAST-1 includes both pico- and nano- plankton (Thaler and Lovejoy, 2014; Logares et al., 2012), while MAST-7 is mostly in the picoplankton range and grazes on small bacteria (Logares et al., 2012). MAST-4 is widely distributed and found in non-polar oceans (Massana et al., 2006). Here, MAST-4 were absent from the colder April waters and recovered in October along with other warmer water species. Temperature influences distribution patterns of MAST-4 (Martinez et al., 2013) and it is also reported to graze on *Synechococcus* (Lin et al., 2012), consistent with MAST-4 co-occurrence with *Synechoccoccus* in October. Sequences belonging to the MAST-3 clade were recovered in both seasons, however these fell into April or October sub-groups (**Figure 2.3**). Phylogenetically different species level clades of MAST-3 were reported earlier (Logares et al., 2012) and the different strains may be exploiting different ecological niches. MAST-3 is reported to be globally the most abundant of the MAST clades (Logares et al., 2012), and although it has been suggested that MAST-3 may be diatom epibionts or even parasites (Gomez et al., 2011), their occurrence in October when diatoms were absent indicates that the group is likely functionally diverse and not limited to co-occurrence with diatoms.

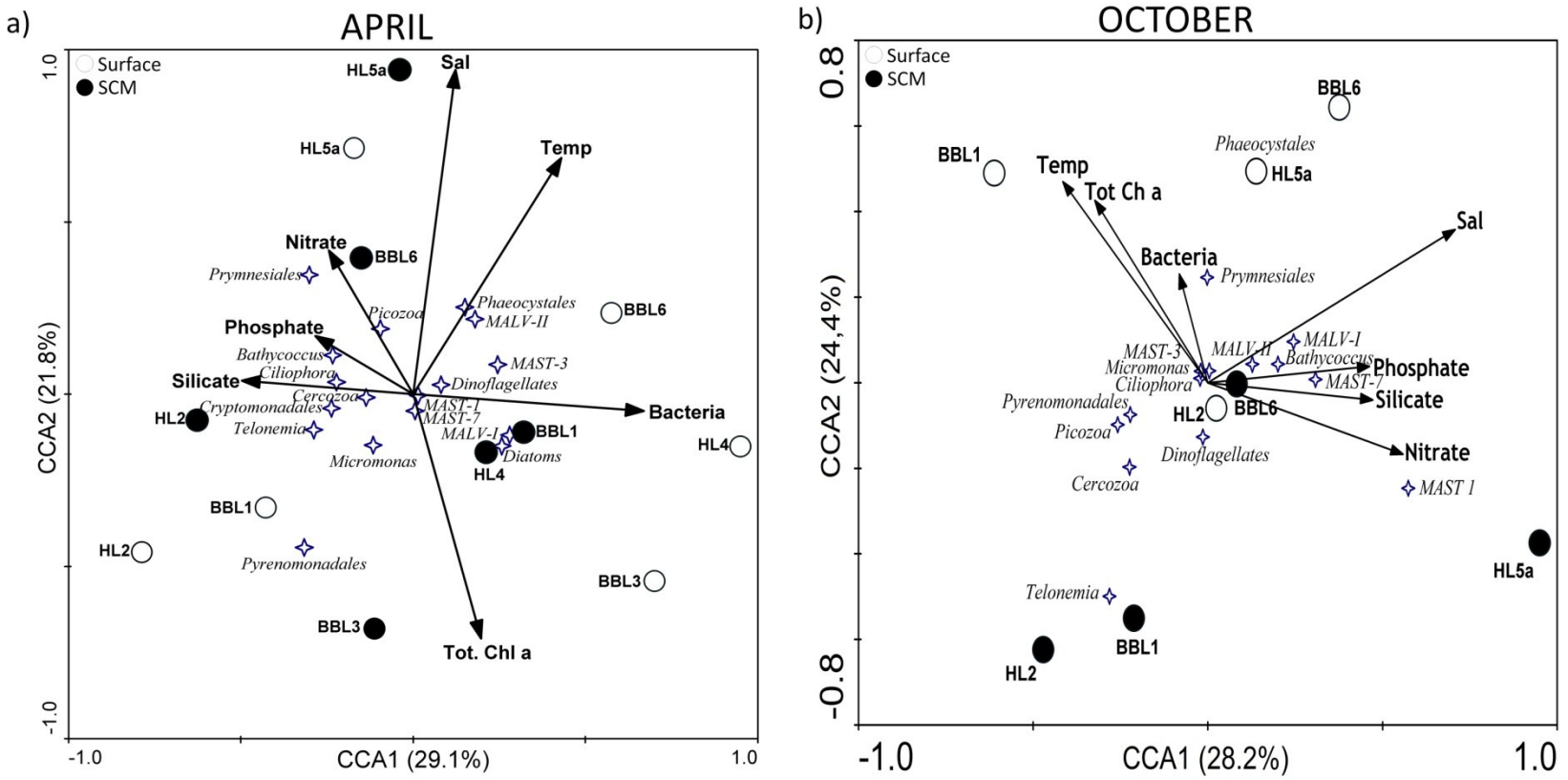


Figure 2.8: Canonical Correspondence Analysis (CCA) plot using relative abundance of taxa (%) detected **a)** in April and **b)** in October run on CANOCO. All taxa detected were used for analysis but only taxa which were responsible to clustering between April and October with a minimum of 1% of total contribution identified by the SIMPER test are displayed. Variables are nutrients: phosphate, nitrate, silicate (µM), salinity (Sal), temperature (Temp, °C), total chlorophyll *a* (Tot. Chl *a*, µg Chl *a* L^{-1}) and bacteria (× 10^6 cells mL^{-1}) from flow cytometry. White circles indicate surface sample and black circles indicated the subsurface chlorophyll maximum.

Although in principle our libraries would consist of only cells < 3 μm in one dimension and able to pass through a 3 μm filter (Vaulot et al., 2008; Seenivasan et al., 2013), ciliates and dinoflagellates were especially common in October (**Figure 2.3**). These groups are often reported from < 3 μm surveys despite the pre-filtration step that should exclude them by size. Their presence in the small fraction could be due to several factors including fragile cells breaking-up on filtration (Terrado et al., 2011) and DNA persisting in the environment (Josephson et al., 1993). Another substantial group of sequences that were recovered especially in October, belonged to uncultivated putatively parasitic alveolates (MALV) (Siano et al., 2011), which can infect many larger cells (Brate et al., 2012; Siano et al., 2011). Parasites could reroute fixed carbon away from zooplankton grazers, and disrupt resource competition between species by preferentially infecting the most actively growing species (Guillou et al., 2008). In which case, sequences belonging to dinoflagellates, ciliates and Polycystinea could be from live (infectable), recently dead cells (killed by the parasite) or persistent DNA. Several groups of MALV have small pico-size zoospores and appear to be ubiquitous members of the marine picoeukaryote communities (Guillou et al, 2008).

2.4.5 Environmental influences

In April, distribution of taxa followed an onshore-offshore pattern, while in October, there was an additional vertical influence, where the SCM offshore communities differed from the surface communities (**Figure 2.8**). Size fractionated Chl *a* and FCM data indicated that picoeukaryotes were a non-negligible component of the photosynthetic communities in April and were also abundant in October, even when nutrients were depleted and *Synechococcus* concentrations were high. Previous flow cytometry studies highlighted the occurrence and likely ecological importance of what were termed eukaryotic ultraphytoplankton (≤ 5 μm) (Li et al., 2011), but the taxonomic identity of these small cells was not known. Dauchez and colleagues (Dauchez et al., 1996) described how ultraphytoplankton communities contributed to new production (nitrate over nitrate + ammonium uptake or *f* ratio) on the SS throughout the year. Clearly the smaller microbial eukaryotes would have contributed to this uptake, since nitrate transporters are found in, for

example *Micromonas* and *Bathycoccus* (Worden et al., 2009; Moreau et al., 2012). While nitrate concentrations were depleted in October surface waters, concentrations were high in the SCM offshore stations. Picophytoplankton in these deeper samples would have an advantage under lower irradiance levels compared to larger cells (Raven et al., 1998).

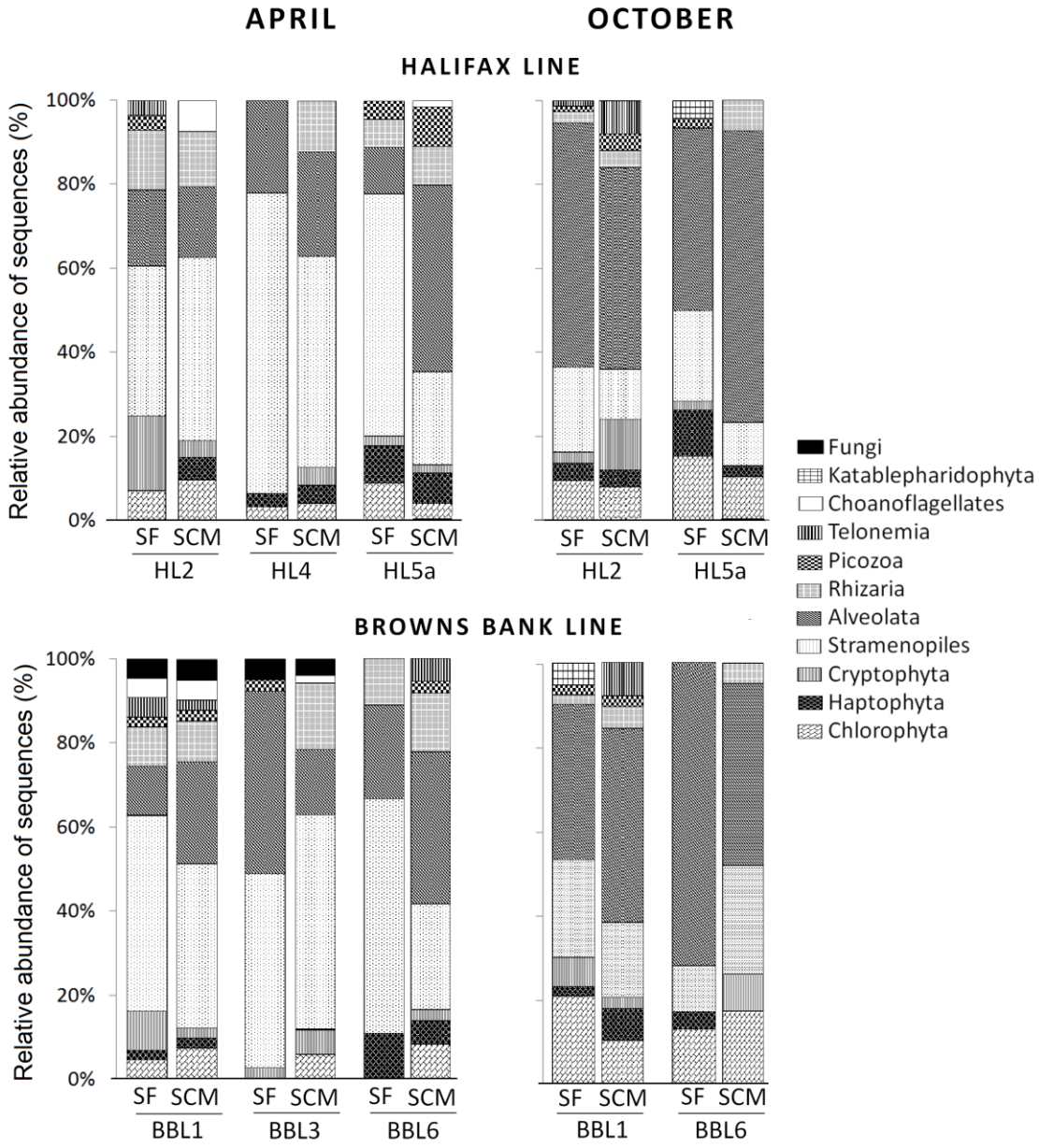
The two sampling dates were selected based on known phenology of larger phytoplankton species succession (Greenan et al., 2008; Song et al., 2011) typical of the temperate North Atlantic (Joint et al., 1993; Longhurst, 1995). In this study, using molecular markers, we were able to identify an underlying mostly flagellate community with some taxa more characteristic of the April spring bloom and others more characteristic of October low nutrient warm water conditions. Other taxa appear to have broader temporal distributions, at least at the level of 18S rRNA gene phylogeny. Among these, were species that are considered true picophytoplankton such as *Micromonas*, *Bathycoccus*, *Florenciella*, *Aureococcus* and mostly uncultivated picoheterotrophic eukaryotes such as Picozoa and MAST-3.

2.5 Acknowledgements

We thank Karen Scarcella and Emmanuelle Medrinal for assistance in the field, Jeff Anning for nutrients and FCM data, Bob Ryan for CTD data, and finally André Comeau, Ramon Terrado, Mary Thaler, and others members of Lovejoy Lab for comments and analytical advice. We acknowledge the professional assistance of the Canadian Coast Guard crew of the CCGS Hudson and are indebted to Compute-Canada and the CLUMEQ Supercomputing Facility. This work is a contribution to the Canadian Healthy Oceans Network (CHONe) funded by the Natural Science and Engineering Research Council of Canada (NSERC) and the Atlantic Zone Monitoring Program (AZMP) of the Department of Fisheries and Oceans Canada Bedford Institute of Oceanography. NSERC Discovery funds to CL enabled the completion of this work.

Annex 2

Annex 2_Figure A1: Relative proportions of the major phylogenetic groups of microbial eukaryote sequences from the 20 clone libraries along Halifax (HL) and Browns Bank Lines (BBL) from surface (SF) and subsurface chlorophyll maximum (SCM).



Annex 2_Table A1: Highest BLAST match, and percent similarity (%) to the non-redundant sequences are given for Chapter 2 with 824 sequences retained were grouped into 292 non-redundant sequences that we have deposited in GenBank.

OTU Accession n°	Name of Reference Sequence by OTU	Highest match (accession number)	Similarity (%)	Origin	List of clones by OTU
CHLOROPHYTA					
KC488346	BBL1SCM04.76	Uncultured marine eukaryote clone UEPAC34p4 (AF525871)	99	Pacific Ocean	HL2SF10.76, BBL6SCM10.35, HL2SF10.62, HL2SF10.25, HL2SCM04.53, HL4SF04.46, BBL1SCM04.33, HL2SCM04.66, HL5aSCM04.66, BBL1SCM04.01, BBL1SCM04.76
KC488347	BBL1SCM10.85	*Micromonas pusilla* CCMP1545 (AY954994)	99		HL5aSF10.43, BBL1SCM10.77, HL2SF10.36, BBL1SCM10.26, BBL1SCM10.85
KC488348	BBL1SF10.93	*Ostreococcus lucimarinus* CCE9901 CP000592)	99		BBL1SF10.93
KC488349	BBL1SF10.96	*Nannochloris* sp. MBIC10055 (AB183583)	99		BBL1SF10.96
KC488350	BBL3SCM04.92	*Bathycoccus prasinos* CCMP1898 (JX625115)	100		HL2SF10.32,HL2SCM10.50,HL4SCM04.65,HL5aSCM10.11,BBL6SCM04.95,BBL6SF10.02,HL5aSF10.61,HL5aSF10.35,HL5aSF04.11,HL5aSCM04.74,HL2SCM04.20,HL5aSF10.29,HL5aSCM10.81,HL5aSF04.76,HL2SCM04.19,HL2SF04.30,HL5aSF04.88, BBL6SCM04.46,HL5aSF04.38,BBL6SCM04.55,HL5aSF10.56,BBL3SCM04.92,BBL3SCM04.81
KC488351	BBL6SCM10.94	uncultured marine picoeukaryotes,ws_159, clone 1815A09 (FR874693)	98	Fjord coastal water	BBL6SCM10.94,BBL6SCM10.65
KC488352	BBL6SF10.79	Uncultured eukaryote clone NPK97_402 (EU371355)	99	Ice from Austre	BBL6SF10.79,BBL1SF10.36
KC488353	HL2SCM10.35	uncultured picoeukaryote ws_54, clone 1801F12 (FR874326)	99	Fjord coastal water	HL2SCM10.35,BBL1SCM10.45
KC488354	HL2SF10.60	Uncultured Chlorophyta clone 6c-H11 (FN690732)	99	sea ice Baltic Sea	HL2SF10.60
KC488355	HL5aSCM10.17	uncultured marine picoeukaryotes ws_159, clone 1815D02 (FR874721)	100	Fjord coastal water	BBL6SCM10.63,BBL1SF10.64,HL2SF10.18,BBL1SF10.37,BBL1SF10.21,BBL6SF10.05,BBL6SF10.91,BBL6SF10.29,HL5aSF10.45,HL5aSF10.18,BBL1SF10.03,HL2SCM04.56,BBL1SF04.18,BBL3SCM04.30,BBL1SF10.75,BBL1SF10.15,HL5aSCM10.17,HL2SF04.33,HL5aSCM10.21,BBL1SF04.04
CHOANOFLAGELLATES					
KC488356	BBL1SCM04.18	Uncultured eukaryote clone NPK2_136 (EU371175)	99	Arctic sea water	HL5aSCM04.35, HL2SCM04.03, HL2SCM04.67, BBL1SCM04.18
KC488357	BBL1SF04.15	Uncultured eukaryote clone clone 4-H9(FN690484)	97	Gulf of Finland	BBL1SF04.15, BBL1SCM04.48
KC488358	BBL1SF04.95	Uncultured eukaryote clone SHAA582 (JQ226502)	96	Saanich Inlet	BBL1SF04.95
KC488359	BBL3SCM04.79	Uncultured eukaryote clone NPK2_136 (EU371175)	99	Arctic sea water	BBL3SCM04.79
KC488360	HL2SCM04.51	Uncultured eukaryote clone NPK2_136 (EU371175)	95	Arctic sea water	HL2SCM04.51
KC488361	HL2SCM04.78	Uncultured eukaryote clone DSGM-66 (AB275066)	99	Sagami Bay	HL2SCM04.78
CILIOPHORA					
KC488362	BBL1SCM10.54	*Pseudotontonia simplicidens* (FJ422993)	99	Chinese waters	HL5aSF10.20, BBL1SF10.09, HL5aSF10.77, BBL1SCM10.54
KC488363	BBL1SCM10.73	Uncultured eukaryote clone SCM16C45 (AY665090)	90	Sargasso Sea	BBL1SCM10.73
KC488364	BBL1SCM10.79	Uncultured eukaryote clone A3 (FN263264)	97	Baltic Sea	BBL1SCM10.79
KC488365	BBL1SF10.19	Uncultured eukaryote clone NPKS2_T13 (EU371394)	99	Kongsfjorden	HL5aSCM04.32, BBL1SF10.54, BBL1SF10.19
KC488366	BBL1SF10.43	*Dadayiella ganymedes* strain DPX7 (JX101853)	94	China Bay	BBL1SF10.43
KC488367	BBL1SF10.57	Uncultured eukaryote clone SCM15C2 (AY665093)	98	Sargasso Sea	BBL1SF10.57
KC488368	BBL1SF10.66	Uncultured eukaryote clone A3 (FN263264)	99	Baltic Sea	BBL1SF10.66
KC488369	BBL3SCM04.07	Uncultured eukaryote clone: RM2-SGM10 (B505518)	87	microbial mats	BBL3SCM04.07
KC488370	BBL6SCM10.28	*Strombidium* sp. JG-2011a clone 30(JX178850)	94	China: Yantai	BBL6SCM10.28
KC488371	BBL6SF10.03	Uncultured eukaryote clone SCM28C124 (AY665085)	99	Sargasso Sea	BBL6SF10.03
KC488372	BBL6SF10.94	Uncultured eukaryote clone SCM28C124 (AY665085)	99	Sargasso Sea	BBL6SCM10.16, HL2SCM10.03, BBL6SF10.96, HL5aSF10.51, BBL6SF10.94, BBL1SCM10.82
KC488373	HL2SCM04.64	Uncultured eukaryote clone CC02A175.103 (JX188358)	99	South China Sea	HL2SCM04.64
KC488374	HL2SCM10.12	*Strombidium* sp. JG-2011a clone 30 (JX178850)	90	China: Yantai	HL2SCM10.12
KC488375	HL2SCM10.53	Uncultured eukaryote clone SCM16C45 (AY665090)	98	Sargasso Sea	HL2SCM10.53
KC488376	HL2SF04.06	Uncultured eukaryote clone SIF_1F2 (EF527106)	99	Fjord Norway	HL2SF04.06

KC488377	HL2SF04.19	Uncultured eukaryote clone CNCIII05.210 (HM581712)	99	Arctic Ocean	HL2SF04.19
KC488378	HL2SF04.75	Uncultured alveolate clone 1b-G6 (FN689960)	99	Gulf of Finland	HL2SF04.75, BBL1SF04.11
KC488379	HL2SF10.12	Uncultured eukaryote clone UEPACAIp5 (AY129053)	98	Pacific Ocean	HL2SF10.12
KC488380	HL2SF10.57	*Mesodinium* sp. LGC-2011 (JN412737)	99		HL5aSF04.05, HL2SF10.57, HL2SCM04.22, BBL1SF04.57, BBL1SCM04.43, BBL1SCM04.39, BBL1SCM04.17
KC488381	HL2SF10.75	Uncultured eukaryote clone SCM16C50 (AY664915)	91	Sargasso Sea	HL2SF10.75
KC488382	HL4SCM04.05	*Strombidium* sp. JG-2011a clone 30 (JX178850)	96	China: Yantai	HL5aSF10.39, HL5aSF10.03, HL4SCM04.05
KC488383	HL5aSCM04.33	Uncultured eukaryote clone SCM15C31 (AY665101)	95	Sargasso Sea	HL5aSCM04.33
KC488384	HL5aSCM04.34	*Rimostrombidium veniliae* (FJ876964)	94	China: Guangzhou	HL5aSCM04.34
KC488385	HL5aSCM10.06	Uncultured eukaryote clone NW414.24 (DQ120009)	99	Arctic ocean	HL5aSCM10.06
KC488386	HL5aSCM10.20	Uncultured eukaryote clone WC-11-MC295-OTU-4 (KC771210)	99	Bering Sea	HL5aSCM10.20, HL2SF04.10
KC488387	HL5aSCM10.24	Uncultured ciliate clone LC22_5EP_20 (DQ504324)	97	Lost City	HL5aSCM10.24
KC488388	HL5aSCM10.27	Uncultured eukaryote clone SCM28C124 (AY665085)	99	Sargasso Sea	HL5aSCM10.27
KC488389	HL5aSCM10.77	Uncultured eukaryote clone CC02A105.053 (JX188331)	95	South China Sea	HL5aSCM10.77
KC488390	HL5aSF04.16	Uncultured eukaryote clone CNCIII51_94 (HM581786)	99	Arctic Ocean	HL5aSF04.16
KC488391	HL5aSF04.59	Uncultured eukaryote clone SHAU437 (JQ222913)	93	Pacific Ocean	HL5aSF04.59
KC488392	HL5aSF10.08	Uncultured eukaryote clone SCM28C124(AY665085)	99	Sargasso Sea	HL5aSF10.08
KC488393	HL5aSF10.09	Uncultured eukaryote clone E_MO_2C12_OTU32 (JN832730)	99	Norwegian Sea	HL5aSF10.09, HL2SF10.51
KC488394	HL5aSF10.34	*Pseudotontonia simplicidens* (FJ422993)	94	Chinese waters	HL5aSF10.34
KC488395	HL5aSF10.60	Uncultured eukaryote clone E_MO_1E09_OTU31 (JN832729)	97	Norwegian Sea	HL5aSF10.60, HL2SF10.05
KC488396	HL5aSF10.73	Uncultured eukaryote clone UEPACAIp5 (AY129053)	97	Pacific Ocean	HL5aSF10.73
		CRYPTOPHYTA			
KC488397	BBL1SCM10.34	*Teleaulax minuta* strain Cr8EHU (JQ966996)	99	Spain water	BBL1SF04.37, HL2SCM10.02, BBL1SCM10.34
KC488398	BBL3SCM04.60	*Hemiselmis cryptochromatica* (AM901354)	100	Maine, West Boothbay Harbor	HL4SCM04.70, HL2SCM04.76, HL5aSF04.74, HL5aSCM04.63, BBL1SF04.17, HL2SF04.68, HL2SF04.01, BBL3SCM04.60, BBL1SCM04.81
KC488399	BBL6SCM04.06	*Rhodomonas* sp. RCC2020 (JN934672)	99	Beaufort Sea	BBL6SCM04.06, BBL1SF04.88, BBL1SF04.06
KC488400	BBL6SCM10.23	*Teleaulax gracilis* strain Cr7EHU (JQ966995)	99	Spain water	HL5aSF10.36, HL2SF10.77, HL2SF10.55, HL2SF04.56, HL2SF04.51, HL2SF04.14, BBL6SCM10.52, HL2SCM04.62, BBL6SCM10.23, BBL3SF04.90
KC488401	HL2SCM10.09	Uncultured eukaryote clone E_MO_2B01_OTU37 (JN832734)	99	Norwegian Sea	BBL3SCM04.04, HL2SCM10.68, HL2SCM10.09, BBL3SCM04.54, BBL1SF10.88, BBL1SF10.80, BBL1SF10.63
		DINOFLAGELLATES			
KC488402	BBL1SCM04.11	Uncultured marine eukaryote clone CNCIII05_232 (HM581744)	99	Arctic Ocean	HL5aSCM10.76, BBL1SCM10.89, BBL1SCM04.11
KC488403	BBL1SCM04.14	Uncultured eukaryote clone SIF_1G6 (EF527108)	99	Fjord Norway	BBL1SCM04.14
KC488404	BBL1SCM10.44	Uncultured eukaryote clone CNCIII51_35 (HM581766)	98	Arctic Ocean	BBL1SCM10.44
KC488405	BBL1SCM10.62	Uncultured eukaryote clone AMT15_33B_22 (EU780630)	99	Ocean water	BBL1SCM10.62
KC488406	BBL1SCM10.87	*Pentapharsodinium tyrrhenicum* (AF022201)	98		BBL1SCM10.87
KC488407	BBL1SCM10.93	*Prorocentrum rhathymum* strain S7 (EU287487)	95		BBL1SCM10.93, BBL1SCM10.35
KC488408	BBL1SF04.03	Uncultured syndiniales clone PROSOPE.E3-25m.129 (EU793218)	99	Mediterranean Sea	BBL1SF04.03
KC488409	BBL1SF04.38	Uncultured eukaryote clone SSRPD84 (EF172986)	99	Sargasso Sea	BBL1SF04.38
KC488410	BBL1SF10.04	Uncultured eukaryote clone MAL106.6.A8(JQ956272)	95	Baffin Bay (Arctic)	BBL1SF10.04
KC488411	BBL1SF10.39	Uncultured eukaryote clone SCM15C24 (AY664889)	99	Sargasso Sea	BBL1SF10.39, BBL1SCM10.78
KC488412	BBL1SF10.50	*Lepidodinium chlorophorum* (AY331681)	99		HL5aSCM10.42, BBL1SF10.50, BBL1SF10.29
KC488413	BBL1SF10.86	*Karlodinium micrum* isolate NEPCC736 (EF492490)	99		BBL1SF10.86
KC488414	BBL6SCM04.27	*Pentapharsodinium tyrrhenicum* (AF022201)	97		BBL6SCM04.27

KC488415	BBL6SCM10.20	Uncultured marine eukaryote clone SIF.1F6 (EF527107)	99	Fjord Norway	BBL6SCM10.20
KC488416	BBL6SCM10.62	*Gyrodinium fusiforme* (AB120002)	96		BBL6SCM10.62
KC488417	BBL6SCM10.78	Uncultured marine eukaryote clone CNCIII51.25 (HM581763)	96	Arctic Ocean	BBL6SCM10.78
KC488418	BBL6SCM10.91	Uncultured eukaryote clone SCM15C29 (AY665034)	97	Sargasso Sea	BBL6SCM10.91
KC488419	BBL6SF04.52	Uncultured eukaryote clone SCM37C27 (AY664962)	99	Sargasso Sea	BBL6SF04.52
KC488420	BBL6SF04.62	Uncultured eukaryote clone AMT15_27B_4 (EU780618)	97	Ocean water	BBL6SF04.62
KC488421	BBL6SF04.81	Uncultured marine alveolate clone North Pole SW70_28 (HQ438156)	99	North Pole Sea Ice	BBL6SF04.81
KC488422	HL2SCM10.01	Uncultured eukaryote clone AMT15_27B_4 (EU780618)	97	Ocean water	HL2SCM10.01
KC488423	HL2SCM10.05	*Dinophyceae sp.* CCMP1878 (AY251287)	99		HL2SCM10.05
KC488424	HL2SCM10.11	*Heterocapsa rotundata* strain CCCM 680 (AF274267)	98		HL2SCM10.11
KC488425	HL2SCM10.15	Uncultured marine eukaryote clone CNCIII51_25 (HM581763)	99	Arctic ocean	HL2SCM10.15, BBL1SF10.28, BBL1SCM10.68
KC488426	HL2SCM10.62	*Lepidodinium chlorophorum* (AY331681)	97		HL2SCM10.62
KC488427	HL2SCM10.65	Uncultured marine alveolate clone North Pole SW70_44 (HQ438161)	97	North Pole Sea Ice	HL2SCM10.65
KC488428	HL2SF10.22	Uncultured eukaryote clone SCM28C103 (AY665030)	97	Sargasso Sea	HL2SF10.22, HL2SCM04.73, BBL1SCM10.39, BBL1SCM04.20
KC488429	HL2SF10.24	*Katodinium rotundatum* strain ccmp1542 (DQ388464)	96		HL2SF10.24
KC488430	HL2SF10.37	Uncultured eukaryote clone SCM28C103 (AY665030)	98	Sargasso Sea	HL2SF10.46,BBL6SCM10.30,HL5aSF10.27,BBL6SCM04.08,HL2SF10.29,HL5aSF10.71,HL5aSF10.01,BBL6SF04.01,BBL6SCM10.01,HL2SF10.37,BBL6SCM04.60,BBL1SCM10.24
KC488431	HL2SF10.40	*Gymnodinium sp.* QZ-2012 (JQ639761)	97		HL2SF10.40
KC488432	HL2SF10.45	Uncultured eukaryote clone CNCIII05_240 (HM581707)	96	Arctic Ocean	HL2SF10.45
KC488433	HL2SF10.47	Uncultured eukaryote clone CNCIII05_20 (HM581705)	98	Arctic ocean	HL2SF10.47
KC488434	HL4SF04.01	Uncultured marine eukaryote clone SA7.56 (FJ626722)	100	Antarctic	HL5aSF10.59,BBL1SF04.29,HL4SCM04.25,HL4SCM04.51,HL4SCM04.12,HL4SCM04.13,HL4SCM04.34,HL4SF04.01,BBL6SF04.39,BBL6SF04.10,BBL3SF04.21,BBL1SCM10.09
KC488435	HL4SF04.27	Uncultured eukaryote clone CNCIII05_222 (HM581706)	98	Arctic Ocean	HL4SF04.27, HL4SF04.26
KC488436	HL4SF04.63	Uncultured eukaryote clone SCM27C15 (AY664924)	98	Sargasso Sea	HL4SF04.63
KC488437	HL5aSCM04.06	Uncultured eukaryote clone SCM38C10 (AY664972)	99	Sargasso Sea	HL5aSCM04.06, HL2SCM04.16
KC488438	HL5aSCM04.13	Uncultured eukaryote clone SCM15C8 (AY664881)	99	Sargasso Sea	HL5aSCM04.13
KC488439	HL5aSCM04.20	Uncultured eukaryote clone E_MO_3D02_OTU27 (JN832724)	99	Norwegian Water	HL5aSCM04.20
KC488440	HL5aSCM04.21	Uncultured eukaryote clone SCM15C8 (AY664881)	99	Sargasso Sea	HL5aSCM04.21, HL2SCM04.10
KC488441	HL5aSCM04.23	Uncultured marine eukaryote clone CNCIII05_240 (HM581707)	99	Arctic Ocean	HL2SCM04.91,HL5aSF04.73,HL2SF10.48,BBL3SCM04.93,BBL3SCM04.77,HL5aSCM04.30,HL5aSCM04.50,HL5aSCM10.05,HL5aSCM04.28,HL5aSCM04.23,BBL3SCM04.71,BBL1SCM10.17
KC488442	HL5aSCM10.03	*Gyrodinium cf. gutrula* (FN669511)	99		HL5aSCM10.03
KC488443	HL5aSCM10.07	Uncultured eukaryote clone CNCIII05_222 (HM581706)	98	Arctic Ocean	HL5aSCM10.07
KC488444	HL5aSCM10.78	Uncultured eukaryotes clone LC22_5EP_33 (DQ504325)	99	Mid-Atlantic	HL5aSCM10.78
KC488445	HL5aSCM10.90	Uncultured eukaryote clone CNCIII51_20 (HM581767)	91	Arctic Ocean	HL5aSCM10.90
KC488446	HL5aSCM10.93	Uncultured alveolate clone 3c-F2 (FN690153)	99	Baltic Sea	HL5aSCM10.93
KC488447	HL5aSF04.51	Uncultured eukaryote clone SCM28C40 (AY664893)	93	Sargasso Sea	HL5aSF04.51
KC488448	HL5aSF10.22	*Gyrodinium fusiforme* (AB120002)	99		HL5aSF10.22
FUNGI					
KF761280	BBL1SF04.25	Uncultured eukaryote clone DSGM-64 (AB275064)	99	Sagami Bay	BBL1SF04.25, BBL1SF04.43
KF761282	BBL3SF04.22	Uncultured eukaryote clone DSGM-64 (AB275064)	99	Sagami Bay	BBL1SCM04.23, BBL3SF04.22
KF761283	BBL3SF04.28	Uncultured eukaryote clone DSGM-64 (AB275064)	99	Sagami Bay	BBL3SF04.28, BBL1SCM04.58
KF761286	BBL3SCM04.08	Uncultured eukaryote clone DSGM-64 (AB275064)	99	Sagami Bay	BBL3SCM04.08, BBL3SCM04.31

HAPTOPHYTA					
KC488449	BBL1SF10.59	uncultured marine picoeukaryote ws_138,clone 1804E01 (FR874463)	99	fjord water	HL5aSF04.56, HL2SF10.68, HL5aSF10.38, HL2SF10.02, BBL1SF10.59
KC488450	BBL6SF04.91	*Phaeocystis* sp. JD-2012 RCC908 (JX660990)	99	Pacific Ocean	HL5aSCM04.16, BBL6SF04.91, BBL6SCM04.91
KC488451	HL2SCM10.72	*Haptolina fragaria* strain UIO029 (AM491013)	99		HL5aSCM04.81, HL2SCM10.72, BBL1SF04.09
KC488452	HL5aSCM04.11	*Phaeocystis pouchetii* isolate P360 (AF182114)	99	Norway Water	HL5aSF10.49, HL5aSCM04.11, HL2SCM04.84, HL5aSCM04.12, HL2SCM04.77, BBL6SF04.72, HL4SF04.33, BBL6SF04.36
KC488453	HL5aSCM10.88	Uncultured *Chrysochromulina* clone Biosope_T60.034 (FJ537341)	99	Pacific Ocean	HL5aSCM10.88, HL2SF10.52, BBL6SCM04.40, BBL1SCM10.92, BBL1SCM10.12
KC488454	HL5aSF04.03	*Phaeocystis pouchetii* isolate P360 (AF182114)	99	Norway water	HL5aSF04.03, HL4SCM04.47
KC488455	HL5aSF04.27	Uncultured eukaryote clone AMT15_15_10m_412 (GQ863817)	98	Atlantic Meridional Transect	HL5aSF04.27, BBL1SCM04.09
KC488456	HL5aSF04.58	Uncultured marine eukaryote CN207St155 (HM581637)	99		HL5aSF04.58, BBL6SF10.13
KC488457	HL5aSF10.47	Uncultured eukaryote Clone B5 (FN263276)	99	Poland:Gulf of Gdansk	HL5aSF10.47, HL5aSF10.33, HL5aSF10.93, HL2SCM04.80, BBL6SF10.11, BBL1SCM10.84
KATABLEPHARIDS					
KF761288	HL5aSF10.02	*Leucocryptos marina* clone lmssu5 (DQ980481)	99		HL5aSF10.85, HL5aSF10.02
KF761290	BBL1SCM10.21	Uncultured eukaryote clone E_MO_3H06_OTU36 (JN832733)	97	Norwegian Sea	BBL1SCM10.74, BBL1SCM10.21
MARINES ALVEOLATES (MALV)					
KC488458	BBL1SCM04.32	Uncultured eukaryote clone DSGM-28 (AB275028)	98	Sagami Bay	BBL1SCM04.32
KC488459	BBL1SCM04.54	Uncultured Marine picoeukaryotes ws_96, clone 1802G10 (FR874416)	99	fjord water	HL4SF04.89, BBL6SCM04.69, BBL1SCM04.54
KC488460	BBL1SCM10.31	Uncultured eukaryote clone CC02SE75.082 (KF031707)	96	South China Sea	BBL1SCM10.31
KC488461	BBL1SCM10.67	Uncultured Alveolate clone RS.12f.10m.163(KC582951)	97	Red Sea	BBL1SCM10.67
KC488462	BBL1SCM10.69	Uncultured eukaryote clone CC02SE75.082 (KF031707)	96	South China Sea	BBL1SCM10.69
KC488463	BBL1SF10.41	Uncultured eukaryote clone SSRPB75 (EF172792)	95	Sargasso Sea	BBL1SF10.41
KC488464	BBL1SF10.47	Uncultured eukaryote clone AMT15_1B_36 (EU780604)	99	ocean water	BBL1SF10.47
KC488465	BBL3SCM04.65	Uncultured eukaryote ws_96, clone 1802G10 (FR874416)	99	fjord water	HL4SCM04.03, BBL6SF10.49, HL5aSF10.68, BBL3SCM04.65
KC488466	BBL3SF04.19	Uncultured eukaryote clone CC02A175.077 (JX188346)	93	South China Sea	BBL3SF04.19
KC488467	BBL3SF04.51	Uncultured eukaryote ws_101, clone 1807F06 (FR874623)	99	fjord water	BBL3SF04.51, BBL3SF04.16
KC488468	BBL3SF04.70	Uncultured eukaryote ws_138, clone 1804C07 (FR874451)	100	fjord water	HL5aSF10.88,HL5aSF04.82,HL2SCM04.49,HL5aSF10.75,BBL3SF04.62,BBL3SCM04.84,HL5aSF10.42,BBL3SF04.70,BBL6SF10.14,BBL3SF04.67,BBL3SF04.60,BBL6SF10.44,BBL3SCM04.20
KC488469	BBL6SCM04.48	Uncultured eukaryote ws_138, clone 1804C07 (FR874451)	99	fjord water	HL5aSF04.01, HL2SF04.87, BBL3SF04.69, BBL3SCM04.27, HL5aSCM04.42, BBL6SCM04.48, BBL1SF04.10, HL5aSF10.05, BBL1SCM04.84
KC488470	BBL6SCM10.04	Uncultured alveolate clone RA010412.5 (DQ186532)	96		BBL6SCM10.04
KC488471	BBL6SCM10.50	Uncultured eukaryote clone SSRPD71 (EF172955)	94	Sargasso Sea	BBL6SCM10.50
KC488472	BBL6SF10.17	Uncultured eukaryote clone CC02A175.068 (KF031839)	99	South China Sea	BBL6SF10.17, BBL1SF10.25, BBL1SF10.08
KC488473	BBL6SF10.19	Uncultured eukaryote clone CC02SE75.082 (KF031707)	99	South China Sea	BBL6SF04.03, BBL6SCM04.16, BBL6SF10.19, BBL3SCM04.87
KC488474	BBL6SF10.24	Uncultured eukaryote clone CC02A740.034 (JX188380)	99	South China Sea	HL4SF04.54, HL5aSCM10.48, HL5aSF10.80, BBL6SCM04.13, HL2SCM04.65, HL4SF04.83, HL5aSF10.53, BBL6SF10.24, HL5aSF10.12, BBL6SCM04.03
KC488475	BBL6SF10.27	Uncultured alveolate clone RA001219.16 (DQ186528)	99		BBL6SF10.27, BBL3SCM04.36
KC488476	BBL6SF10.35	Uncultured eukaryote clone SHBB427 (JQ223103)	99	Saanich Inlet	BBL6SF10.35, BBL6SF10.06, BBL3SF04.65, BBL3SCM04.46
KC488477	BBL6SF10.63	Uncultured alveolate clone RS.12f.10m.00141 (KC582884)	98	Red Sea	BBL6SF10.63, BBL1SCM04.77
KC488478	BBL6SF10.77	Uncultured eukaryote clone SCM37C36 (AY664989)	99	Sargasso Sea	HL4SF04.88, HL4SF04.55, HL5aSF10.90, BBL6SF10.77
KC488479	BBL6SF10.90	uncultured marine picoeukaryotes ws_96, clone	100	fjord water	HL2SCM04.88, HL4SF04.34, BBL6SF10.90

		1802G10 (FR874416)			
KC488480	HL2SCM04.27	Uncultured alveolate clone 3c-B4 (FN690288)	96	Sweden Sea	HL2SCM04.27
KC488481	HL2SCM04.63	Uncultured alveolate clone 3c-B4 (FN690288)	99	Sweden Sea	HL2SCM04.63, BBL6SF10.10
KC488482	HL2SCM10.08	*Syndinium* sp. ex Corycaeus sp. (DQ146406)	99		HL2SCM10.08
KC488483	HL2SCM10.59	Uncultured alveolate clone RS.12f.10m.2 (KC582943)	96	Red Sea	HL2SCM10.59, BBL6SF10.60
KC488484	HL2SF10.04	Uncultured eukaryote clone Q2A12N10 (EF172973)	93	Sargasso Sea	HL2SF10.04, BBL6SF10.38
KC488485	HL2SF10.16	Uncultured eukaryote clone CC02SE75.027 (KF031679)	94	South China Sea	HL2SF10.16
KC488486	HL2SF10.20	Uncultured alveolate clone RA010412.70 (DQ186533)	99		HL4SF04.82, HL2SF10.20, BBL6SF10.72
KC488487	HL2SF10.35	Uncultured eukaryote clone Q2B03N5 (EF173008)	99	Sargasso Sea	HL2SF10.35
KC488488	HL2SF10.39	Uncultured phytoplankton clone Q1-11 (JQ420088)	94	Sieburth in China	HL2SF10.39
KC488489	HL2SF10.44	Uncultured eukaryote clone CC02SE75.027 (KF031679)	99	South China Sea	HL2SF10.44, BBL6SF10.20, BBL3SCM04.53, BBL1SF10.13
KC488490	HL2SF10.49	Uncultured alveolate clone RS.12f.10m.95 (KC582949)	89	Red Sea	HL2SF10.49
KC488491	HL2SF10.63	Uncultured eukaryote clone NPK2_163 (EU371150)	98	Arctic waters	HL2SF10.63
KC488492	HL2SF10.70	Uncultured alveolate clone RS.12f.10m.0035 (KC582892)	97	Red Sea	HL2SF10.70
KC488493	HL2SF10.78	Uncultured syndiniales clone PROSOPE.ED-50m.30 (EU793718)	97		HL2SF10.78
KC488494	HL2SF10.84	Uncultured eukaryote clone UEPAC41p3 (AY129042)	97	Pacific Ocean	HL2SF10.84
KC488495	HL2SF10.86	Uncultured alveolate clone 3c-B4 (FN690288)	98	Baltic Sea	HL2SF10.86
KC488496	HL2SF10.88	Uncultured eukaryote clone UEPAC41p3 (AY129042)	97	Pacific Ocean	HL2SF10.88
KC488497	HL2SF10.95	Uncultured eukaryote clone Q2A12N10 (EF172973)	94	Sargasso Sea	HL2SF10.95, BBL6SF10.56
KC488498	HL2SF10.96	Uncultured eukaryote clone Q2A12N10 (EF172973)	99	Sargasso Sea	HL5aSF10.67, BBL1SCM10.37, HL2SF10.96, BBL6SF10.32, BBL1SCM10.19, HL2SF10.30, BBL1SCM10.88, BBL1SCM10.13
KC488499	HL4SF04.23	Uncultured eukaryote clone MAL106_6_H2 (JQ956297)	99	Baffin Bay Arctic	HL4SF04.23
KC488500	HL4SF04.67	Uncultured eukaryote clone CC02A740.049 (KF031896)	99	South China Sea	HL4SF04.67
KC488501	HL4SF04.75	Uncultured marine alveolate Group II DH147-EKD6 (AF290068)	94	Antarctica	HL4SF04.75
KC488502	HL5aSCM04.05	Uncultured alveolate clone RS.12f.10m.163 (KC582951)	99	Red Sea	HL2SCM10.16, BBL6SF10.89, BBL6SCM10.26, HL5aSCM04.05, HL5aSF10.10, HL5aSCM04.01, BBL6SCM04.04
KC488503	HL5aSCM04.08	Uncultured alveolate clone RS.12f.10m.00117 (KC582893)	91	Red Sea	HL5aSCM04.08
KC488504	HL5aSCM04.19	Uncultured eukaryote clone SSRPC30 (EF172883)	99	Sargasso Sea	HL5aSCM04.19, BBL6SF10.80
KC488505	HL5aSCM04.39	Uncultured eukaryote clone CC02A175.020 (JX188344)	96	South China Sea	HL5aSCM04.39
KC488506	HL5aSCM04.47	Uncultured alveolate clone RS.12f.10m.0016 (KC582890)	98	Red Sea	HL5aSCM04.47
KC488507	HL5aSCM04.53	Uncultured eukaryote clone UEPACDp5 (AY129058)	98	Pacific Ocean	HL5aSCM04.53
KC488508	HL5aSCM10.02	Uncultured eukaryote clone CC02A175.077 (JX188346)	99	South China Sea	HL5aSCM10.02, HL4SF04.61, HL4SF04.45
KC488509	HL5aSCM10.33	Uncultured alveolate clone RA010412.5 (DQ186532)	95		HL5aSCM10.33
KC488510	HL5aSCM10.41	Uncultured eukaryote clone CC02A740.105 (JX188381)	99	South China Sea	HL5aSCM04.72, HL5aSCM10.41, BBL6SCM04.38, HL5aSF10.19, HL4SF04.28, BBL6SF10.54, BBL1SF04.07, BBL1SF04.02
KC488511	HL5aSCM10.68	Uncultured eukaryote clone CC02A175.003 (JX188333)	99	South China Sea	HL5aSCM10.68
KC488512	HL5aSF10.40	Uncultured alveolate clone RS.12f.10m.114 (KC582922)	93	Red Sea	HL5aSF10.40, BBL1SF10.48
KC488513	HL5aSF10.58	Uncultured eukaryote clone SCM16C19 (AY664988)	97	Sargasso Sea	HL5aSF10.58, HL2SCM04.70
KC488514	HL5aSF10.70	Uncultured eukaryote clone ws_164, clone 1816F11 (FR874821)	95	fjord water	HL5aSF10.70, BBL1SF04.59
KC488515	HL5aSF10.86	Uncultured eukaryote clone CC02A175.020 (JX188344)	94	South China Sea	HL5aSF10.86, HL2SF10.09
MARINES STRAMENOPILES (MAST) – MAST-1					

KC488516	BBL1SF04.12	Uncultured marine eukaryote clone CD8.09 (DQ121426)	99	Sea Water	HL5aSF04.15, HL5aSCM04.44, BBL6SCM04.52, BBL1SF04.12, BBL1SF04.77, BBL1SF04.05, BBL1SCM04.86, BBL1SCM04.92, HL5aSF04.26, BBL1SCM04.80
KC488517	BBL1SF04.74	Uncultured stramenopile clone 155D3Be4vg (JQ782004)	99	Pacific Ocean	BBL6SF04.30, BBL3SCM04.01, HL4SF04.91, BBL1SCM04.40, BBL6SF04.85, HL2SCM04.29, BBL1SF04.74, BBL1SCM04.15
KC488518	BBL1SF04.80	Uncultured stramenopile clone 155D3Be4vg (JQ782004)	99	Pacific Ocean	HL2SF10.01, HL4SF04.93, BBL1SF04.80
KC488519	HL4SF04.91	Uncultured stramenopile clone 155D3Be4vg (JQ782004)	99	North Pacific Ocean	HL4SF04.91
KC488520	HL5aSCM10.37	Uncultured eukaryote clone AMT15_15_10m_378 (GQ863791)	99	Atlantic Meridional Transect	HL5aSCM10.37
KC488521	HL5aSF04.94	Uncultured marine eukaryote clone AD6.02 (DQ121419)	99	seawater	HL4SF04.37,HL4SF04.09,HL5aSCM04.25,BBL3SCM04.56,HL2SF04.12,HL2SF04.27 ,HL5aSF04.37,HL2SCM04.81,HL5aSF04.94,HL2SF04.20,BBL3SCM04.40
MARINES STRAMENOPILES (MAST)–MAST-3					
KC488522	BBL1SCM04.60	Uncultured stramenopile clone RA010516.17 (AY295631)	99	English Channel	BBL1SCM04.60
KC488523	BBL1SF10.16	Uncultured eukaryote clone Q2C03N10 (EF172976)	99	Sargasso Sea	HL2SCM10.39, BBL1SF10.42, HL2SF10.10, BBL1SF10.16, BBL1SCM10.46
KC488524	BBL3SF04.83	Uncultured stramenopile clone 905st23-32 (JQ782094)	99	Pacific ocean	BBL3SF04.83, BBL3SF04.79
KC488525	BBL6SCM04.32	Uncultured stramenopile clone 70S8Be8Op (JQ782032)	99	Pacific ocean	BBL6SCM04.32, BBL1SF04.65
KC488526	BBL6SCM10.72	Uncultured stramenopile clone 70S1Ae4eE (JQ782047)	98	Pacific ocean	BBL6SCM10.72
KC488527	BBL6SF04.07	Uncultured eukaryote clone E_NY_1C02_OTU33 (JN832829)	99	Norwegian Sea	BBL6SF04.07
KC488528	BBL6SF10.62	Uncultured eukaryote clone E_NY_1C02_OTU33 (JN832829)	99	Norwegian Sea	BBL6SF10.62, BBL6SF10.07
KC488529	HL2SF10.28	Uncultured eukaryote clone IND58.19 (EU561890)	99	Indian Ocean	HL2SF10.28
KC488531	HL4SF04.72	Uncultured stramenopile clone 155S1Ae6UF (JQ781958)	99	north pacific ocean	HL4SF04.72
KC488532	HL5aSCM04.09	Uncultured eukaryote clone E_NY_1C02_OTU33 (JN832829)	99	Norwegian Sea	HL5aSCM04.09, BBL1SCM04.59
KC488533	HL5aSCM04.24	Uncultured stramenopile clone 155S1Ae6We (JQ781963)	99	Pacific ocean	HL5aSCM04.24, HL5aSCM04.79, HL5aSCM04.03, HL4SF04.70, BBL6SCM10.53
KC488534	HL5aSF04.08	Uncultured eukaryote clone UEPACRp5 (AY129069)	99	Pacific Ocean	HL5aSF04.08
KC488535	HL5aSF04.61	Uncultured stramenopile clone 905st23-32 (JQ782094)	99	Pacific ocean	HL5aSF04.61, HL5aSF04.55
KC488536	HL5aSF10.63	Uncultured marine eukaryote clone IND58.19 (EU561890)	99	Ocean water	BBL1SF10.22,BBL1SCM10.30,HL2SF10.74,HL2SF10.67,HL2SF10.65,BBL6SCM10.9 5,HL5aSF10.50,BBL1SF10.44,HL5aSF10.63,HL2SF10.89,BBL6SCM10.82,BBL1SF10. 11,BBL1SCM10.90,BBL1SCM10.25,BBL1SCM10.40,BBL1SCM10.01
KC488537	HL5aSF10.64	Uncultured eukaryote clone AMT15_1_10m_135 (GQ863783)	99	Atlantic Meridional Transect	HL2SF10.38, HL5aSF10.48, HL5aSF10.06, HL5aSF10.64, HL5aSCM10.38, HL2SF10.26, HL5aSF10.82, HL5aSF10.37, HL5aSF10.32, BBL1SCM10.03
MARINES STRAMENOPILES(MAST)–MAST-4					
KC488538	BBL1SF10.51	Uncultured marine eukaryote clone IND1.6 (EU561666)	99	Sea Water	BBL1SF10.51
KC488539	BBL6SF10.04	Uncultured stramenopile clone 155D3Ae6i8(JQ782003)	99	Pacific ocean	HL5aSCM10.16, BBL6SF10.04, BBL1SF10.30
KC488540	HL2SCM10.14	Uncultured stramenopile clone 14H3Te6O0 (JQ782065)	99	Pacific ocean	HL2SCM10.14
KC488541	HL5aSCM04.43	Uncultured stramenopile clone 155D1Ae4Ay (JQ782010)	99	Pacific ocean	HL5aSCM04.43, HL2SF10.72
MARINES STRAMENOPILES(MAST)–MAST-7					
KC488542	BBL1SCM04.91	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	98	Fjord water	BBL1SF04.42, BBL6SCM04.33, BBL6SF04.16, BBL1SCM04.91
KC488543	BBL1SF04.33	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	99	fjord water	BBL1SF04.33
KC488544	BBL3SCM04.88	Uncultured marine eukaryote clone BL010320.6 (AY381207)	99	Mediterranean, Blanes Bay	BBL3SF04.88, BBL3SF04.55, HL5aSF04.92, BBL3SCM04.88
KC488545	BBL3SF04.33	Uncultured marine eukaryote clone BL010320.6	99	Mediterranean,	BBL3SF04.33, BBL3SF04.06

		(AY381207)		Blanes Bay	
KC488546	BBL3SF04.76	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	98	fjord water	BBL6SCM04.12, BBL3SCM04.59, BBL1SF04.35, BBL3SF04.76, BBL3SF04.40, BBL3SF04.02, BBL1SCM04.78
KC488547	BBL6SF04.93	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	99	fjord water	HL5aSF04.53, BBL6SCM04.51, BBL6SF04.93, BBL1SF04.83, BBL1SF04.70, BBL1SF04.16
KC488548	HL2SCM04.05	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	98	fjord water	BBL3SCM04.48, HL2SCM04.05, BBL1SF04.26
KC488549	HL2SCM04.26	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	98	Fjord water	HL2SCM04.26, HL2SCM04.09, HL2SCM04.12, HL2SCM04.01
KC488550	HL2SCM04.92	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	100	fjord water	HL5aSF04.46,HL5aSF04.54,BBL6SF04.88,HL4SF04.69,HL4SCM04.20,HL5aSF04.52 ,HL5aSF04.32,BBL6SCM04.75,BBL6SF04.28,HL2SF10.07,HL2SCM04.61,HL2SCM0 4.59,HL2SCM04.75,HL2SCM04.83,HL2SCM04.38,HL5aSF10.54,HL2SCM04.92,HL2 SCM04.33,HL5aSCM04.37,BBL6SF04.80,BBL6SCM04.56,BBL1SCM04.73
KC488551	HL2SF04.70	Uncultured marine eukaryote clone BL010320.6 (AY381207)	99	Mediterranean, Blanes Bay	HL5aSF04.79,BBL1SF04.40,HL5aSF04.66,HL5aSF04.65,HL4SF04.53,BBL1SF04.47, BBL1SF04.20,BBL1SF04.01,BBL1SF04.34,BBL1SF04.24,BBL1SF04.14,BBL3SCM04. 37,BBL3SCM04.23,BBL3SCM04.09,BBL3SCM04.39,BBL6SF04.02,HL5aSF04.91,HL 5aSF04.84,HL2SCM04.58,HL2SF04.70,HL2SF04.03,HL2SCM04.42,HL5aSCM10.10, HL2SF04.63,HL2SF04.38,HL2SF04.53,BBL1SCM04.75,HL5aSCM04.14,HL2SCM04. 18,HL2SCM04.46,HL2SCM04.15,BBL1SCM04.67,HL5aSF04.42,HL5aSF04.41,BBL1 SCM04.41,BBL1SCM04.05
KC488552	HL2SF04.76	Uncultured eukaryote clone BL010320.6 (AY381207)	99	Mediterranean Blanes Bay	HL2SF04.76, BBL3SCM04.72
KC488553	HL4SF04.19	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	99	Fjord water	HL2SCM04.71, BBL6SF04.40, HL5aSF04.40, HL4SF04.19, HL4SF04.05, BBL3SCM04.38, BBL3SCM04.34
KC488554	HL5aSF04.09	uncultured picoeukaryotes ws_59, clone 1806A01 (FR874496)	99	Fjord water	HL5aSF04.09, BBL3SCM04.66
KC488555	HL5aSF04.72	uncultured picoeukaryotes ws_143, clone 1809G01 (FR874674)	99	fjord water	HL5aSF04.72
KC488556	HL5aSF04.80	uncultured picoeukaryotes ws_143, clone 1809G01 (FR874674)	99	Fjord water	HL5aSF04.80, BBL6SF04.26, BBL6SCM10.67, HL5aSF04.45, BBL3SCM04.35, BBL3SCM04.24, BBL1SCM04.10
PICOZOA					
KC488334	BBL1SCM04.08	Uncultured eukaryote clone NW617.02 (DQ060525)	95	Arctic Ocean	BBL1SCM04.08
KC488335	BBL1SF04.52	Picobiliphyte sp. MS609-66 (JN934893)	99	Maine	BBL1SF04.52
KC488336	BBL1SF10.02	Uncultured eukaryote clone MAL106_6_H9 (JQ956302)	98	Arctic sea	BBL1SF10.02
KC488337	BBL6SCM04.10	Uncultured marine eukaryote clone EN351CTD040_4mN11 (EU368038)	99	marine euphotic zone	HL5aSF10.30, BBL6SCM04.10, BBL1SCM10.27
KC488338	HL2SCM10.71	Uncultured eukaryote clone NW617.02 (DQ060525)	99	Arctic Ocean	HL2SCM10.71
KC488339	HL2SF10.08	Uncultured eukaryote clone CC02A175.091 (KF031848)	98	South China sea	HL2SF10.08
KC488340	HL5aSCM04.10	Uncultured eukaryote clone MAL106_6_H9 (JQ956302)	99	Baffin Bay, Arctic	HL5aSCM04.10
KC488341	HL5aSCM04.17	Uncultured eukaryote clone NW617.02 (DQ060525)	99	Arctic Ocean	HL5aSCM04.17, BBL3SF04.86
KC488342	HL5aSCM04.38	Uncultured eukaryote clone MAL106_6_H9 (JQ956302)	97	Baffin Bay, Arctic	HL5aSCM04.38
KC488343	HL5aSCM04.49	Uncultured phototrophic eukaryote clone OR000415.159 (DQ222875)	100	North Atlantic Ocean	HL5aSCM04.49
KC488344	HL5aSF04.30	Picobiliphyte sp. MS609-66 (JN934893)	99	USA: Maine	HL5aSF04.30
KC488345	HL5aSF04.93	Uncultured marine eukaryote clone EN351CTD039_30mN9 (EU368037)	98	marine euphotic zone	HL5aSF04.93, HL5aSCM04.26, HL2SF04.32
RHIZARIA					
KC488557	BBL1SCM04.28	Uncultured cercozoan clone 7-B3 (FN690367)	99	Baltic Sea	BBL1SF04.68, HL2SF10.03, BBL1SF04.19, BBL1SCM04.28
KC488558	BBL1SCM04.63	Uncultured eukaryote Clone DSGM-45 (AB275045)	99	Sagami Bay	BBL3SCM04.52, HL4SCM04.49, BBL3SCM04.80, BBL1SCM04.63
KC488559	BBL1SCM04.83	Uncultured eukaryote Clone CYSGM-12 (AB275095)	99	Sagami Bay	BBL1SCM04.83
KC488560	BBL1SCM10.60	Uncultured eukaryote Clone CYSGM-15 (AB275098)	95	Sagami Bay	BBL1SCM10.60

KC488561	BBL1SCM10.84	Uncultured cercozoan clone 3b-F5 (FN690383)	96	Sweden Sea	BBL1SCM10.84
KC488562	BBL1SF04.90	Uncultured eukaryote clone NB30_23A9G (JN048125)	99	North Baffin Bay	BBL1SF04.90, BBL1SF04.53
KC488563	BBL1SF10.49	Uncultured eukaryote clone NOR46.27 (DQ314814)	99	arctic waters	BBL1SF10.49
KC488564	BBL3SCM04.21	Unnamed chlorarachniophyte BC52 (AF076172)	90		BBL3SCM04.21
KC488565	BBL3SCM04.43	Uncultured marine eukaryote NAMAKO-4 (AB252744)	99	Japan Lake	BBL3SCM04.43
KC488566	BBL3SCM04.45	Uncultured marine eukaryote NAMAKO-3 (AB252743)	99	Japan Lake	BBL3SCM04.45
KC488567	BBL3SCM04.82	*Cryothecomonas* sp. APCC MC5-1Cryo (GQ144679)	98		HL4SCM04.37, HL2SF10.27, HL4SCM04.90, BBL3SCM04.82
KC488568	BBL3SF04.23	*Minorisa minuta* isolate 6C1 (JX272635)	92	Spain: Blanes Bay	HL2SCM04.14, BBL3SF04.23, BBL3SF04.15
KC488569	BBL3SF04.39	*Minorisa minuta* isolate 6C1 (JX272635)	92	Spain: Blanes Bay	BBL3SF04.45, BBL3SF04.39, BBL3SF04.10, BBL3SCM04.63
KC488570	BBL6SCM04.44	Uncultured eukaryote clone DSGM-49 (AB275049)	98	Sagami Bay	BBL6SCM04.44
KC488571	BBL6SCM04.67	Uncultured eukaryote clone DSGM-49 (AB275049)	98	Sagami Bay	BBL6SCM04.67
KC488572	BBL6SCM04.74	Uncultured eukaryote clone DSGM-49 (AB275049)	98	Sagami Bay	BBL6SCM04.74
KC488573	BBL6SF04.08	Uncultured cercozoanclone 3b-F5 (FN690383)	93	Baltic Sea	BBL6SF04.08
KC488574	BBL6SF04.09	Uncultured eukaryote clone NAMAKO-8 (AB252748)	98	Japan Lake	BBL6SF04.09
KC488575	BBL6SF04.17	uncultured marine picoeukaryote ws_96, clone 1802E03 (FR874390)	98	fjord water	BBL6SF04.17
KC488576	BBL6SF04.18	Uncultured eukaryote clone NOR46.14 (DQ314811)	96	arctic waters	BBL6SF04.18
KC488577	BBL6SF10.84	Uncultured eukaryote clone UEPACOp5 (AY129068)	95	Pacific Ocean	BBL6SF10.84
KC488578	HL2SCM04.02	Uncultured eukaryote clone DSGM-47 (AB275047)	99	Sagami Bay	HL2SCM04.02, BBL3SCM04.94, BBL1SCM04.87
KC488579	HL2SCM04.21	Uncultured cercozoan clone 3b-F5 (FN690383)	98	Baltic Sea	HL2SCM04.21
KC488580	HL2SCM04.45	Uncultured eukaryoteClone CYSGM-12 (AB275095)	99	Sagami Bay	HL2SCM04.45, HL2SCM04.36
KC488581	HL2SCM04.74	Uncultured eukaryote clone NB30_9B8D (JN048124)	99	North Baffin Bay	HL2SCM04.74
KC488582	HL2SCM10.63	Uncultured eukaryote clone DSGM-49 (AB275049)	98	Sagami Bay	BBL6SCM04.21, HL2SCM10.63, BBL6SCM04.58, BBL3SCM04.13
KC488583	HL2SF04.26	Uncultured eukaryote clone NOR46.27 (DQ314814)	99	Arctic ocean	HL2SF04.26
KC488584	HL4SF04.20	Uncultured archaeon clone GG101008Arch13 (JN591988)	99	Puget Sound	HL4SF04.20
KC488585	HL4SF04.25	Uncultured eukaryote clone NPK2_163 (EU371150)	92	Arctic sea	HL4SF04.25
KC488586	HL4SF04.95	Uncultured marine eukaryote clone MO010.880.0133 (GU246586)	95	Pacific Ocean	HL4SF04.95
KC488587	HL5aSCM04.27	Uncultured eukaryote clone SHAA473 (JQ226414)	99	Pacific Ocean	HL5aSCM04.18, HL5aSCM04.27, HL5aSCM04.02, BBL1SCM04.16
KC488588	HL5aSCM04.40	Uncultured eukaryote clone NOR46.14 (DQ314811)	99	Artic waters	HL2SF04.16, HL2SF04.13, HL2SF04.34, HL2SCM04.57, HL5aSF04.33, HL5aSCM04.40, HL2SCM04.04
KC488589	HL5aSCM04.59	Uncultured eukaryote Clone DSGM-46 (AB275046)	99	Sagami Bay	HL5aSCM04.59
KC488590	HL5aSCM10.58	Uncult. rhizarian clone A95F15RM3G06(GU822605)	99	Caribbean Sea	HL5aSCM10.58
KC488591	HL5aSCM10.64	Uncult.rhizarian clone BCI3F14RM1E07 (GU820872)	99	Caribbean Sea	HL5aSCM10.64, HL5aSCM10.39, HL5aSCM10.19
KC488592	HL5aSF04.10	Uncultured eukaryote clone DSGM-47 (AB275047)	98	Sagami Bay	HL5aSF04.10
KC488593	HL5aSF04.14	Uncultured eukaryote clone SHAU671 (JQ223021)	97	Saanich Inlet	HL5aSF04.14
KC488594	HL5aSF04.36	Uncultured eukaryoteClone DSGM-52 (AB275052)	94	Sagami Bay	HL5aSF04.36
KC488595	HL5aSF04.67	Uncultured eukaryote clone OR000415.113 (AY381186)	93	North Atlantic	HL5aSF04.67
KC488596	HL5aSF10.24	Uncultured eukaryote clone AMT15_33_5m_126 (GQ863824)	96	Atlantic Meridional Transect	HL5aSF10.24
KC488597	HL5aSF10.31	Uncultured eukaryote clone CC02A740.069 (JX188377)	96	South China Sea	HL5aSF10.31
AUTOTROPHS STRAMENOPILES					
KC488598	BBL3SF04.92	*Triparma* sp. TOY-0807 (AB546639)	99	Japan: Oyashio	BBL3SF04.92, BBL3SF04.57, HL4SF04.18, BBL3SF04.17
KC488599	BBL3SF04.96	*Minidiscus trioculatus* strain CCMP 496 (FJ590768)	100		HL5aSCM04.29, HL4SF04.30, HL4SF04.16, HL4SF04.38, BBL3SF04.96, BBL3SF04.03
KC488600	BBL6SF04.33	Uncultured Pedinellales Clone B7 (FN263029)	98	Baltic Sea	BBL6SF04.33
KC488601	HL2SCM04.28	*Aureococcus anophagefferens* clone Q1-4 (JQ420078)	99	China coast	BBL6SF10.12, HL2SCM04.28, BBL6SF10.86, BBL6SCM10.08
KC488602	HL2SCM04.60	*Pseudo-nitzschia australis* isolate 10249 10AB (JN599166)	99	Monterey Bay	BBL3SCM04.10, BBL3SCM04.85, BBL3SCM04.78, BBL3SCM04.69, BBL3SCM04.51, HL2SCM04.60, BBL1SCM04.30

KC488603	HL2SF04.24	*Aureococcus anophagefferens* clone Q1-4 (JQ420078)	99	China coast	HL5aSF10.11, HL2SF10.34, HL2SF04.24
KC488604	HL2SF10.31	*Dictyocha speculum* (U14385)	99		HL2SF10.31, HL2SCM10.77
KC488605	HL2SF10.58	Uncultured stramenopile clone RS.12f.10m.71 (KC583004)	99	Red Sea Coast	HL2SF10.58
KC488606	HL4SCM04.58	*Thalassiosira aestivalis* strain CCMP 975 (DQ093369)	99		HL4SCM04.85, HL4SF04.40, HL4SF04.42, HL4SCM04.75, HL4SCM04.41, HL4SCM04.57, HL4SCM04.58, HL4SCM04.22, HL4SCM04.18, HL4SCM04.10
KC488607	HL4SF04.11	*Bolidomonas pacifica* strain CCMP1866 (HQ912557)	99		HL4SF04.11
KC488608	HL4SF04.90	*Chaetoceros* sp. CCAP 1010/16 (FR865486)	94	North Sea	BBL6SCM04.42, HL4SF04.48, HL4SF04.47, HL2SCM04.25, HL4SF04.90, HL4SF04.13, BBL3SCM04.05
KC488609	HL5aSCM04.22	*Fragilariopsis curta* (EF140623)	99	Antarctica	BBL3SF04.35, BBL3SF04.80, HL5aSCM04.22, BBL3SF04.72, BBL3SCM04.22
KC488610	HL5aSF04.39	Uncultured stramenopile clone RS.12f.10m.71 (KC583004)	99	Red Sea Coast	HL5aSF04.39, HL5aSCM04.95
TELONEMIA					
KC488611	BBL1SCM04.64	Uncultured Telonema clone MALINA_St320_ES065 (JF698774)	99	Beaufort Sea Arctic	BBL1SCM04.64
KC488612	BBL1SCM10.28	Uncultured eukaryote clone RA000412.136 (AJ564768)	94	France	BBL1SCM10.28
KC488613	BBL1SCM10.41	Uncultured eukaryote clone RA000412.136 (AJ564768)	97	France	BBL1SCM10.41, BBL1SCM10.02
KC488614	BBL1SF04.91	Uncultured Telonema clone MALINA_St320_ES065 (JF698774)	99	Beaufort Sea Arctic	BBL1SF04.91, BBL1SF04.61
KC488615	BBL6SCM04.14	Uncultured marine eukaryote clone CNCIII05_8 (HM581741)	98	Arctic Ocean	BBL6SCM04.14
KC488616	BBL6SCM04.57	Uncultured marine eukaryote clone CNCIII05_8 (HM581741)	97	Arctic Ocean	BBL6SCM04.57
KC488617	HL2SCM10.10	Uncultured eukaryote 18S rRNA gene, clone BL010625.25 (AJ564770)	99	Spain	HL2SCM10.10
KC488618	HL2SCM10.17	Uncultured eukaryote clone NPK97_54 (EU371378)	97	Svalbard	HL2SCM10.17
KC488619	HL2SF04.80	Uncultured Telonema clone STFeb_122(HM135058)	94	Lake Germany	HL2SF04.80
KC488620	HL2SF10.61	Uncultured Telonema clone STFeb_122(HM135058)	95	Lake Germany	HL2SF10.61

Chapitre 3 - Tracking environmental selection of marine microbial eukaryotes using 18S rRNA genes and 18S rRNA in the Gulf of Maine

Résumé

Comprendre les relations entre la sélection de l'environnement et la structure de la communauté microbienne implique l'identification des taxons qui sont actifs dans des conditions spécifiques et les facteurs environnementaux impliqués. Dans cette étude, nous étions deux régions distinctes, Jordan Basin (JB) et North East (NE) Channel et Fan dans le golfe du Maine (Canada) en Juillet 2010. Les petits eucaryotes de la communauté microbienne marine (MME) échantillonnés à la surface et au maximum de chlorophylle subsurface (SCM) sont identifiés par séquençage à haut débit ciblant la région V4 du gène 18S ARNr (ADN) et par transcription inverse de l'ARNr (ARN). Nos résultats soulignent de grandes différences entre les communautés MME des deux régions. Toutes les communautés issues de JB inférées à partir de l'ADN étaient différentes des communautés inférées à partir de l'ARN. Il y avait également des signaux ARN et ADN distincts aux SCM des stations NE, alors qu'en surface de NE les communautés étaient plus semblables entre ADN et ARN. Globalement, la fréquence relative des reads d'ADN et d'ARN par taxons étaient corrélées ($R^2 = 0,613$, $p < 0,001$), mais avec des différences significatives au niveau des ratios ADN/ARN reads à plusieurs niveaux taxonomiques. Les Chlorophytes, diatomées, Picozoa et les marines alvéolées (groupe -I,- II) étaient sur-représentés avec l'ADN par rapport à l'ARN tandis que d'autres, en particulier les ciliées, haptophytes, Telonemia, et Chrysophyceae étaient sur-représentés avec l'ARN. Cette étude a soulignée la façon dont les séquences environnementales obtenues à partir de l'ADN vs ARN, ne sont pas uniformes dans les environnements ou groupes taxonomiques, mais que combinées elles permettent potentiellement de fournir des informations supplémentaires sur l'état des communautés.

Abstract

Understanding the relationships between environmental selection and microbial community structure implies both knowledge of relevant environmental drivers and identification of taxa that are active under specific conditions. In this study, we investigated two distinct regions, Jordan Basin (JB) and the North East Channel and Fan (NE) in the Gulf of Maine (Canada) in July 2010. Small marine microbial eukaryotic (MME) assemblages from the surface and subsurface chlorophyll maxima (SCM) were identified using high-throughput amplicon sequencing targeting the V4 region of the 18S rRNA gene (DNA template) and reverse transcribed rRNA (RNA template). Our results highlighted major differences between MME communities from the two regions. All JB communities inferred from DNA clustered apart from the RNA sourced communities. There were also distinct RNA and DNA signals in the SCM across the NE stations, while in surface samples the MME communities identified from the two templates were more similar. Overall, relative occurrence of DNA and RNA reads by taxa were correlated (R^2= 0.613, p<0.001); however, there were significant differences in the ratio of DNA/RNA reads at multiple taxonomic levels. Chlorophytes, diatoms, Picozoa, and marine alveolates (groups I and II) were over represented in the DNA compared to RNA, while others, especially ciliates, haptophytes, Telonemia, and Chrysophyceae were over represented in RNA. This study highlighted how environmental sequence data from DNA and RNA are not uniform across environments or taxonomic groups. The approach targeting both has the potential to provide additional information on the state of the communities.

3.1 Introduction

Because of the universality of the gene coding for small subunit ribosomal RNA (SSU rRNA), high throughput environmental SSU rRNA gene surveys are an efficient way to explore microbial communities over space and time (Sogin et al., 2006; Not et al., 2007; Comeau et al., 2011). Such surveys targeting the eukaryotic 18S rRNA gene have highlighted community level differences in marine microbial eukaryote (MME) communities in different oceans and depths (Brown et al., 2009; Stoeck et al., 2010; Comeau et al., 2011; Monier et al., 2013). However, the majority of such surveys have been carried out using DNA as a template, and identifying whether these reads are in fact from active or viable cells has been questioned (Stoeck et al., 2007). DNA from both living and dead microorganisms, or even as extracellular DNA, can be amplified and sequenced (Josephson et al., 1993; Sørensen et al., 2013), and inferring ecologies of organisms from DNA records is problematic. For example, Charvet et al. (2012) found sequences matching marine diatoms in the anoxic waters of a coastal Arctic meromictic lake where no living diatoms had ever been reported and suggested that the diatoms had most likely been transported to the lake from the nearby ocean by wind or precipitation. Similarly Harding et al. (2011) amplified and cloned DNA from marine diatoms in glacial snow 200 km from the Arctic Coast.

Another approach to address such concerns is to amplify RNA converted to cDNA and target 18S rRNA, which is found in the ribosomes of living cells. Since ribosomes are more numerous in cells that are actively producing proteins needed for growth and division, the taxa identified from 18S rRNA are often inferred to be from the active community, with some caveats, such as resting stages being able to retain their ribosomes (Blazewicz et al., 2013). In principal, as RNA degrades more quickly than DNA, the relative similarity of the two communities identified from RNA compared to DNA could provide information on the turnover of communities in the environment. To date, MME studies using both templates have reported sometimes surprising, but inconsistent differences in communities detected from RNA compared to genomic DNA (Stoeck et al., 2007; Not et al., 2009; Terrado et al., 2011), suggesting there is no consistent bias between the two templates and differences (or not) between them may well be ecologically informative. For example, heterotrophic marine stramenopiles (MAST) were recently reported to show little difference in the

proportions of sequences from DNA compared to cDNA (Logares et al., 2012). This would be consistent with balanced growth and cellular losses, for example from grazing, in which case the signal from DNA would be closely coupled to the number of ribosomes and rRNA. In other cases, over representation of taxa in DNA compared to RNA could be the result of dead or dormant cells persisting in the environment following nutrient depletion and cessation of growth. Alternatively, over representation of taxa from RNA compared to DNA could indicate the activation of taxa at the start of a bloom (Terrado et al., 2011). At the community level, large differences between DNA and RNA-derived communities could be ecosystem specific, resulting from low loss rates and high DNA preservation such as found in anoxic marine waters (Not et al., 2009; Charvet et al., 2012). Alternatively the differences could be due to the successional state of the systems. In this case, communities in transition would have greater disequilibrium between the DNA and RNA, while in systems where loss and growth processes were closer to equilibrium, the two communities would be predicted to be more similar.

The Gulf of Maine (GoMA), bordered by the New England coastline of the United States and the eastern maritime provinces of Canada (Northwest Atlantic Ocean), is ecologically diverse and an important habitat for many marine species (Mills, 1989). The complex oceanographic features in the GoMA and its connection with open Atlantic waters over short distances provided an opportunity to investigate MME communities with different histories and potential successional stages. While the morphological and taxonomic diversity of the photosynthetic microplankton (20 - 200μm) and larger nanoplankton (10 - 20μm) is well-described in the GoMA (Li et al., 2011), less is known about smaller eukaryotic picoplankton (0.8 - 3μm) and flagellates in the smaller nanoplankton (2 - 10μm) size class. The main objective of this study was to explore differences between small communities inferred from DNA compared to those from RNA. Our working hypothesis was that certain taxa would always be over represented in DNA compared to RNA. The alternative would be that over representation in DNA compared to RNA is not taxon specific. In the latter case, the environment would be the primary factor determining over or under representation. We would then predict that taxa that were over represented in RNA were those that were responding to recent environmental changes that

promote high species turnover. Our goal was to link biological species turnover to physical processes.

3.2 Materials and methods

3.2.1 Study region and general sampling

Samples were collected in the GoMA aboard the *CCGS Hudson* from 28 July to 11 August 2010 (**Annex 3_figure A1)** in conjunction with the Discovery Corridor Program of the Canadian Department of Fisheries and Oceans, the Centre for Marine Biodiversity, and the Gulf of Maine Census of Marine Life. Six stations were selected. Three stations were from the Jordan Basin (JB), which is one of three large basins in the GoMA. Three other stations were selected starting at the outer edge of North East Channel (NEC), which is the major connection between the Gulf and the rest of the Northwest Atlantic, and proceeding to the North East Fan (NEF) where the continental shelf slopes into the deep Atlantic; these three stations are referred to as North East (NE) stations.

Discrete water samples were collected directly from the 12-L Niskin bottles mounted on a Rosette equipped with profilers for conductivity, temperature and depth (CTD; SBE-911 CTD; Sea-Bird Inc., Bellevue, WA, USA), oxygen (Seabird Model SBE-43 dissolved oxygen sensor), *in situ* chlorophyll fluorescence (Chelsea Aquatracka MK III), relative nitrate concentrations using the *in situ* ultraviolet spectrophotometer (ISUS) probe (Satlantic, Halifax NS Canada), and a LI-COR spherical Quantum Sensor (model LI-193SA) for data on photosynthetically available radiation (PAR: 400-700 nm). Profiles were acquired on the downcast. The depth of the subsurface chlorophyll maximum (SCM) was identified from the chlorophyll fluorescence profile, and water samples were collected on the upcast. The depth of euphotic zone (Z_{eu}) was calculated as 1% surface irradiance (PAR), and the mixed layer depth (Z_{mix}) was calculated from CTD profile as the shallowest depth at which we detected the maximum rate of change in density (σ_θ).

Samples for nutrients, were collected directly from the Niskin bottles, samples for chlorophyll *a* (Chl *a*) were collected into 2-L polycarbonate (PC) bottles, and water samples for nucleic acids were collected into 7-L cleaned and sample rinsed PC carboys.

Samples for nutrients were immediately frozen at -20°C. Material for Chl *a* was filtered onto GF/F filters (Whatman, GE Healthcare, Piscataway, NJ, USA) and were stored at -80°C until analysis. For nucleic acids, 6 L of water was sequentially filtered through a 50-µm nylon mesh, a 47-mm diameter 3-µm pore size PC filter, and finally a 0.2-µm pore size Sterivex unit (Millipore Canada Ltd, Mississauga, ON, Canada). RLT buffer (Qiagen) with β-mercaptoethanol was added to the Sterivex units that were then kept at -80°C until nucleic acid extraction. Only material in the 3 to 0.2 µm fraction was used for this study. This was meant to enrich for cells <3 µm in one dimension (Vaulot et al., 2008; Seenivasan et al., 2013).

3.2.2 Laboratory protocols

Nutrients (nitrite and nitrate referred to together as nitrate, phosphate and silicate) were analyzed spectrophotometrically as in Mitchell et al. (2002). For ammonium, the fluorometric method of Kerouel and Aminot (1997) was used. Filters for Chl *a* were extracted in 90% acetone (Mitchell et al., 2002), and concentration was determined by fluorescence on a Turner Design Model 10 fluorometer, before and after acidification.

DNA and RNA were extracted from Sterivex units using AllPrep DNA/RNA Micro Kit (Qiagen). Extraction followed the manufacturer's instructions with a cell lysis step, 600 µl of RLT Plus, 85 µl SDS 10%, and 30 µl proteinase K were added to the Sterivex units and incubatedat 65°C for 15 min. Extracted RNA was immediately reverse transcribed to cDNA using a High Capacity Reverse Transcriptase Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's recommendations. Extract quality was checked on a 1% agarose gel. Extracted DNA and cDNA were amplified with primers, E572F (5'-CY GCG GTA ATT CCA GCT C-'3) and E1009R (5'-CRA AGA YGA TYA GAT ACC RT-'3) targeting the V4 region of the 18S rRNA gene as described in Comeau et al. (2011). Primers for individual samples included a sample specific Roche multiplex identifier (MID).Equal quantities of amplicons from each sample were pooled, then run on a 1/4 plate on the Roche 454 pyrosequencer using GS-FLX Titanium series chemistry at the IBIS/Université Laval Plateforme d'Analyses Génomiques. The raw reads are available in the NCBI Sequence Read Archive (SRA) with accession number SRP040423. For the

remainder of the paper we will refer to the V4 18S rRNA gene as the DNA reads and the 18S rRNA converted to cDNA as the RNA reads.

3.2.3 Quality control and taxonomic assignments

Raw reads were quality controlled and chimeras were detected using UCHIME (Edgar et al., 2011). In addition, reads were quality screened as follows: reads with one or more ambiguous bases (Ns), <150 bp, longer than expected amplicon size or lacking a perfect match with the forward PCR primer were removed. The remaining reads were aligned using mothur version 1.21.1 (Schloss et al., 2009) against the SILVA eukaryotic alignment (Pruesse et al., 2007). Misaligned reads were removed and aligned reads were clustered at the ≥98% similarity level into Operational Taxonomic Units (OTUs) using furthest-neighbor clustering with mothur as described in Comeau et al. (2011). Singletons, OTUs with only one sequence in the total dataset, were discarded, and the dataset randomly re-sampled to ensure the same number of reads from each sample. OTUs were taxonomically identified with a 50% bootstrap cut-off (Comeau et al., 2011) in mothur against our in-house reference database that includes Atlantic coastal and open ocean reference sequences (Dasilva et al., 2013). Sequences classified as belonging to Metazoa (4123 reads) and Streptophyta (4 reads) represented together 3.5% of total dataset and were not treated in this study.

3.2.4 Statistical analyses

A nonparametric analysis of variance (Kruskal-Wallis; $p < 0.05$) (Kruskal and Wallis, 1952) run using the Paleontological Statistics Software Package (PAST) version 2.12 (Hammer et al., 2001) was used to test differences in environmental variables (temperature, salinity and nutrient concentrations) among stations and depths. Rarefaction curves, Chao1 estimator, Shannon diversity indices (Clarke, 1993), and shared OTUs were derived using mothur. Differences between communities was estimated using the Fast UniFrac Web Server (Hamady et al., 2010), to compute all UniFrac distances. The UniFrac distances themselves were calculated from a large-scale phylogenetic tree based on the mothur/Silva alignment of all retained reads (118 525 reads) using FastTree version 2.1

(Price et al., 2010) in "accurate mode" (-mlacc 2 -slownni) with the general time reversible (GTR) model and pseudo counts. Principal Coordinate Analysis (PCoA) plots (weighted) and UPGMA clusters analysis (unweighted) were also generated. We identified the five main taxonomic groups responsible for the clustering by UniFrac using a SIMPER test, run within PAST based on the relative abundance of reads.

The relationship between environmental variables and the potentially active eukaryotic community recovered by RNA was explored by a Canonical Correspondence Analysis (CCA) triplot distribution using the Canoco for Windows package version 4.5 (Ter Braak and Šmilauer, 2002). The variables used were: temperature, salinity, nutrient concentrations, total Chl *a* concentrations, and pycnocline depth. The taxa for this analysis were selected by frequency of occurrence >1%, and significance levels of environmental variables were tested using the forward selection criteria.

Finally, scatter plots of paired comparisons between number of DNA and RNA reads were generated for individual taxonomic groups. The significance of DNA and RNA read ratios for each taxon occurrence was inferred from the coefficient of determination (R^2), and standard error run in SigmaPlot 11.0 (Systat Software, San Jose, CA). DNA to RNA ratios by taxon were compared among sampling sites and depths.

3.3 Results

3.3.1 Environmental parameters

In summer, Jordan Basin (JB) stations in the Gulf of Maine (GoMA) were mostly temperature stratified but with slightly greater salinity at depth. Along the North East (NE) Channel and Fan stations the thermocline and halocline were at similar depths, creating a well-stratified water column. The temperature difference between the surface and thermocline was on average < 3°C for JB, compared to > 6 °C for NE stations (**Figure 3.1, Annex 3_figure A2**). Overall, JB upper waters were cooler with lower salinity compared to the NE stations (**Table 3.1, Figure 3.1**). Station NEC211, where shelf water is exchanged with offshore waters, marked a transition in physical characteristics from inside to outside the GoMA (**Figure 3.1, Annex 3_figure A2**). SCM were detected at or just above the pycnocline at around 15 m depth in JB and between 30 to 50 m depth in the NE stations

Table 3.1 : Water column characteristics of the six stations from Jordan Basin (JB), North East Channel (NEC) and North East Fan (NEF) with coordinates of each station (Latitude: Lat., and Longitude: Long.); date; and local time of sampling. Bottom depth is for bathymetry comparison between two regions; depth of subsurface chlorophyll maximum (SCM depth); depth of euphotic zone (Euph depth); depth of pycnocline (Pycno depth, see materials and methods); and euphotic zone/mixture zone ratio (Z_{eu}/Z_{mix}, see materials and methods).

Stations	Date	Local time	Lat. (°N)	Long. (°W)	Bottom depth (m)	SCM depth (m)	Euph depth (m)	Pycno depth (m)	Zeu/Zmix ratio
JB121	31 Jul	07:00	43°19	67°03	81	15	34	12	2.83
JB700	31 Jul	20:00	43°17	67°22	162	15	36	12	3.00
JB601	30 Jul	20:00	43°11	67°14	209	16	32	6	5.30
NEC211	09 Aug	05:30	41°54	65°30	1500	51	-	46	-
NEF43	05 Aug	20:30	41°47	65°20	1933	30	41	24	1.71
NEF512	07 Aug	18:20	41°16	65°15	2951	39	39	28	1.39

(-) Euph. depth at station NEC211 no determined: PAR surface was 0.4 µmol/m^2 sec.

(**Table 3.1, Figure 3.1**). JB surface and SCM samples were from the same relatively warm (10°C to 17°C) low-salinity (32.1 to 32.3) water mass (**Annex 3_figure A2**). Nutrient concentrations ranged from 0.7-2.3 µM for nitrate, 0.4-0.5 µM for ammonium, 0.3-0.7 µM for phosphate, and 1.6-3.6 µM for silicate. Chl *a* concentrations were 0.32-1.33 µg Chl *a* L^{-1} (**Table 3.2, Figure 3.1**).

In contrast, surface and SCM samples from the NE stations were from different water masses (**Annex 3_figure A2**). The near surface waters were low in nutrients (0.7-0.8 µM nitrate, 0.4-0.6 µM ammonium, 0.1-0.3 µM phosphate, 0.3-0.8 µM silicate) and Chl *a* (0.11- 0.18 µg Chl *a* L^{-1}). The surface layer was also warmer (19°C to 22°C) with higher salinity (32.5 to 33.5) compared to the JB waters (**Table 3.2, Figure 3.1**). NE station temperatures below the pycnocline were colder (11°C to 16°C), salinity greater (32.8 to 34.8), nutrient concentrations higher, and Chl *a* concentrations greater compared to the surface (**Table 3.2, Figure 3.1**). Overall >there was no statistical difference between the physical metrics of the JB surface and SCM waters (Kruskal-Wallis; $p > 0.05$). From NE stations, temperatures were significantly different between surface and SCM ($p = 0.013$). Surface temperatures were significantly different between the two regions ($p = 0.011$) along with salinity ($p = 0.039$), while at the SCM only salinity was significantly different between the two regions ($p = 0.044$).

3.3.2 Pyrosequencing results, quality control and diversity index

The bioinformatics filtering process removed about 24% of raw DNA reads, and 33% of raw RNA reads, leaving 90 198 DNA and 70 118 RNA high-quality reads (**Annex 3_figure A4**). After checking the alignment and performing additional BLAST searches, 63 850 DNA and 54 675 RNA reads with ≥95% query coverage remained. After clustering at ≥98% similarity, based on the total of 118 525 reads (DNA and RNA), 15 717 OTUs were generated. On average, there were 357 OTUs per 1000 DNA reads and 487 OTUs per 1000 RNA reads (**Annex 3_figure A4**). Rarefaction curves (**Annex 3_figure A5**) also indicated more OTUs from RNA than from the DNA. None of the rarefaction curves reached saturation. The Shannon Index and Chao1 estimator (**Annex 3_figure A6**) suggested that community diversity was higher in the JB compared to NE stations, and was higher in

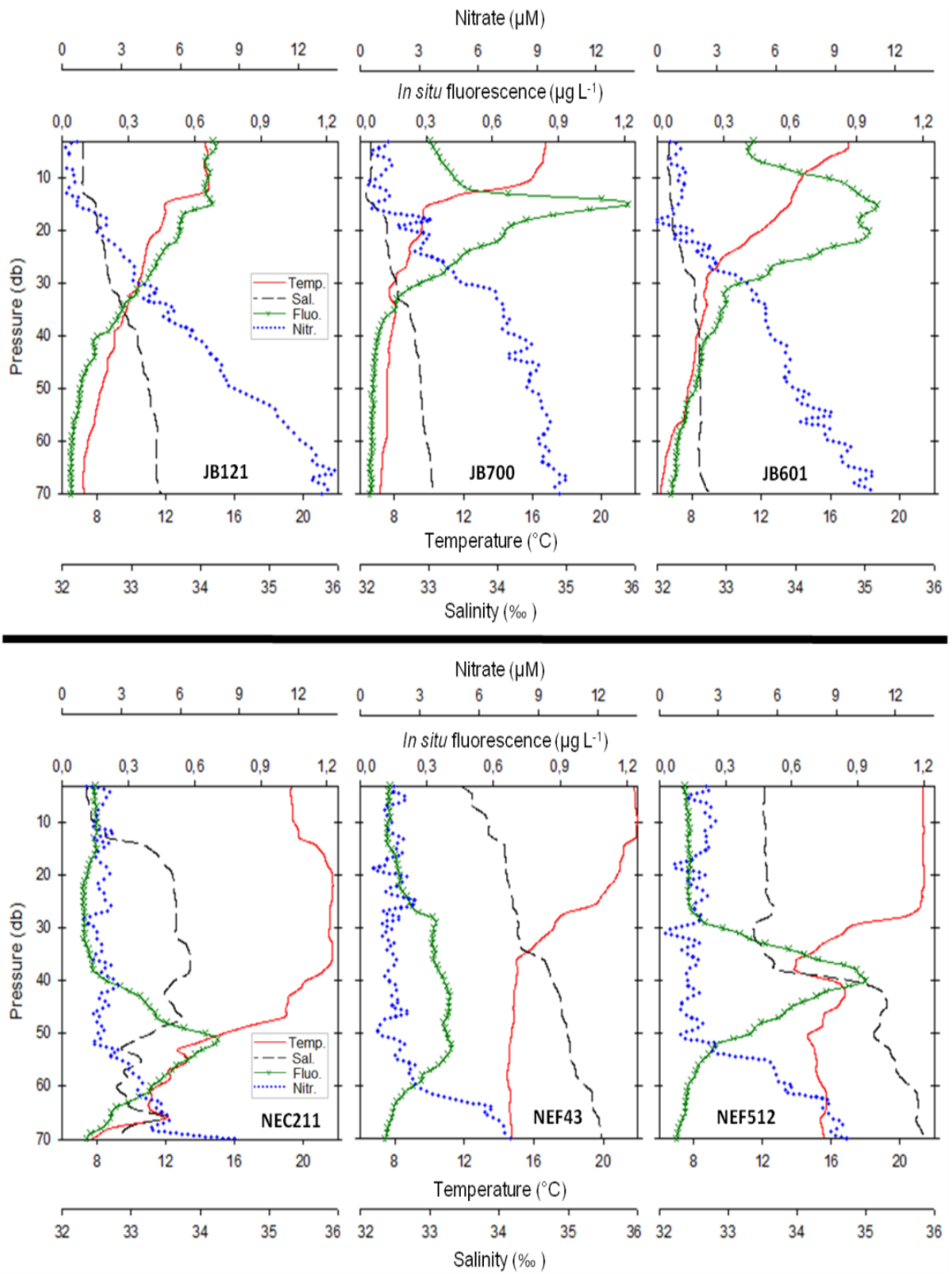


Figure 3.1: CTD profiles of the upper 70m of the six stations: temperature (°C, red solid line), Salinity (black dashed line), *in situ* fluorescence (µg L^{-1}, green heavy hatched line), and *in situ* nitrate concentrations (µM, blue dotted line) calibrated from bottle sample data (**Annex 3_figure A3**).

samples from RNA compared to DNA. Diversity measures also differed between the DNA and RNA templates, but were only significant between the JB samples (two-sample t-test, $p \leq 0.05$).

3.3.3 Biogeography and biodiversity of MME

Overall, alveolates dominated the reads, especially those from the NE stations. Most of the alveolate reads were dinoflagellates representing 31% of reads, followed by ciliates at 13%, and diverse uncultivated marine alveolates (MALV) at 11% (**Figure 3.2a**). Photosynthetic and mixotrophic groups in both samples were haptophytes, chlorophytes, and cryptophytes, as well as photosynthetic stramenopiles. Stramenopiles also include heterotrophic flagellates, especially the polyphyletic uncultivated marine stramenopiles (MAST) (**Annex 3_figure A7**). Other heterotrophic flagellates that were well represented included Picozoa, katablepharids, and Choanoflagellida (**Figure 3.2a**). The category "other", accounting for 0.25% (297) of reads, was made up of heterotrophic groups including fungi, Amoebozoa, Centroheliozoa and Apusozoa (**Figure 3.2a**).

This grouping into major eukaryotic lineages, highlighted the similarity of the JB surface and SCM samples and suggested higher representation of several groups in the RNA compared to the DNA. For example, in all cases Telonemia and Haptophyta were a higher proportion of the total in the RNA compared to the DNA (**Figure 3.2a**). The NE surface samples were characterized by a higher proportion of ciliate reads in the RNA, and the relative proportion of autotrophic groups was low compared to JB surface samples (**Figure 3.2a**). In the NE SCM, haptophyte reads seemed to be more abundant in RNA compared to DNA similar to JB, while MALV tended to be lower in the RNA compared to DNA (**Figure 3.2a**). At finer taxonomic levels similar differences among the stations and groups were also evident (**Annex 3_figure A7**).

From the total of 15 717 OTUs clustered at ≥98% sequence identity, ca. 49% of OTUs recovered in RNA from JB were shared with the DNA, and ca. 45% of NE RNA OTUs and DNA OTUs were shared (**Figure 3.2b**). We found more OTUs restricted to the RNA template (28%) compared to the DNA (15 %) from JB. Similarly for the NE stations, 30% of the OTUs were only found from the RNA (21 % from DNA). Combining both regions and templates, there were 958 OTUs (6%) in common (**Figure 3.2b**).

Table 3.2 : Chlorophyll *a* concentrations and physicochemical parameters of the samples selected for tag pyrosequencing from Jordan Basin (JB) and North East Channel (NEC) or Fan (NEF). Extracted total chlorophyll *a* (Tot. Chl *a*); temperature (Temp); salinity (Sal); and nutrient concentrations: ammonium (NH_4^+), nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^{4-}).

Stations	Depth (m)	Tot. Chl *a* (µg Chl *a* L^{-1})	Temp (°C)	Sal.	NH_4^+ (µM)	NO_3^- (µM)	PO_4^{3-} (µM)	SiO_4^{4-} (µM)
JB121	2	0.53	15.7	32.2	0.37	1.04	0.49	2.21
	14	0.70	13.6	32.3	0.50	2.28	0.69	2.84
JB700	2	0.32	17.3	32.1	0.49	0.81	0.27	1.59
	14	1.33	10.7	32.1	0.40	0.87	0.31	1.69
JB601	2	0.40	17.1	32.2	0.37	0.99	0.33	2.05
	12	0.55	15.4	32.2	0.40	2.49	0.66	3.56
NEC211	2	0.18	19.3	32.5	0.62	0.90	0.26	0.77
	51	0.35	11.5	32.8	0.60	2.66	0.64	2.89
NEF43	2	0.13	21.8	33.5	0.41	0.87	0.09	0.41
	30	0.39	14.9	34.8	0.48	0.90	0.23	1.43
NEF512	2	0.11	21.2	33.4	0.44	0.87	0.11	0.27
	39	0.87	16.0	34.7	0.70	2.33	0.48	2.38

PCoA of UniFrac weighted distances grouped the communities first by region and second by template (**Figure 3.3a**). The P1 axis explained 35% of the variance in the data, reflecting differences in communities between the JB and NE stations. Around 20% of the variance in the data was explained by the P2 axis, reflecting differences in communities recovered from the DNA and RNA templates. Samples from JB were relatively close to each other, indicating more similar communities in terms of both abundance and phylogeny compared to a greater spread among the NE stations (**Figure 3.3a**). We then used to presence/absence of OTUs to reveal any biogeography and template specific patterns. Presence/absence data minimizes the effect of multiple copies of the 18S rRNA gene in the genomes of different taxa, and provided additional weight to rare taxa from the RNA template. The resulting unweighted UniFrac cluster analysis also revealed the pronounced separation of MME communities from the two regions (**Figure 3.3b**). The JB communities clustered by starting template (DNA and RNA), with no separation of the surface and SCM samples. In contrast, within the NE stations, the SCM communities were separated from communities from the surface, and the influence of starting templates was more complex.
The NE SCM DNA and RNA communities clustered apart from each other, whereas surface communities from the two templates were more similar and grouped by station. Moreover, the surface community from station NEC211 formed its own distinct cluster, with DNA and RNA communities relatively similar and separate from the other NE samples (**Figure 3.3b**).

We then investigated taxonomic composition based on relative abundance of OTUs of the taxa in the unweighted UniFrac cluster analysis and ran a SIMPER test to identify taxa that explained the clustering (**Figure 3.3b**). Although an unweighted test was used initially to provide additional weight to rare versus dominant taxa, we noted that at a given taxonomic rank, those with more OTUs were responsible for the clustering by sampling site and template. The ciliate contribution was much greater in the NE regions; representing 24% of the NE reads compared to 3.2% of JB reads. Telonemia reads were most abundant in JB, and Haptophyta were recovered everywhere, but were less abundant in the surface of the NE stations (**Figure 3.3b**).

CCA indicated a significant relationship between communities recovered from RNA, and environmental factors explained 55.2% of the variability in the first axis and

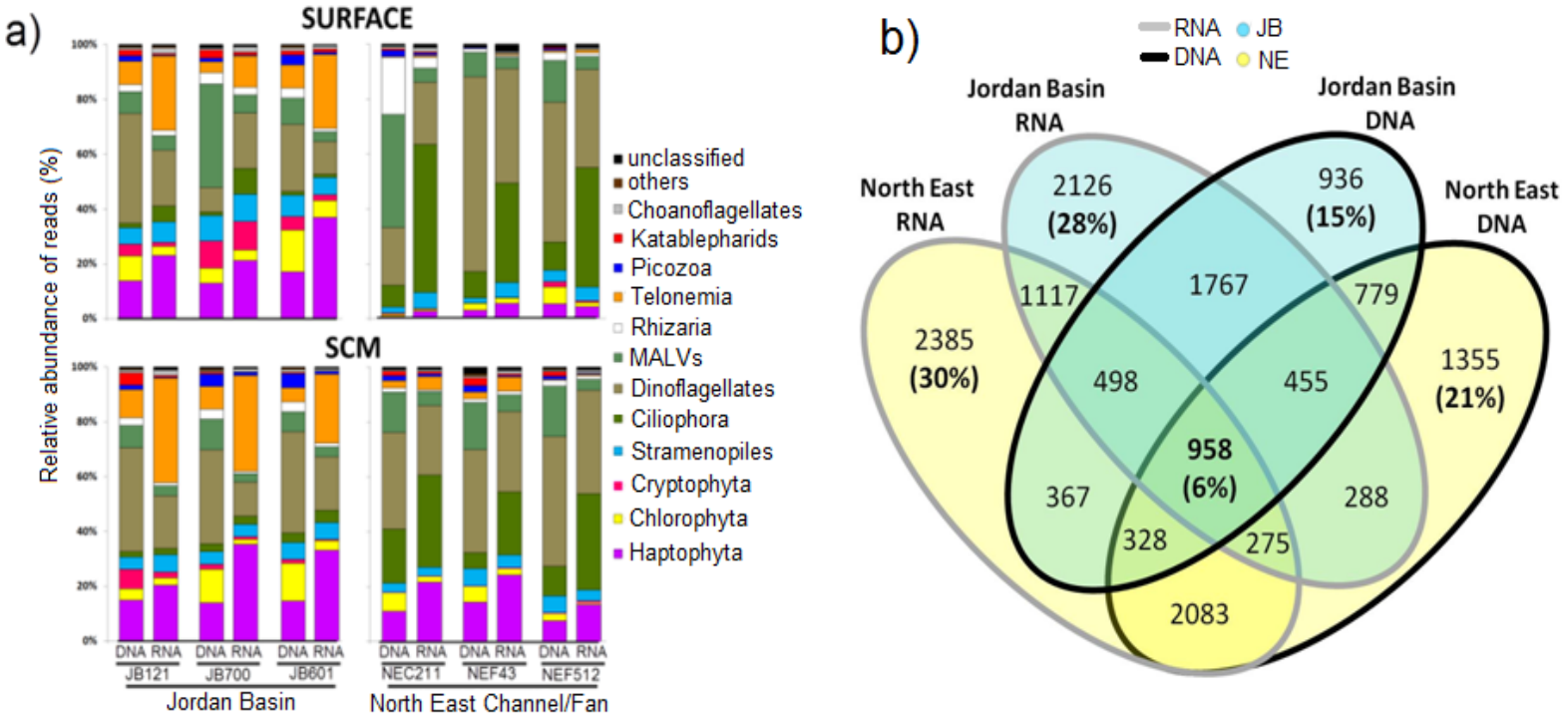


Figure 3.2: **a)** Relative abundance of reads according major groups. **b)** Number of OTUs shared between regions and templates (total OTUs 15 717).

15.9% of the variability in the second axis (**Figure 3.4**). Using forward selection, we noted that water temperature (p = 0.024) and salinity (p = 0.010) were the two most significant variables responsible of this clustering. Communities from the NE stations were most influenced by these physical factors, corresponding to higher temperatures in the surface and higher salinity in the SCM. Moreover, SCM from both regions were associated with higher nutrient concentrations. We note that the JB700 surface station clustered apart (**Figure 3.3b**), but showed little association with any environmental variable tested.

3.3.4 Global differences between the DNA and RNA templates

With the aim of identifying any trends of DNA:RNA ratios by taxa across the sampling region, we plotted read numbers from RNA versus DNA by station and by taxonomic group (**Figure 3.5a**). If a point fell above the 1:1 line there were more reads from a given taxonomic group in a given sample in the RNA template compared to the DNA template. There was an overall significant correlation between the RNA and DNA reads (R^2 = 0.61, $p <$ 0.001). There was no evident pattern when the points were identified by station and sample depth (**Figure 3.5a left**), but when taxonomic groups were identified in the same plot (**Figure 3.5a right**), the ratio of DNA:RNA reads appeared to vary by taxon. We tested the significance of these trends using linear correlation (**Table 3.3**). Among the major groups responsible for UniFrac clustering in Figure 3.3b, reads from both DNA and RNA were significantly correlated for Telonemia, Haptophyta, Ciliophora, and dinoflagellates (R^2 > 0.66, $p <$ 0.02), but not for MALV (**Table 3.3**). Telonemia, Haptophyta, and Ciliophora tended to be over represented in the RNA compared to the DNA (above the 1:1 line). In contrast, dinoflagellates were mainly recovered from DNA (below the 1:1 line) (**Figure 3.5a right**). In addition Chlorophyta, Cryptophyta, Rhizaria, Picozoa, and Katablepharida were mainly recovered from DNA (below the 1:1 line), while choanoflagellates was mostly recovered from RNA (above 1:1 line) (**Figure 3.5a right**).

We then tested for significance between the DNA and RNA reads at the level of taxa identified from the CCA (**Figure 3.4**), which were those that most contributed to the communities distribution. At these mostly finer taxonomic levels, among taxa with very significant correlations (R^2 > 0.85, p < 0.001) were Pelagophyceae, *Pyramimonas*

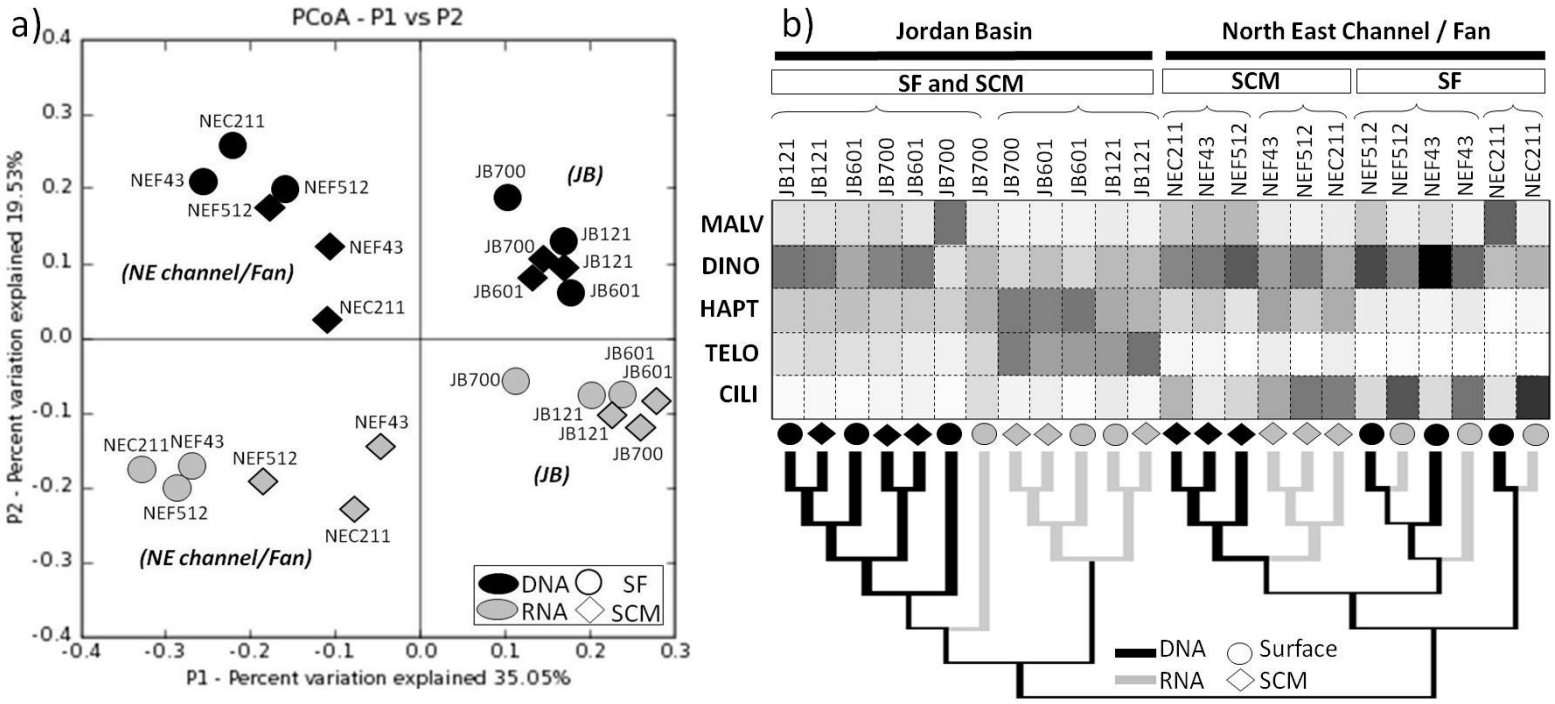


Figure 3.3: Phylogenetic structure of eukaryotes communities based on OTUs clustered at ≥98% sequences identity. **a)** Corresponding Principal Coordinate Analysis (PCoA) using UniFrac weighted distance metric. OTU relative abundances were normalized across samples. **b)** Unweighted Unifrac cluster analysis associated with matrix of relative abundance of taxa which are mainly responsible of difference between stations and/or templates identified by SIMPER test (MALV, marine stramenopiles; DINO, dinoflagellates; HAPT, Haptophyta; TELO, Telonemia; CILI, Ciliates).

(Chlorophyta), *Polycystinea* (Rhizaria), and two haptophyte orders (Zygodiscales and Prymnesiales) (**Table 3.3**). The grouping by similarity of the DNA to RNA ratios emphasized that most of these taxa were predominantly recovered from the RNA template, except for Zygodiscales and *Polycystinea* (**Figure 3.5b**). At the significance level of $p \leq 0.05$ with R^2 between 0.58 and 0.83, another 15 clades were identified as having significant correlations between the DNA and RNA reads (**Table 3.3**). Within these were several haptophyte orders (Phaeocystales, Syracosphaerales, and Coccosphaerales), Chrysophyceae, MAST-4 and MAST-7 that were over-represented by RNA (**Figure 3.5b**), while Mamiellophyceae (*Bathycoccus*, *Micromonas*, *Ostreococcus*), diatoms, MALV-I and MALV-II were over-represented in DNA (**Figure 3.5b**). For other taxa, the DNA and RNA reads were not significantly correlated, and were more sporadic. This category included five other MAST groups (-1, -2, -8, -9, -10), choanoflagellates, bolidophytes, Dictyochophytes, several Rhizaria (Cercozoa, *Polycystinea* and *Sticholonche*), Zygodiscales in the Haptophyta, and *Pterosperma* a chlorophytes, (**Table 3.3, Figure 3.5b, Annex 3_figure A7**).

3.4 Discussion

3.4.1 Communities from rRNA gene and rRNA

Diversity assessed from RNA and DNA led to a more complex view of the MME community compared to using a single template. The high throughput approach yielded high recovery of OTUs from the two templates and the better coverage increased access to rare taxa (Sogin et al., 2006), making it possible to track many of the same taxa in both DNA and RNA. We report ca. 50% overlap in the OTUs retrieved as RNA and DNA reads, much higher than previously found in cloning and Sanger sequencing studies. For example, in a deep anoxic basin this overlap was 27% (Stoeck et al., 2007), in the Beaufort Sea (Canadian Arctic) there was ca 10% overlap (Terrado et al., 2011), and for the Mediterranean Sea the overlap was only 4% (Not et al., 2009). In principal, all of the organisms from RNA should also be recovered from the DNA surveys. The discrepancy is proposed to be because: 1) taxa with few gene copies may be masked by taxa with high 18S rRNA gene copy numbers, and 2) dead microorganisms or even extracellular DNA can be

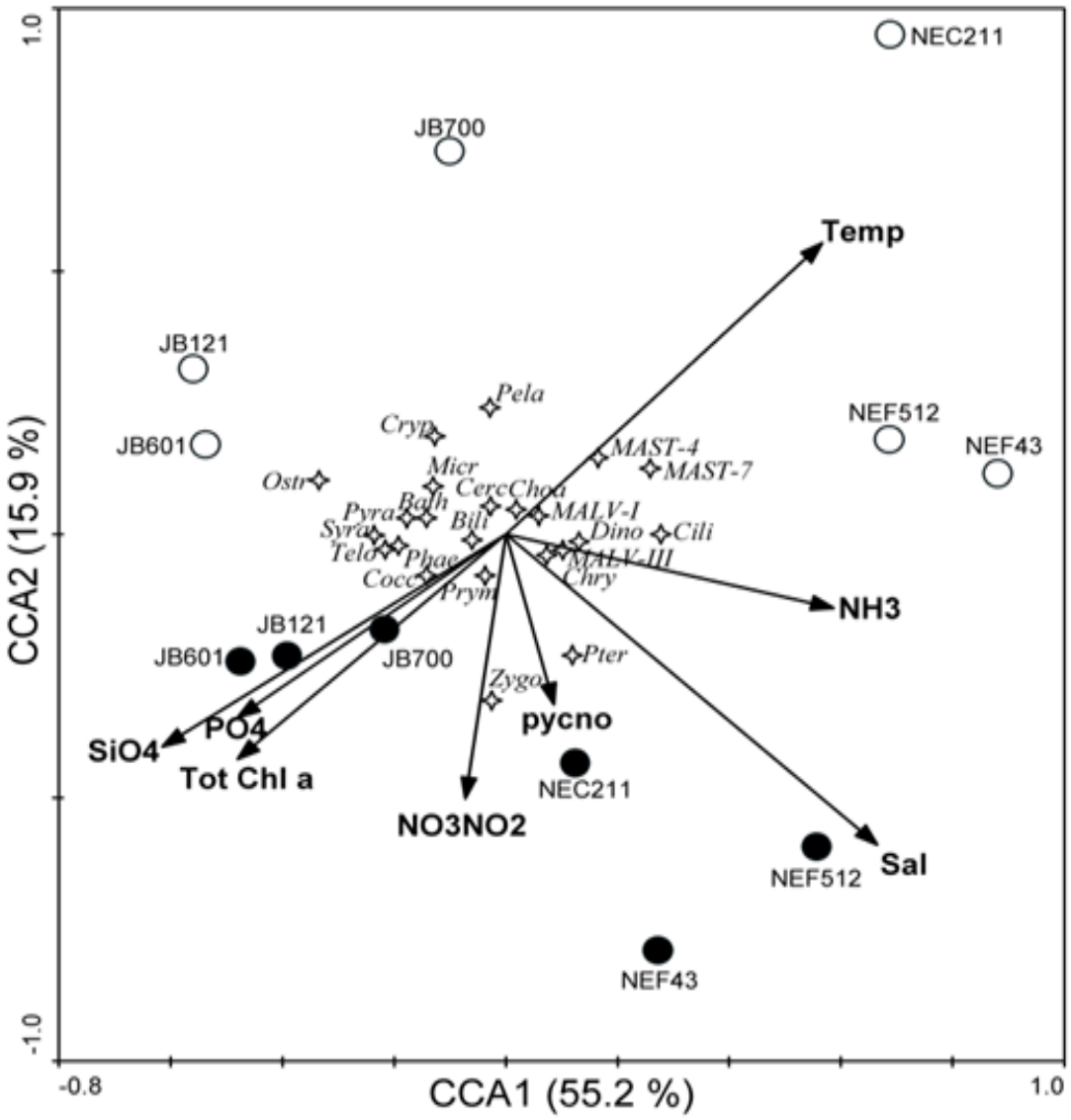


Figure 3.4: Canonical correspondence analysis (CCA) showing the relative abundance of marine microbial eukaryotes taxa (%) recovered by RNA in relation to environmental parameters. Only taxa with a minimum of 1% of total contribution are used and displayed. Surface samples are in white circles and SCM samples are in black circles. Physical factors included temperature (Temp); salinity (Sal); total chlorophyll *a* (Tot Chl *a)*; pycnocline depth (Pycno). Taxa identified were *Bathycoccus* (Bath); Picozoa (Bili); Cercozoa (Cerc); Choanoflagellates (Choa); Chrysophyceae (Chry); Ciliates (Cili); Coccosphaerales (Cocc); Cryptophyta (Cryp); Dinoflagellates (Dino); *Micromonas* (Micr); *Ostreococcus* (Ostr); Pelagophyceae (Pela); Phaeocystales (Phae); Prymnesiales (Prym); *Pterosperma* (Pter); *Pyramimonas* (Pyra); Syracosphaerales (Syra); Telonemia (Telo); Zygodiscales (Zygo).

amplified and sequenced using DNA template (Josephson et al., 1993; Sørensen et al., 2013). In the second case, diversity from DNA surveys will be overestimated; however, our results show that RNA generally provided a higher estimate of total diversity with fewer phylotypes responsible for the majority of reads in the DNA.

At the level of major groups, the number of DNA reads was significantly correlated with number of RNA reads ($R^2 = 0.61$, $p < 0.001$). This result could be due to ribosome numbers being a function of cell concentrations of a particular organism. An often repeated caveat for estimating cell numbers is that larger cells have more 18S rRNA gene copies. Dinoflagellates have multiple rRNA gene copies (Cavalier-Smith, 2005; Zhu et al., 2005), and for that reason are thought to be over-represented in 18S rRNA gene surveys compared to smaller cells. In our case, dinoflagellates were consistently under-represented in the RNA reads despite having high DNA read representation. In contrast, ciliates, which are also large cells with very high 18S rRNA gene copy numbers (Gong et al., 2013), were consistently over-represented in the RNA reads. For small cells, for example Picozoa and *Bathycoccus* which are ≤3 μm, were consistently under-represented in RNA compared to DNA, whereas other smaller taxa were over-represented in RNA, for example MAST -7 and MAST-4, and some taxa showed no consistent correlations, for example the bolidophytes, MAST-1 and MAST-8. Another factor that could influence gene to ribosomal RNA ratios may be polyploidy within a genome, and since polyploidy may result in phenotypic plasticity, the correlation with rRNA copies would not be expected to be significant; the high numbers of gene copies would be maintained independent of protein coding activity since different life stages could conceivably give rise variable numbers of ribosomes actively transcribing mRNA. Clearly, controlled experiments are needed to assess these ratios within individual taxa but overall our results point to the ratio of DNA to RNA as a promising indicator of how the individual taxa may be fairing in their environment.

The global trend masked important differences among taxa. As mentioned above, some taxa were always or nearly always over-represented in DNA compared to RNA. These included, several taxa recovered in all samples with significant correlation between DNA and RNA reads; dinoflagellates, diatoms, Cryptophyta, *Bathycoccus*, *Micromonas*, Cercozoa, MALV (groups I and II), Picozoa, Katablepharida, and *Pyramimonas*.

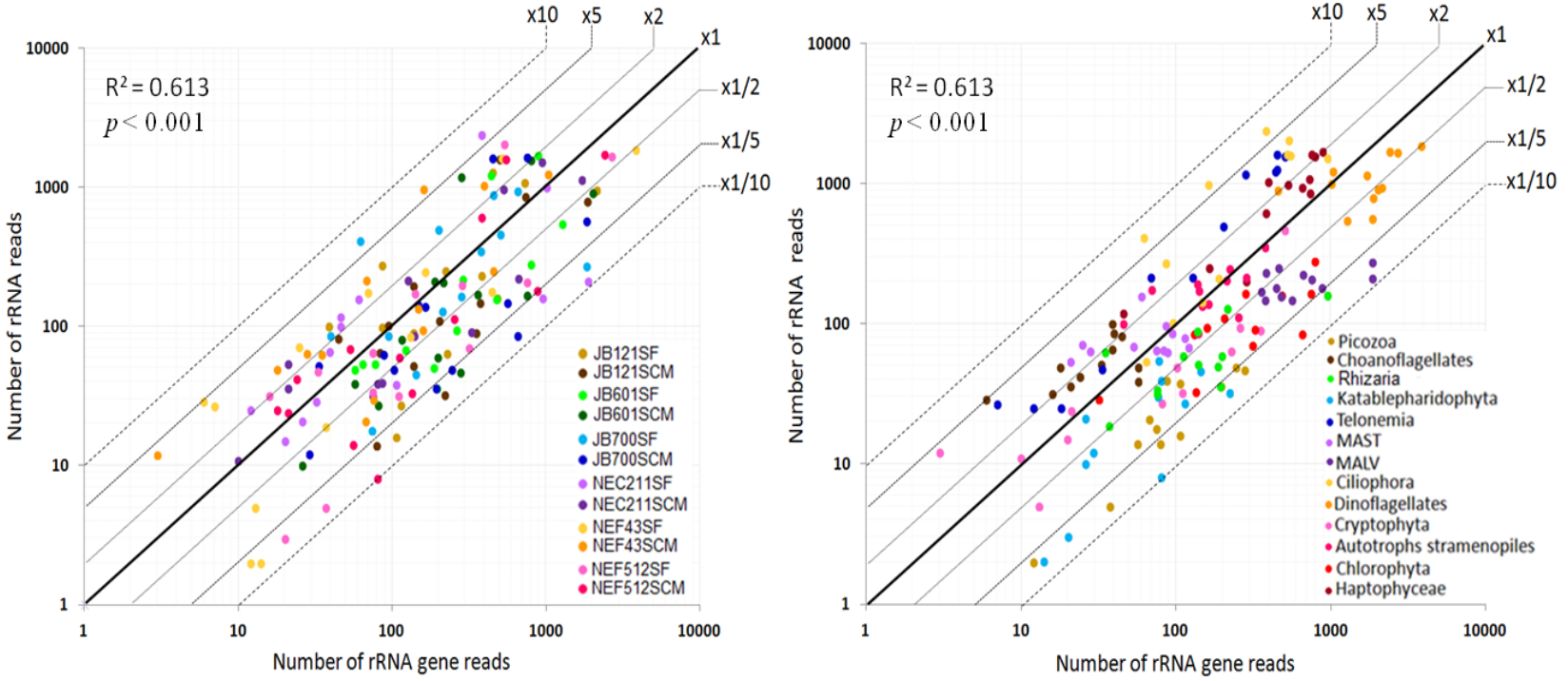


Figure 3.5a: Scatterplots of paired comparisons between number of DNA-reads along the X axis and RNA-reads along the Y axis according to sampling stations (left) and major groups (right).

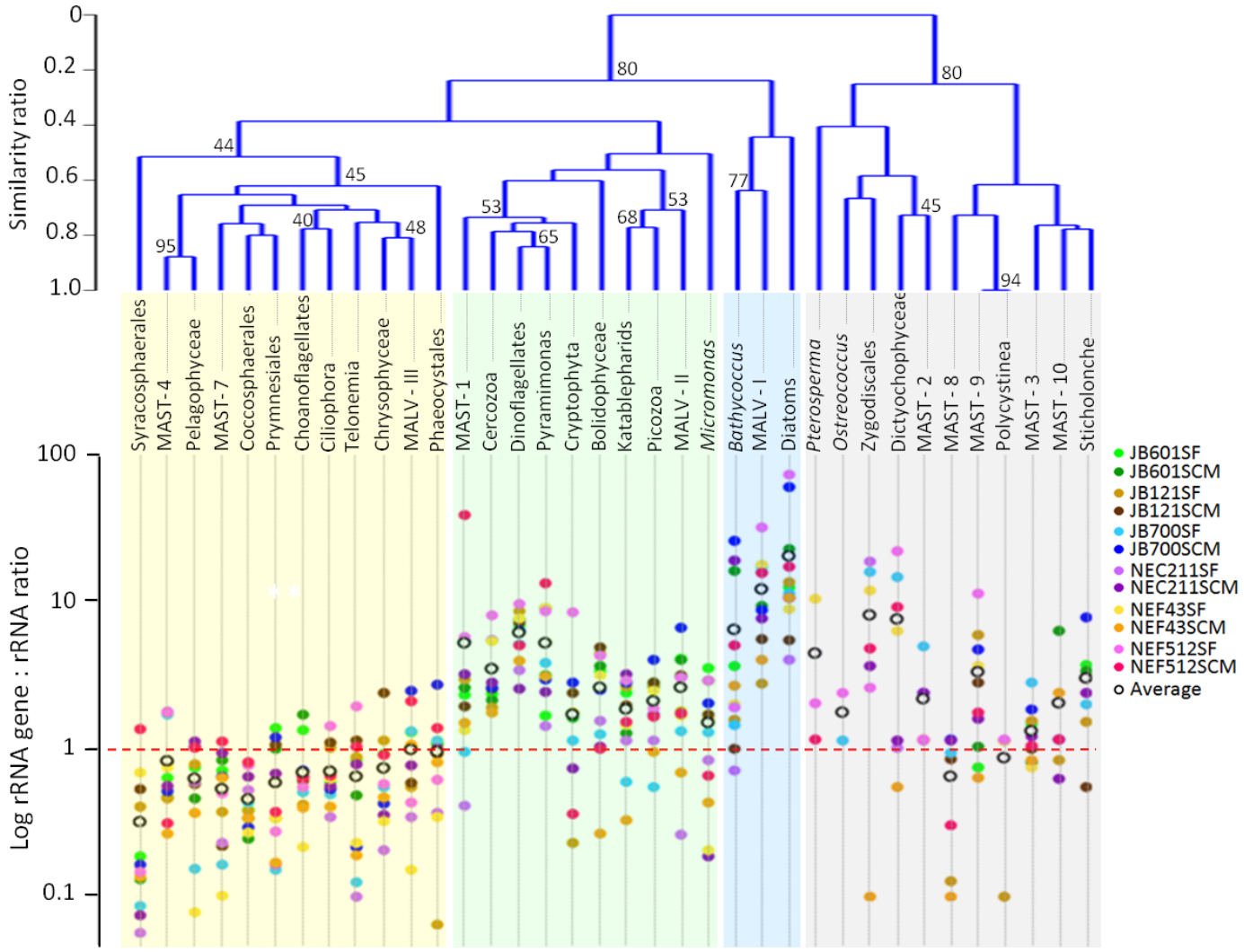


Figure 3.5b: Comparison of DNA to RNA ratio (Y axis, log scale) by taxon (X axis). Taxa were clustered by similarity of the DNA to RNA ratio using cluster analysis (Bray Curtis) run on PAST software. The average of the ratios is shown as an open circle for each taxon, and the individual samples are identified by the same color code in Figure 3.5a left. Red line is 1:1 (where number of DNA reads = RNA reads). Below the line RNA reads > DNA reads, and above the line RNA reads < DNA reads.

These taxa are regularly reported in DNA surveys, and some are often dominant. Terrado et al. (2011) in a cloning and sequencing study following the beginning and decline of an Arctic spring bloom, also noted some of the same groups were rare in the RNA libraries compared to the DNA libraries, in particular dinoflagellates, Picozoa, Cercozoa, Cryptophyta, and MALV. In the same study diatoms, *Bathycoccus* and *Micromonas* were more common using RNA as a template (Terrado et al 2011). We found that several groups were always or nearly always over-represented in RNA, and also showed significant correlations between DNA and RNA reads, these included several haptophyte orders, ciliates, Telonemia, Pelagophyceae, *Pyramimonas*, Chrysophyceae, and MAST-4 and MAST-7. Terrado et al. (2011) also found ciliates and Haptophyta dominant in their RNA libraries. The tendency for the same groups to be over-represented in RNA suggests that when they are present they are normally active, and do not persist using different life strategies such as cysts and spores. Finally, some taxa were detected at only one station or depth. These included *Ostreococcus* (Chlorophyta), Dictyochophyceae, *Polycystinea* (Rhizaria), and Zygodiscales (haptophyta). The variability in the ratios of DNA to RNA of these taxa suggested a capacity to respond to different environmental conditions, consistent with environmental selection.

3.4.2 Biogeography and taxonomic composition of MME community

Microbial community composition is presumably a result of past and current environmental selection. To fully understand factors influencing the distribution and diversity of MME in the GoMA the complex hydrography of the region is relevant. General circulation in the GoMA is relatively well described and spatial patterns are influenced by bathymetry and tidal mixing (Garrett et al., 1978).

Within the JB, strong currents mix the upper water column, a situation that is common in this region in spring and summer (Xue et al., 2000). In the surface layer, there was little difference in the surface and SCM communities, with higher Chl *a* levels in the SCM. At these stations, the euphotic zone was deeper than the surface mixed layer defined by the depth of the pycnocline. In these cases growth was favored by light availability in the surface waters, and re-supply of nutrients. These favorable but constantly changing

Table 3.3 : Linear correlation between DNA-reads and RNA-reads (Significant $p < 0.05$). Standard error (St. Err.) was calculated only for taxa found in all stations; others are no determinate (nd).

	R^2	*p* value	St. Err.
Ciliophora	0.74	0.006	0.76
Telonemia	0.97	<0.001	0.43
Haptophyta	0.90	<0.001	0.44
Dinoflagellates	0.66	0.019	0.43
MALVs	0.49	0.102	0.57
Pelagophyceae	0.98	<0.001	0.63
Pyramimonas	0.97	<0.001	0.48
Zygodiscales	0.93	<0.001	nd
Polycystinea	0.92	<0.001	nd
Prymnesiales	0.87	<0.001	0.43
Cryptophyta	0.85	<0.001	0.55
Picozoa	0.85	<0.001	0.33
Dictyochophyceae	0.83	0.010	nd
MALV -II	0.82	0.001	0.51
Chrysophyceae	0.80	0.002	0.56
Micromonas	0.77	0.003	1.73
Phaeocystales	0.75	0.005	0.67
Syracosphaerales	0.74	0.006	1.10
MAST -3	0.73	0.039	nd
Bathycoccus	0.72	0.009	0.98
Ostreococcus	0.70	0.034	nd
MAST -7	0.66	0.019	0.70
Coccosphaerales	0.65	0.022	0.77
Diatoms	0.62	0.030	1.20
MALV -I	0.60	0.039	0.61
MAST -4	0.59	0.041	0.71
Katablepharids	0.58	0.047	0.67
MAST -1	0.53	0.073	0.70
MAST -8	0.63	0.096	nd
Choanoflagellates	0.45	0.137	0.60
Cercozoa	0.43	0.166	0.62
Sticholonche	0.82	0.183	nd
MALV -III	0.31	0.328	0.48
MAST -2	0.22	0.497	nd
Bolidophyceae	0.18	0.570	1.18
MAST -9	0.19	0.590	nd
MAST -10	0.20	0.629	nd
Pterosperma	0.15	0.669	nd

conditions could explain the co-occurrence of persistent DNA communities and emergent RNA communities (**Figure 3.5**). We noted in that the majority of points from JB stations were below or on the 1:1 line (**Figure 3.5a**) also consistent with a community in transition or flux, with the DNA community being overtaken by taxa more able to take advantage of the immediate conditions. Taxonomic identifications showed that DNA communities were dominated by dinoflagellates, but with more haptophytes and the phagotrophic predator Telonemia mainly recovered from RNA. The favorable conditions for photosynthesis combined with active grazing by bacterivores suggest the production and release of fixed carbon from the phytoplankton fueled bacterial growth that benefited Telonemia (**Figure 3.3**), with dinoflagellates persisting but less active.

Thermoclines were deeper at the NE station (**Figure 3.1, Annexe 3_figure A2**), but the stations were strongly stratified with warmer temperatures in the surface and higher salinity at and below the SCM. In addition, the surface mixed layer was deeper than the euphotic zone, thus the community may have been more light limited. The decrease in diversity within the NE region (**Annexe 3_figure A6**) could also be explained by isolation of the upper layer, resulting in less exchange with adjacent water-masses via mixing and a loss of slower growing or less optimally adapted community members (Howeth et al., 2010). The surface itself was variable, with two of the surface RNA and DNA communities very similar, but more differences in the NEF43 surface samples (**Figure 3.3b**). The community from surface NEC211 station separated from others NEF surface communities perhaps reflecting surface circulation where NEC211 is influenced by surface currents exiting the Gulf, while NEF stations are influenced by surface currents from the south (**Annexe 3_figure A8**).

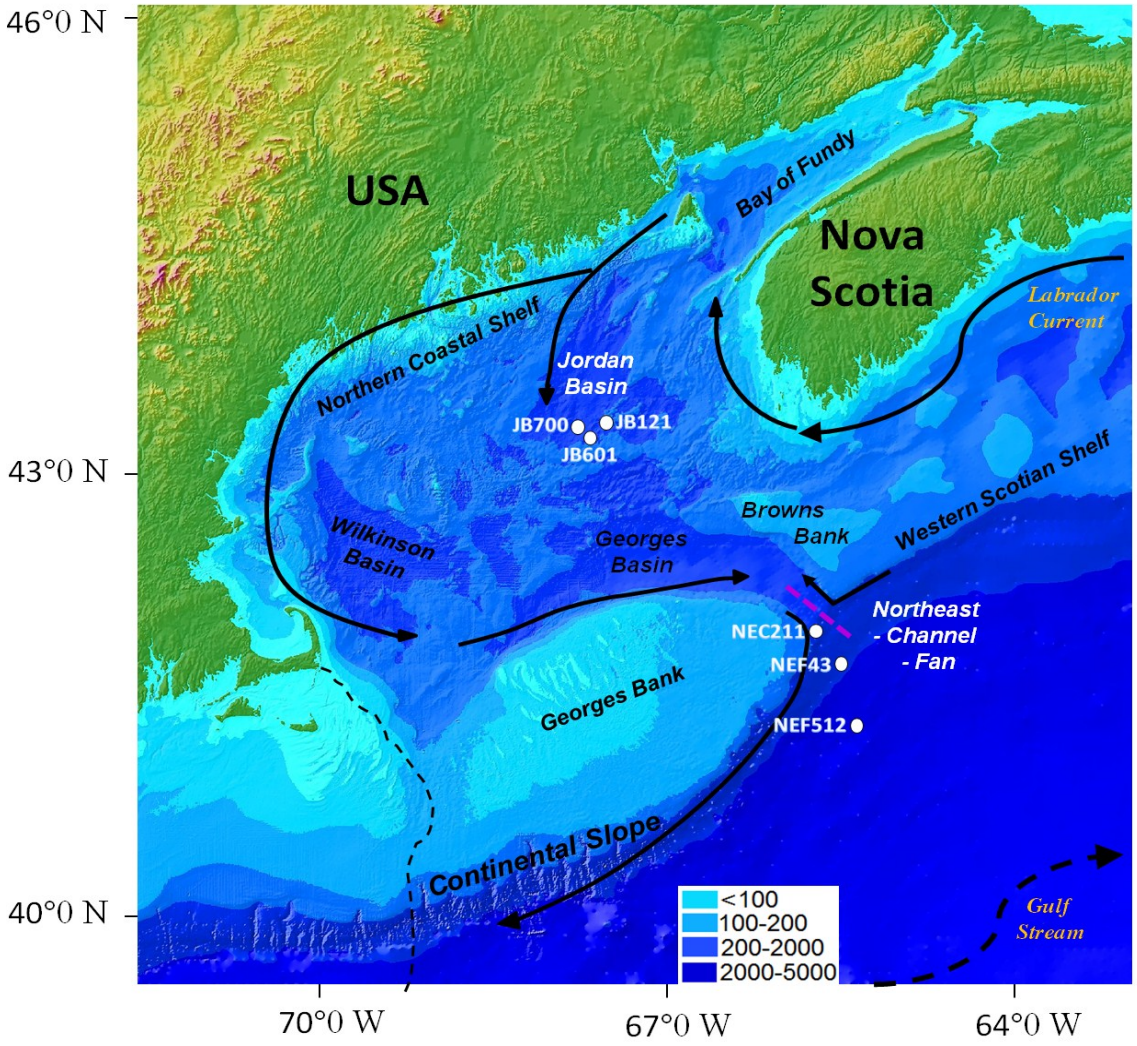
Although it has been suggested that using RNA as a template in 18S rRNA based diversity studies reduces biases inherent from DNA and is as a measure of the active community, we found that using both templates provided additional insights. This study highlighted that results are highly taxon-dependent, and the comparison of DNA and RNA provided complementary information on the state of the communities. We found that the diversity and make up of communities detected by RNA and DNA was highly variable between specific sites and was influenced by environmental conditions. This was consistent with taxa themselves being characteristic of the different environments sampled.

3.5 Acknowledgements

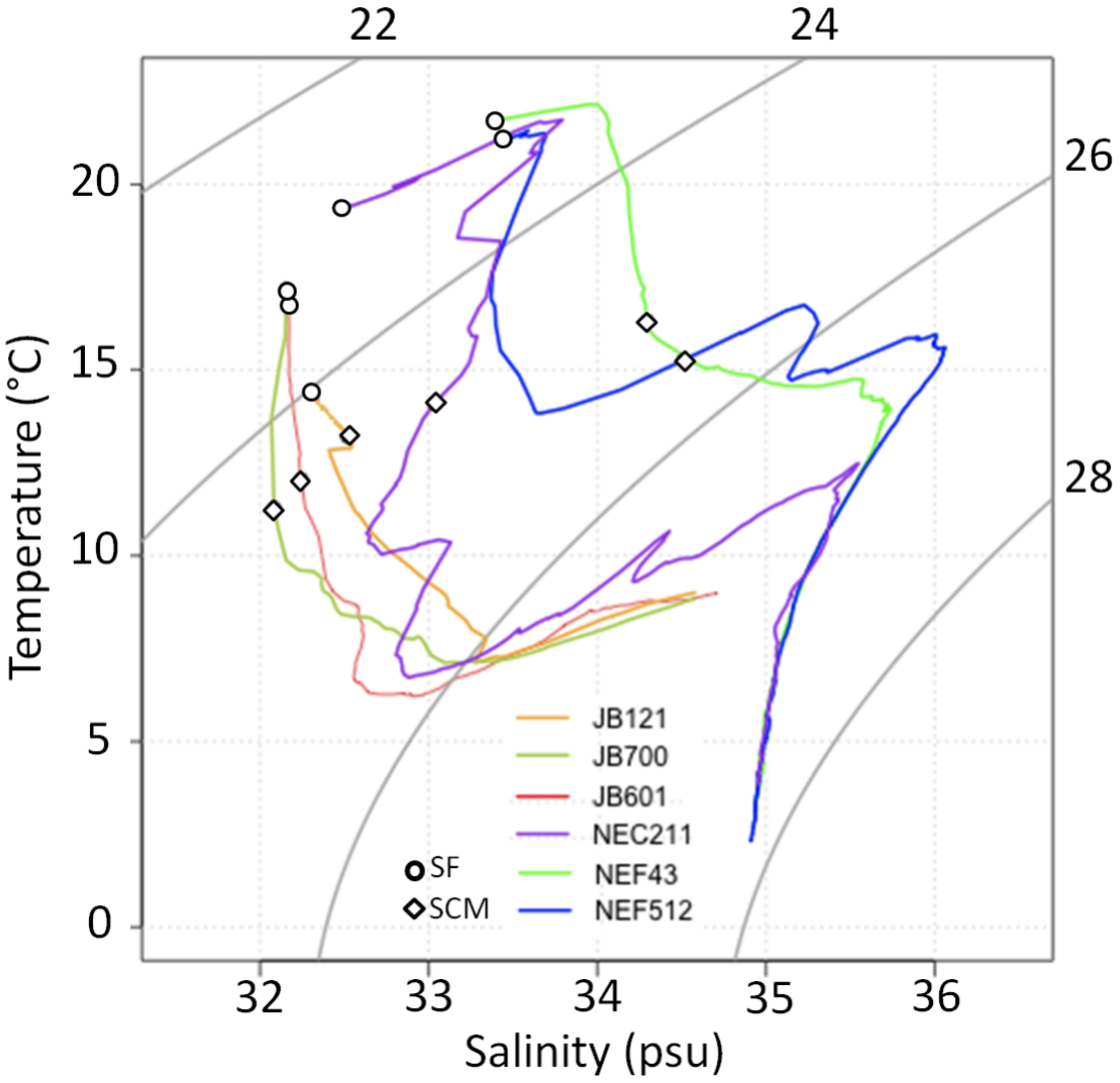
We thank the Canadian Coast Guard officers and crew of the CCGS Hudson for their assistance and support as well as the chief scientist Dr Peter Lawton. We thank Emmanuelle Medrinal for field assistance, Jeff Anning for nutrients and FCM data, Bob Ryan for CTD data, Adam Monier and André Comeau for tutoring in bioinformatics, and Mary Thaler for comments and analytical advice. We thank our colleagues at the IBIS/Université Laval Plateforme d'Analyses Génomiques for the DNA pyrosequencing; the IBIS Centre de Bio- informatique et de Biologie Computationnelle, Compute-Canada and the CLUMEQ Supercomputing Facility. This work is a contribution to the Canadian Healthy Oceans Network (CHONe) funded by the Natural Science and Engineering Research Council of Canada (NSERC). Addition funds were provided by Fonds de recherche du Québec (FQRNT) to Québec Océan and an NSERC discovery grant to CL.

Annex 3

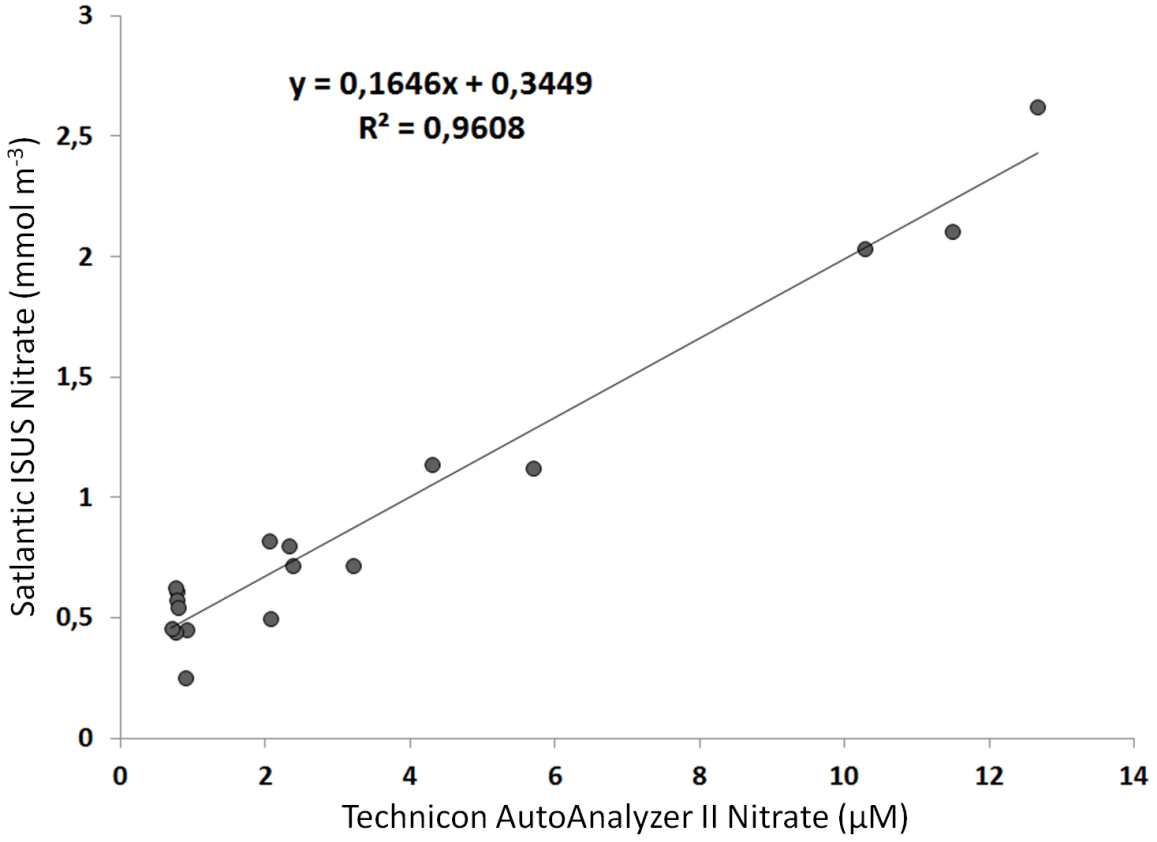
Annex 3_figure A1: Map of sampling sites from Jordan Basin (JB), Northeast Channel (NEC) and Northeast Fan (NEF) regions in the Gulf of Maine. Arrows represent the major surface currents. Map is modified from image created by Gulf of Maine Area Program, Census of Marine Life 15 arc sec**.** *http://pubs.usgs.gov/of/1998/of98-801/bathy/data.htm*.



Annex 3_figure A2: Temperature salinity diagram of sampling stations (surface to bottom) from Jordan Basin (JB) and Northeast channel (NEC) and Fan (NEF). Selected samples for diversity study are represented by point circle in surface and point diamond in subsurface chlorophyll maximum.



Annex 3_figure A3: Relationship between corrected ISUS optical estimates of nitrate concentration determined by UV spectrophotometry and conventional Technicon autoanalyzer II measurements. The optical estimates are highly correlated (r = 0.96) with laboratory measurements.

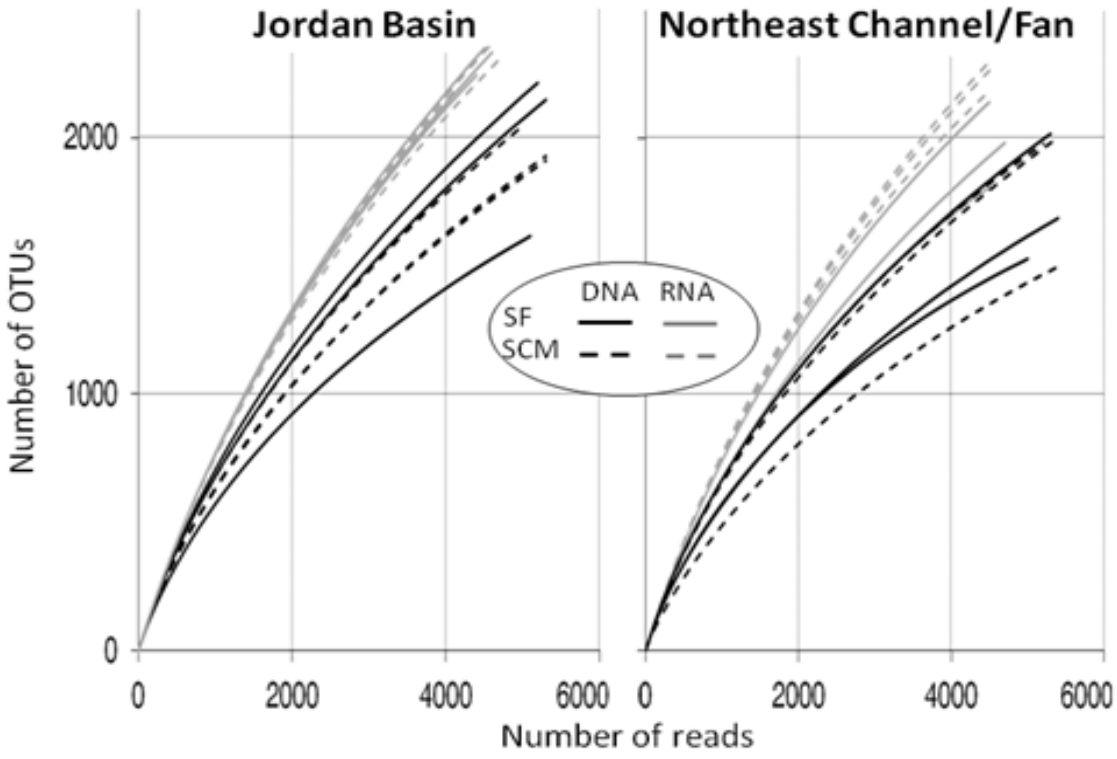


Annex 3_figure A4: Pyrotag raw data, filtering and OTU statistics.

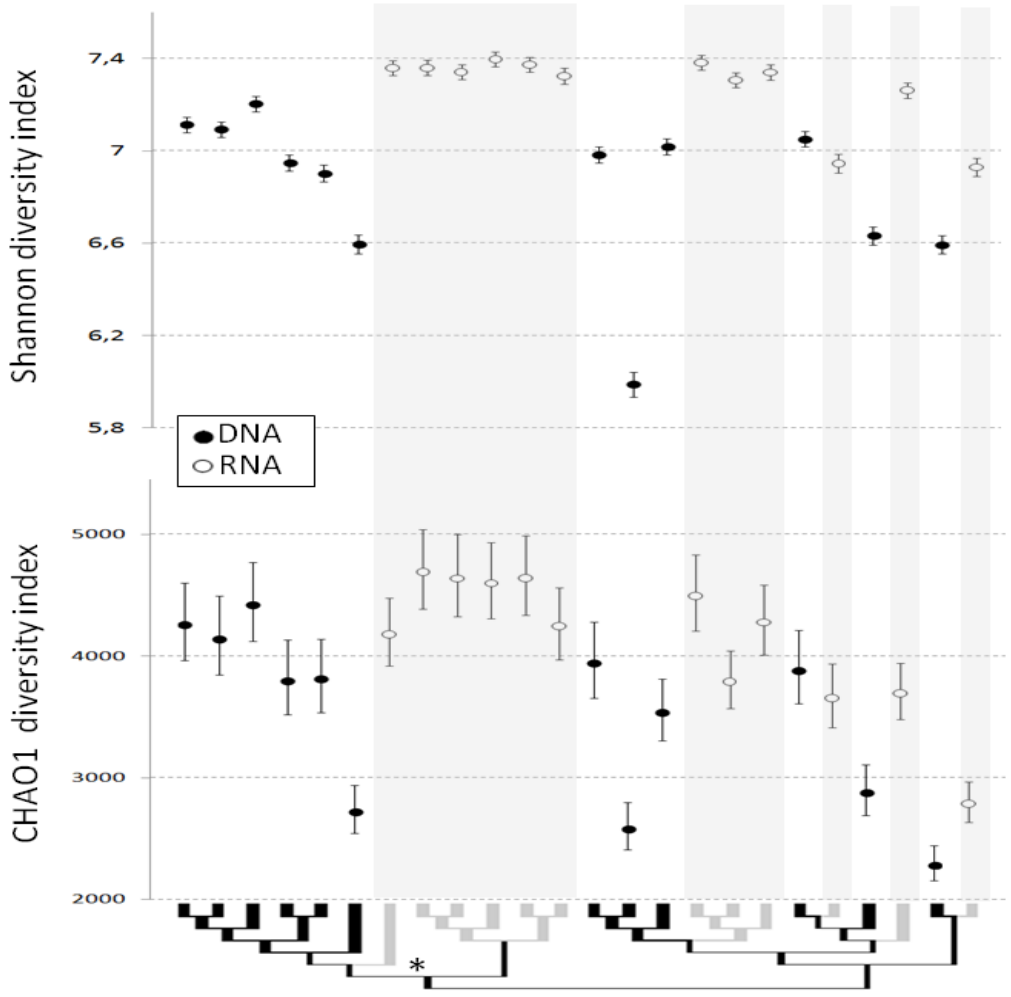
	DNA	RNA
Raw Data		
reads	118 407	106 041
total sequence	43.1 Mb	37.3 Mb
mean length	364 bp	352 bp
Quality Filters		
reads with Ns,<150bp,longer than expected	14 692 (12%)	15 266 (14%)
reads with bad Forward primer	4 347 (4%)	3 376 (3%)
singletons	9 170 (8%)	17 281 (16%)
Final reads		
final good reads (after checking alignment)	63 850	54 675
mean length	401 bp	392 bp
median length	431 bp	426 bp
OTUs Analysis		
mean OTUs/1000 reads	357	487
OTUs (cumulative)[x]	22 763	26 621
OTUs total	15 717	

[x] Cumulative totals when OTUs are splits for each of the 12 bar-codes

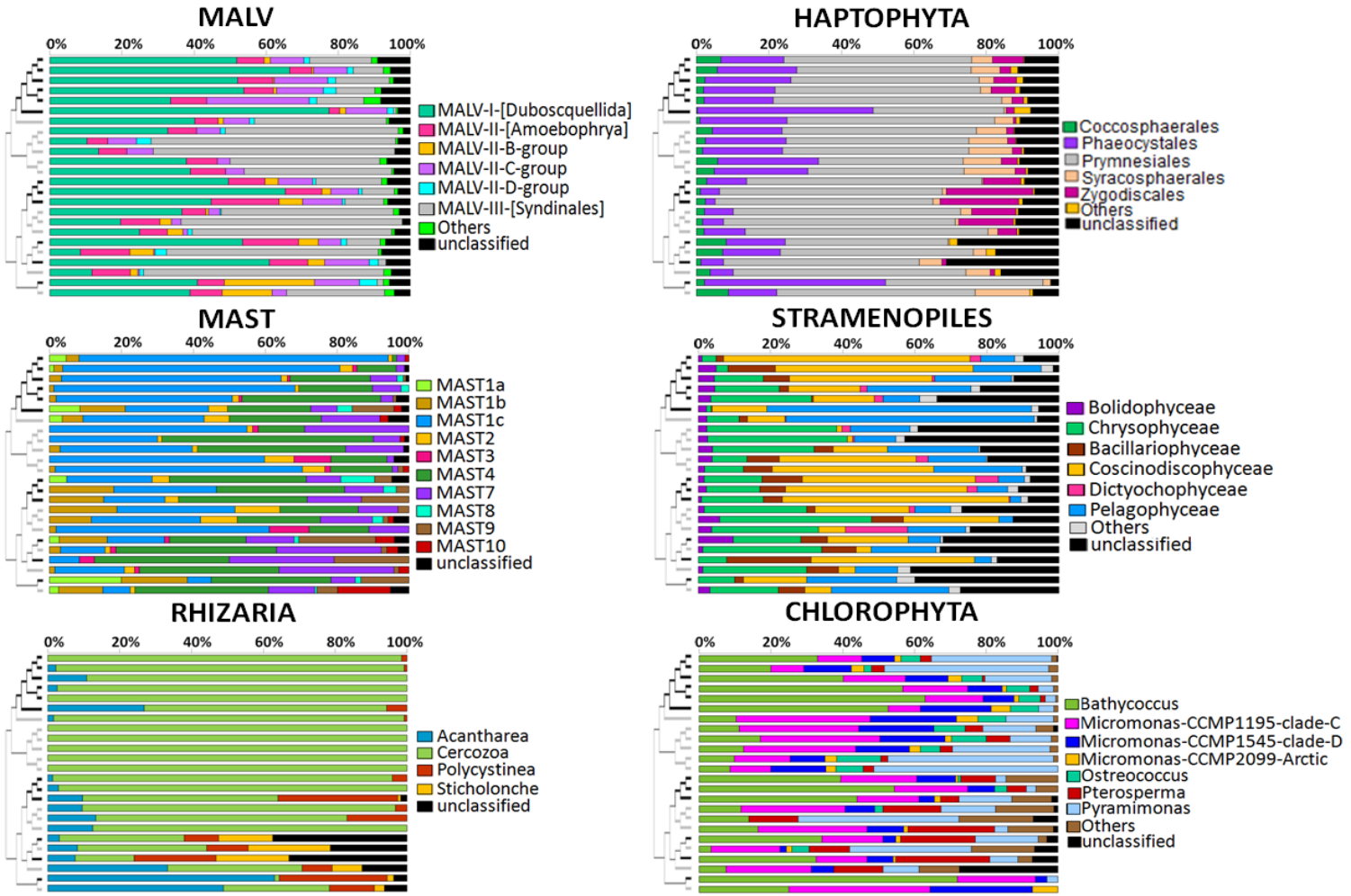
Annex 3_figure A5: Rarefaction curves of similarity-based operational taxonomic units (OTUs) at cluster distance values of 0.02 from both DNA- and RNA-based pyrosequencing.



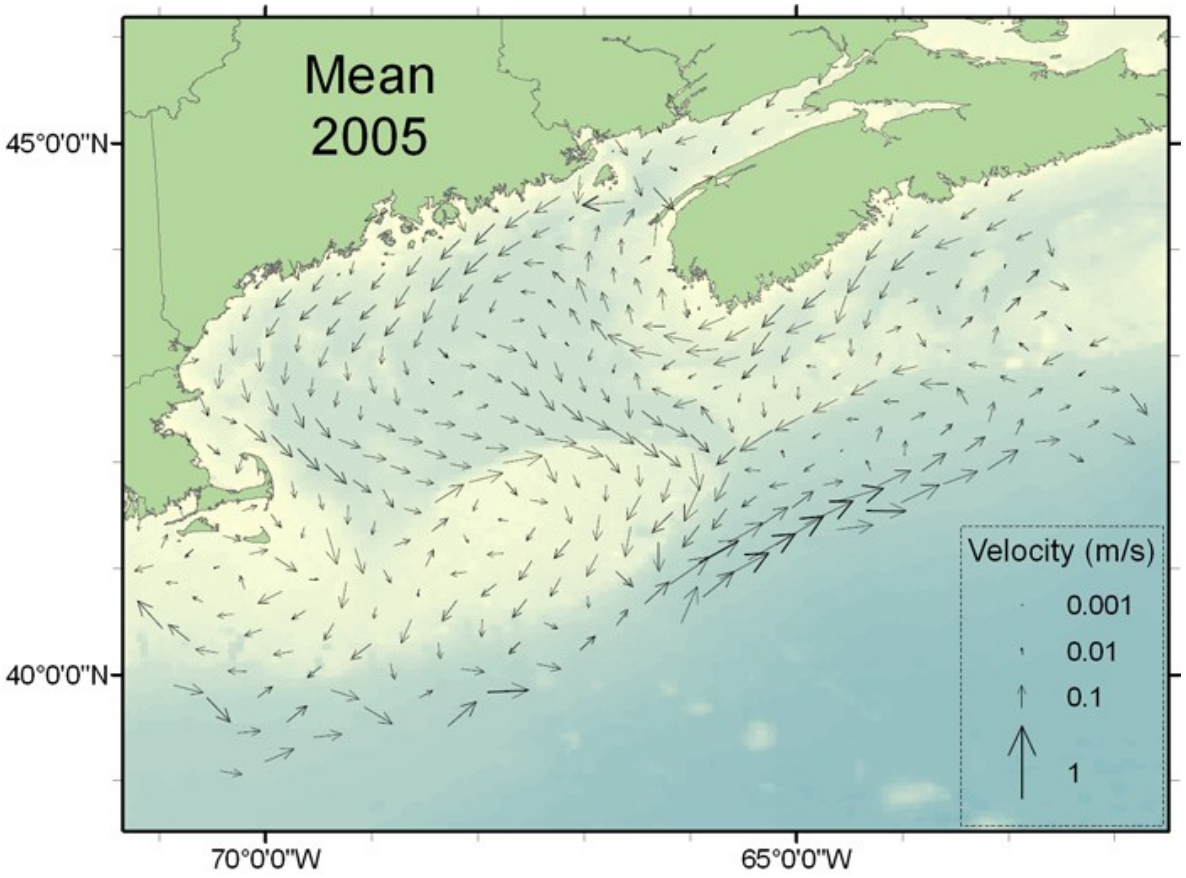
Annex 3_figure A6: Comparisons of species richness estimates of each community sample (at ≥98% OTUs level) from both DNA- and RNA-based pyrosequencing according to UniFrac clustering (figure 2b). The symbol '*' indicates that the diversity index (Shannon and Chao1) between DNA and RNA from Jordan Basin were significant (two-sample t-tests, P≤0.05).



Annex 3_figure A7: Distribution of relative abundance of taxa within heterotrophs (left) and phototrophs (right) phyla known to contain smaller cells. Classification others corresponding to MALV (marine Alveolates): group-II-E, -II-F, -II-G, -IV; Haptophyta: Coccolithales, Isochrysidales, Pavlovales, Reticulosphaerales; Chlorophyta: *Dunaliella*, *Dolichomastix*, *Mamiella*, *Mantoniella*, *Halosphaera*, *Monomastix*, *Nephroselmis*, *Prasinoderma*, *Pseudoscourfieldia*, *Nannochloris*.



Annex 3_figure A8: Mean 2005 surface current velocity vectors for the Gulf of Maine. Data are from the Gulf of Maine Princeton Ocean Model. Tiré de http://www.gulfofmaine-census.org/about-the-gulf/oceanography/circulation/.



Chapitre 4 – Structure and phylogenetic diversity of active photosynthetic picoplancton by photosystem-II psbA transcript in the Gulf of Maine

Résumé

Le picophytoplankton (PPP) marin est une composante fondamentale dans les processus biogéochimiques océaniques. Bien que les facteurs environnementaux qui déterminent la distribution et la diversité phylogénétique des picocyanobactéries soient relativement bien identifiés, la situation est moins claire pour les picoeucaryotes photosynthétiques (PPE). Dans cette étude, nous avons étudié les variations spatiales et la diversité des PPE et des picocyanobactéries (*Synechococcus* et *Prochlorococcus*) en utilisant le clonage et le séquençage Sanger des transcrits *psbA* à la surface et au maximum de chlorophylle subsurface (SCM) dans le golfe du Maine (GoMA) en Juillet 2010. *Synechococcus* était presque toujours numériquement dominant, et *Prochlorococcus* n'a été détecté que dans les stations situées à la sortie du GoMA. Les clades de picocyanobactéries avaient une distribution restreinte, où la majorité des *Prochlorococcus* appartenaient à l'écotype HL-II principalement reporté au SCM, et l'écotype LL-I principalement reporté dans les échantillons de surface. Les *Synechococcus* étaient représentés dans cinq clades du sous-groupe 5.1 (clades II, III, IV, IX principalement reportés au SCM, et le clade I dominant dans les eaux de surface). Les PPE ont été présents dans tous nos échantillons avec de grandes différences entre les stations et les profondeurs mais la mise en évidence de la séparation des niches était moins claire. La majorité des séquences *psbA* étaient affiliées à Chlorophyta, tandis que les Haptophyta étaient les plus divers phylogénétiquement. D'autres séquences *psbA* étaient affiliées aux Cryptophyta et à plusieurs classes de stramenopiles (Dictyochophyceae, Chrysophyceae, diatomées, et Pelagophyceae). Un certain nombre de clones n'a pas pu être identifié au niveau du genre suggérant la nécessité d'analyses phylogénétiques du *psbA* à partir de plusieurs cultures, et l'utilisation du « tri cellulaire » suivi du séquençage afin d'augmenter le nombre de taxons dans les bases de données de référence.

Abstract

Marine picophytoplankton (PPP) are essential components of oceanic biodiversity and important in biogeochemical processes. While environmental factors that drive the distribution and phylogenetic diversity of picocyanobacteria are relatively well known, the situation is less clear for photosynthetic picoeukaryotes (PPE). In this study, we investigated spatial variability and diversity in PPE and picocyanobacteria (*Synechococcus* and *Prochlorococcus*) by way of cloning and Sanger sequencing *psbA* transcripts from the surface and subsurface chlorophyll maximum (SCM) in the Gulf of Maine (GoMA) from samples collected summer 2010. Flow cytometry data indicated that *Synechococcus* were almost always numerically dominant, and *Prochlorococcus* was detected only at stations outside GoMA. Picocyanobacteria clades had restricted distributions. The majority of *Prochlorococcus* belonged to the HL-II ecotype mainly recovered from SCM, and LL-I ecotype was mainly recovered from surface samples. *Synechococcus* were represented by 5 clades of sub-cluster 5.1 (clades II, III, IV, and IX mainly from SCM, and clade I mainly from surface waters). PPE were found in all samples with large differences among stations and depths but evidence for niche partitioning was less clear among picoeukaryotes. The majority of *psbA* sequences were affiliated to Chlorophyta, while Haptophyta were the most diverse. Other *psbA* sequences were from Cryptophyta and photosynthetic stramenopiles (Dictyochophyceae, Chrysophyceae, Bacillariophyta, and Pelagophyceae). A number of clones could not be identified to genus level suggesting the need for phylogenetic analyses of *psbA* from more cultures and possible single cell sorting and sequencing to increase the number of taxa in reference databases.

4.1 Introduction

Picoplankton are defined as cells between 0.2 and 2 μm (Sieburth et al., 1978) but in practice cells up to 3.0 μm in one dimension are also referred to as picoplancton (Charpy and Blanchot, 1998; Vaulot et al., 2008). Picophytoplankton (PPP) are ubiquitous across marine ecosystems, contribute to in carbon cycle, and account for much of the primary production in the ocean (Agawin et al., 2000). PPP include picocyanobacteria with two main genera *Synechococcus* and *Prochlorococcus*; and diverse and polyphyletic microbial eukaryotes. Marine *Synechococcus* and *Prochlorococcus* have been extensively studied in terms of diversity and global distributions, with a large number of studies describing the environmental selective forces that determine distribution of genetically distinct ecotypes (Campbell et al., 1994; Partensky et al., 1999b). In comparison, studies of photosynthetic picoeukaryotes (PPE) are fewer; diversity has been recognized only more recently (Vaulot et al., 2008; Worden and Not, 2008; Massana, 2011). Several studies have shown the significant contribution by PPE to picophytoplankton biomass and annual primary production by PPE may exceed that of cyanobacteria (Li, 1995; Worden, 2004). An important step in understanding the distribution and competitive advantages of PPE compared to cyanobacteria is to first identify phylogenetic differences under different environmental conditions and whether specific taxa could be under direct environmental selections.

The majority of studies on PPE diversity have used the 18S rRNA gene as a molecular marker (Lovejoy et al., 2006; Not et al., 2007, Vaulot 2008), and these studies recovered heterotrophs as well. Alternative approaches such as plastid 16S rRNA gene phylogenies, often in conjunction with flow cytometric sorting, have allowed more direct targeting of PPE (Fuller et al., 2006; Kirkham et al., 2011; Shi et al., 2011). Other studies have focused on functional photosynthetic protein-coding genes, in particular *psbA* which codes for the D1 protein in the photosystem-II reaction center (Morden et al., 1992; Scherer et al., 1991; Zhang et al., 2000; Man-Aharonovich et al., 2010). This gene is involved in photosystem (PS) II, as part of the structure of photosynthetic machinery highly conserved. PS II is prone to light-induced oxidative damage due to the highly oxidative chemistry of splitting water (Barber et al, 1992). PS II is composed of core proteins D1 and D2. The D1 protein is the primary target of oxidative damage, and it is sacrificed to ovoid complete

inactivation and disassembly of PS II. Therefore, under normal photosynthetic growth conditions the D1 protein, encoded by the *psbA* gene, is constantly degraded and re-synthesized in the multistep PS II repair cycle (Aro et al., 1993b; Mulo et al., 2012). The *psbA* gene is a good taxonomic indicator for a variety of photosynthetic marine microbial groups, both the Eukaryota and Cyanobacteria (Zeidner et al., 2005). The *psbA* gene is also found in cyanophages infecting *Prochlorococcus* (Lindell et al., 2004) and *Synechococcus* (Mann et al., 2003).

The Gulf of Maine (GoMA) is a large, semi-enclosed shelf sea of the Northwest Atlantic Ocean bordered by the New England coastline of the United States and the eastern maritime provinces of Canada. This distinct large marine ecosystem has a long history of plankton studies (Mills, 1989), including records from the Continuous Plankton Recorder project, which since 1948 provides a record of zooplankton and large phytoplankton in the northwest Atlantic (Jossi et al., 1993; Kane, 2011). Less attention has been paid to small cells including PPE (Li et al., 2011) that occur year round (Li, 2009). In this study, six stations from the GoMA were selected because of their contrasting hydrographic characteristics. Three stations were selected with relatively low temperature, low salinity from the Jordan Basin (JB), one of three large basins in the GoMA. Three other stations sampled from the North East (NE) Channel and Fan were warmer with higher salinity. The NE channel is the major corridor between the GoMA and the Northwest Atlantic Ocean. Nutrients are low at this time of year in both regions, favoring picophytoplankton over larger cells. Our first goal was to estimate phylogenetic diversity of probable PPE taxa and picocyanobacteria that would be potentially metabolically active and therefore involved in primary production. We chose to target *psbA* transcripts from RNA collected at the six stations. Next, we determined spatial distribution and variability in PPP assemblages by comparing proportions of picocyanobacteria and PPE transcripts. Finally, communities were analyzed along with physical-chemical characteristics to determine what environmental factors may have influenced the microbial eukaryote distributions.

4.2 Materials and methods

4.2.1 Study region and general sampling

Samples were collected aboard the *CCGS Hudson* from 28 July to 11 August 2010 from two distinct regions in the Gulf of Maine: three stations in Jordan Basin (JB) and three stations along North East (NE) Channel and Fan (**Annex 4_figure A1**). Discrete surface water samples were collected at 2 or 3 m, and the subsurface chlorophyll maximum (SCM) determined from the fluorescence profile, directly from 12-L Niskin bottles mounted on a Rosette equipped with profilers for conductivity, temperature and depth (CTD; SBE-911 CTD; Sea-Bird Inc., Bellevue, WA, USA), oxygen (Seabird Model SBE-43 dissolved oxygen sensor), *in situ* chlorophyll fluorescence (Chelsea Aquatracka MK III), relative nitrate concentration using the *in situ* ultraviolet spectrophotometer (ISUS) probe (Satlantic, Halifax NS Canada), and a LI-COR spherical Quantum Sensor (model LI-193SA) that provided photosynthetically available radiation (PAR) measurements (400-700 nm). Profiles to 100m were acquired on the downcast, and water samples were collected on the upcasts.

Samples for nutrients, extracted chlorophyll *a* (Chl *a*), and nucleic acids were collected directly from the Niskin bottles. Samples for nutrients were immediately frozen at -20°C. Chl *a* samples were collected by filtering 500 mL of seawater onto GF/F filters (Whatman, GE Healthcare, Piscataway, NJ, USA) and were stored at -80°C until analysis. For nucleic acids, 6 L of water was sequentially filtered through a 50-µm nylon mesh, a 47-mm diameter 3-µm pore size polycarbonate (PC) filter!(Millipore), and finally a 0.2-µm pore size Sterivex unit (Millipore Canada Ltd, Mississauga, ON, Canada). RLT™ buffer with β-mercaptoethanol was added to the Sterivex units which were stored at -80°C until nucleic acid extraction. Only material from the Sterivex units (3 to 0.2 µm fraction) was used for this study. This was meant to enrich for cells <3 µm in one dimension (Vaulot et al., 2008; Seenivasan et al., 2013).

Maximal PS II photochemistry efficiency (Fv/Fm) was measured using a pulse-amplitude-modulated (PAM) fluorometer (Water-PAM, Walz). Water collected from all stations both depths was filtering on 3-µm filters (small fraction) and realized in triplicate. Prior to measurement, the PAM was zeroed with 4 mL of sample water filtered through a

0.2-µm pore size sterile filter. For each measure 4 mL of sample water was dark-adapted for 30 min to relax photosynthetic activity. Next, a low beam was used to measure minimum fluorescence level F_0 of the dark-adapted sample, followed by a saturating white-light pulse (4000 µmol quanta m^{-2} s^{-1}) to measure maximal fluorescence, Fm. The maximum quantum yield of PS II photochemistry was then estimated as: Fv/Fm = (Fm-F_0)/Fm.

4.2.2 Laboratory Protocols

Nutrients (nitrite and nitrate referred to together as nitrate, soluble reactive phosphorus, and silicic acid) were analyzed spectrophotometrically as in Mitchell et al. (2002). For ammonium, the fluorometric method of Kerouel and Aminot (1997) was used. Nutrient ratios were compared with Redfield ratios calculated on the basis of the following stoichiometries: Si : N : P = 15 : 16 : 1 (Redfield, 1958; Arrigo, 2005).

Total Chl *a* and small size fraction (≤3µm) Chl *a* were determined by high performance liquid chromatography (HPLC) (Hooker et al., 2005). FCM samples were run at Bedford Institute of Oceanography on a FACSort Flow cytometer (Becton Dickinson) as described in Li and Dickie (2001). Three groups of autotrophs (cells with chlorophyll autofluorescence) were distinguished: *Prochlorococcus*, *Synechococcus*, and photosynthetic picoeukaryotes (PPE), which were nominally 0.8 to 2.0 µm. We compared PPE to picocyanobacteria in terms of average carbon content by converting cell concentrations to estimated carbon biomass (C) for the three picophytoplankton categories using the suggested average conversion factors from Buitenhuis et al. (2012). The values used were: 36 fg C cell-1 for *Prochlorococcus*, 255 fg C cell-1 for *Synechococcus*, and 2590 fg C cell-1 for PPE.

RNA was extracted from the Sterivex units using AllPrep DNA/RNA Micro Kits (QIAGEN). Extraction followed the manufacturer's instructions with a cell lysis step, where 600 µl of RLT Plus, 85 µl SDS 10%, and 30 µl proteinase K were added to the Sterivex units for 15 min at 65°C. Extracted RNA was immediately reverse transcribed to cDNA using a High Capacity Reverse Transcriptase Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's recommendations. The resulting cDNA quality was checked on a 1% agarose gel. We then amplified *psbA* from cDNA using degenerate

PCR primers (Wang and Ghen, 2008) that target the conserved YPIWEA and HNFPLD regions (~750pb, **Annex 4_figure A2**). These primers are considered universal and amplify *psbA* from eukaryotes, cyanobacteria, and cyanophages (Tzahor et al., 2009). PCR was performed in a total volume of 25µl containing 10 ng of template, 2.5 µl of 10X buffer, 2 µl of dNTP, 1.3 µl of 50 mM $MgCl_2$, 1µl of 25 uM psbA-1F (5'TAY CCN ATY TGG GAA GC'3) and 1 µl of 25 uM psbA-2R (5'TCR AGD GGG AAR TTR TG'3), and 1.2U of Feldan *Taq* DNA polymerase (Feldan-Bio). Amplification conditions included steps at 95°C for 2 min, and 30 cycles at 94°C for 1 min, 55°C for 1 min, and 68°C for 1 min followed by a final 10 min at 72°C using the iCyclerTM Thermal Cycler (Bio-Rad Laboratories, Inc., CA, USA). An aliquot of the PCR product from each sample was ligated and cloned using the Qiagen PCR cloning kit following the manufacturer's instructions. Putative positive colonies were picked and transferred to 96-multiwell plates with Luria-Bertani medium, ampicillin and 7% glycerol and incubated 8 to 16 h at 37°C before being stored at –20°C until screening. One 96-well plate for each library was screened and clone inserts were verified by PCR amplification using the same primers as for the clone libraries. Overall, total of 1112 clones from 12 clone libraries were sequenced using the original forward primers at Service de séquençage et génotypage du Centre Hospitalier de l'Université Laval (CHUL) using an ABI 3730xl system (Applied BioSystems), which included a purification step.

4.2.3 Phylogenetic community structure analyses

PsbA sequences were aligned using the MAFFT alignment software version 6 with default parameters (Katoh and Frith, 2010) and manually inspected using the BioEdit Sequence Alignment Editor (Hall, 1999). This process identified 11 poor sequences (1 % of sequences) that were discarded. Next, we applied a rapid initial taxonomic assignment to identify major divisions (eukaryotes, bacteria, and cyanophages) using *Mothur* v1.21.1 (Schloss et al., 2009) against our in-house curated reference database of 113 *psbA* sequences based on NCBI taxonomy. A total of 117 of the *psbA* sequences were from cyanophages (11 % of sequences) and were excluded for following analysis. Operational taxonomic units (OTUs) were generated using furthest neighbor clustering by Mothur at the 1% cutoff level difference in nucleic acid sequence in the alignments (Mothur; Schloss et

al., 2009). For each OTU, the closest match and published reference sequences were found using the basic local alignment search tool (BLAST) (Altschul et al., 1997) against the nt/ nr NCBI data base (**Annex 4_table A1**). The presence of chimeras was tested using UCHIME (Edgar et al., 2011) and sequences with low BLAST scores were cut into three separate sections and independently checked against in the GenBank nr database using BLASTn. All suspected chimeras (18 % of sequences) were removed and not analyzed further.

Diversity and phylogenetic relationships among *psbA* transcripts was visualized by constructing Maximum Likelihood (ML) trees of OTUs at 99% similarity, for *Synechococcus*, *Prochlorococcus*, and PPE level using MEGA version 5.1 (Tamura et al., 2011). The ML method was based on the General Time Reversible Model (Nei and Kumar, 2000). All positions with less than 95% site coverage; meaning sequences with fewer than 5% alignment gaps, missing data, or ambiguous bases at any position, were removed. Node support was assessed via 100 bootstrap replicates.

The taxonomic composition and similarities of PPE and picocyanobacteria communities were explored using phylogenetic diversity analyses in UniFrac (Hamady et al., 2010). UniFrac distances were based on the *psbA* phylogenies of each domain, with 622 sequences for eukaryotes and 158 sequences for cyanobacteria, and visualized in FastTree version 2.1 (Price et al., 2010) run in "accurate mode" (-mlacc 2 –slownni) with the general time reversible (GTR) model and pseudo-counts. Weighted UniFrac distance metrics for both cluster analysis and Principal Coordinates Analysis (PCA) were generated separately for PPE and picocyanobacteria.

4.2.4 Statistical analyses

Rarefaction curves and OTUs were calculated in *Mothur* v1.21.1 (Schloss et al., 2009). Determination of the similarity level of the OTUs is important for performing diversity studies using environmental sequences, and several cutoffs were used for statistical analysis (3%, 2% and 1%, **Annex 4_figure A3**). Following this, sequences with 99% similarity were clustered as OTUs and used for phylogenetic analysis (Koeppel and Wu, 2013).

Table 4.1: Sampling locations, Chlorophyll *a* concentrations, and physicochemical parameters. Total (Tot Chl *a*) and small size fraction (Chl *a* ≤3µm) chlorophyll *a* concentration; Temperature (Temp); Salinity (Sal); Photosynthetically Available Radiation (PAR); nutrients: ammonium (NH_4^+), nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^{4-}).

Stations	Lat. (°N)	Long. (°W)	Date	Local time	Depth (m)	Tot. Chl *a* (μgL^{-1})	Chl *a* **≤3µm (µgL^{-1})**	Temp (°C)	Sal	PAR ($\mu mol/m^{-2}$ sec^{-1})	NH_4^+ (µM)	NO_3^- (µM)	PO_4^{3-} (µM)	SiO_4^{4-} (µM)
JB121	43°19	67°03	31 Jul	7:00	2	1.09	0.88	15.7	32.2	44.0	0.37	1.04	0.49	2.21
					14	0.72	0.34	13.6	32.3	4.1	0.50	2.28	0.69	2.84
JB700	43°17	67°22	31 Jul	20:00	2	1.00	0.59	17.3	32.1	61.3	0.49	0.81	0.27	1.59
					14	0.01	0.01	10.7	32.1	6.5	0.40	0.87	0.31	1.69
JB601	43°11	67°14	30 Jul	20:00	2	0.42	0.26	17.1	32.2	55.2	0.37	0.99	0.33	2.05
					12	0.93	0.67	15.4	32.2	10.5	0.40	2.49	0.66	3.56
NEC211	41°54	65°30	09 Aug	05:30	2	0.23	0.23	19.3	32.5	0.4	0.62	0.90	0.26	0.77
					51	0.57	0.56	11.5	32.8	0.0	0.60	2.66	0.64	2.89
NEF43	41°47	65°20	05 Aug	20:30	2	0.22	0.17	21.8	33.5	5.1	0.41	0.87	0.09	0.41
					30	0.53	0.38	14.9	34.8	0.2	0.48	0.90	0.23	1.43
NEF512	41°16	65°15	07 Aug	18:20	2	0.14	0.12	21.2	33.4	819.9	0.44	0.87	0.11	0.27
					39	0.28	0.23	16.0	34.7	8.2	0.70	2.33	0.48	2.38

Canonical Correspondence analysis (CCA) was carried out in the Paleontological Statistics Software Package (PAST) to identify significant correlations among physical variables (temperature, salinity, PAR, Redfield ratios) with biological variables (Total Chl *a*, Small fraction Chl *a*, relative abundance of three picophytoplankton categories from FCM: PPE, *Synechococcus,* and *Prochlorococcus*).

4.3 Results

4.3.1 Environmental variables

The July surface waters from Jordan Basin (JB) were relatively warm (10°C to 17°C) with lower salinity (32.1 to 32.3) compared to the North East (NE) stations, where waters were warmer (19°C to 22°C) and more saline (32.5 to 33.5) (**Table 4.1**). Subsurface chlorophyll maximum (SCMs) were detected at around 15 m in JB, and between 30 to 50 m in the NE stations (**Table 4.1**). Nutrient concentrations in surface waters were low in both regions, and higher below the pycnocline. Nutrients concentrations at the JB stations ranged from 0.8-2.5 μM for nitrate, 0.4-0.5 μM for ammonium, 0.3-0.7 μM for phosphate, and 1.6-3.6 μM for silicate (**Table 4.1**). At the NE stations concentrations ranged from 0.9-2.7 μM for nitrate, 0.4-0.7 μM for ammonium, 0.1-0.6 μM for phosphate, and 0.3-2.9 μM for silicate (**Table 4.1**). The Redfield ratios suggest potential nitrogen limitation (**Figure 4.1**). The second potential limitation was phosphorus and silica. A first group corresponding to surface offshore stations (NEF512 and NEF43) was characterized by a relatively high N:P ratio (11.9 to 14.2) close to potential silica limitation. In contrast, the second group of NE stations had a low N:P stoichiometry (5.1 to 6.3) and were also limited by silica. Finally, the last group clustering stations from JB with lower N:P ratios (2.8 to 4.8) and a relatively high ratio of Si:N (1.0 to 1.5) (**Figure 4.1**).

Photosynthetically available radiation (PAR) values within the water column were derived from incident PAR data collected at each station. From JB, incident PAR varied between 44 and 61 μmol/m^{-2} sec^{-1} at 2 m where the surface sample was collected, and on average 1% of the surface irradiance was detected at 34 m, which was used to estimate the depth of the euphotic zone. At the NE stations, incident PAR was extremely low because

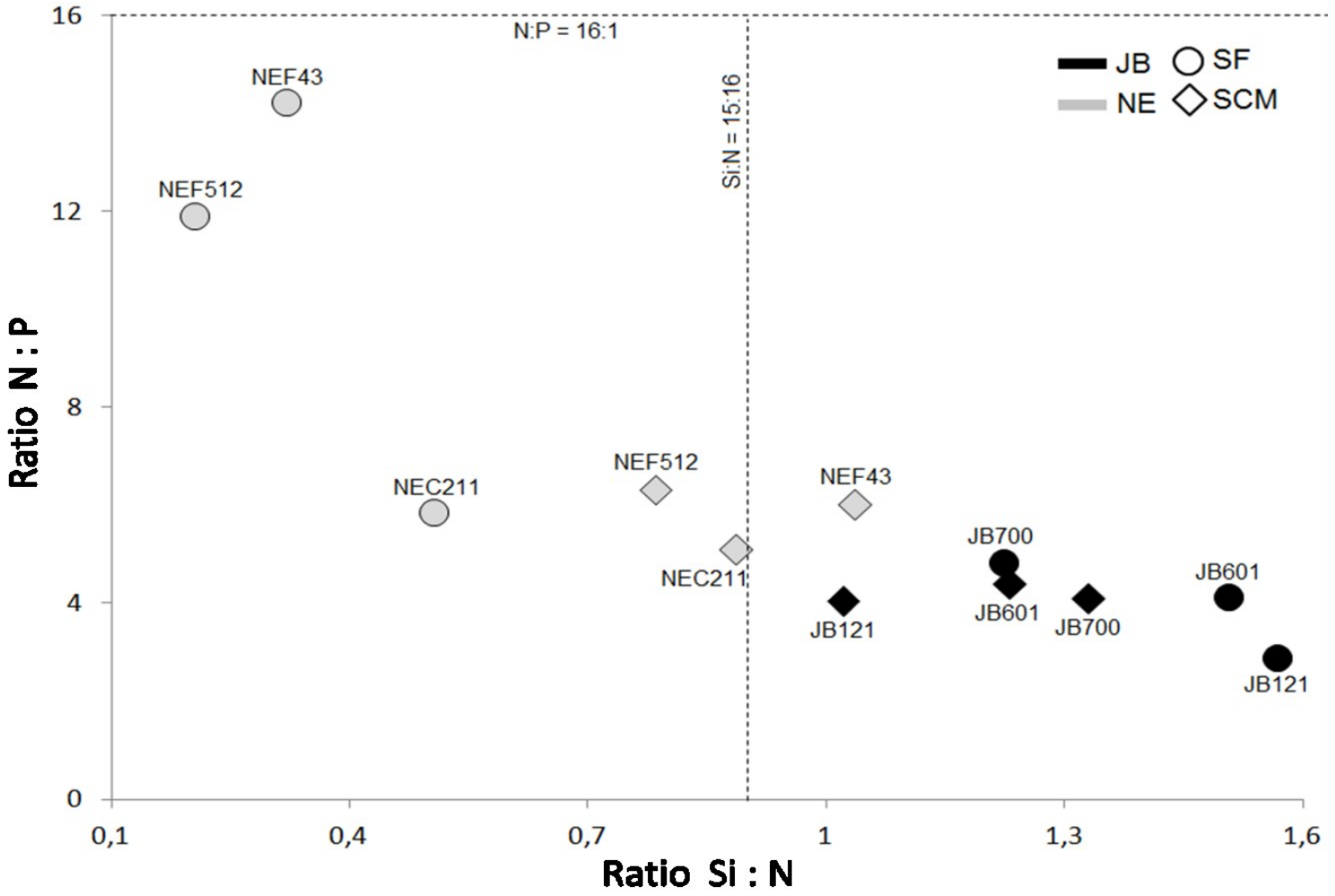


Figure 4.1: Si:N:P molar ratio of samples collected in surface and SCM during GoMA cruise, delimited by the Redfield ratios Si:N:P = 15:16:1.

sampling was done near sunrise (NEC211) or sunset (NEF43) for logistical reasons (<6 μmol/m-2 sec^{-1}). The maximum PAR was taken at the most offshore station NEF512 with ca. 820 μmol/m-2 sec^{-1}. On average, the 1% of the surface irradiance was deeper, around 40 m (**Table 4.1**) at the NE stations.

4.3.2 Chlorophyll *a* concentration and picophytoplanktonic categories

Total chlorophyll *a* (Chl *a*) concentrations differed across the sampling region and on average were greater at the JB stations compared to the NE stations. In the JB, at the stations sampled for clone libraries, total Chl *a* estimated from HPLC ranged from 0.42 to 1.09 μg Chl *a* L^{-1} in surface and 0.01 to 0.93 μg Chl *a* L^{-1} at the SCM (**Table 4.1**). In the NE region, total Chl *a* concentrations were less than half those of JB in the surface with values from 0.14 to 0.23 μg Chl *a* L^{-1} and at the SCM values were from 0.28 to 0.57 μg Chl *a* L^{-1} (**Table 4.1**). The small fraction contribution was lower in JB than NE. From JB surface, on average it accounted for 67% of the total and 73% of the total from NE SCM. In NE, values varied between 72 to 100% and the higher contribution were measured at NEC211 both depths (**Table 4.1**).

The concentration of the three categories of picophytoplankton estimated from FCM varied across stations and depths (**Table 4.2**). Photosynthetic picoeukaryotes (PPE) and *Synechococcus* were present in all samples, whereas *Prochlorococcus* were detected only from the SCM offshore stations (NEC211, NEF43, NEF512) (**Table 4.2**). *Synechococcus* was numerically dominant, except in the surface of the station NEF43, where PPE exceeded *Synechococcus*. *Synechococcus* concentrations ranged from 13 to 117×10^3 cells mL^{-1}. *Prochlorococcus* ranged from 3×10^3 cells mL^{-1} at NEC211 and reached 30×10^3 cells mL^{-1} at NEF43. PPE concentrations were 4 to 43×10^3 cells mL^{-1}, most abundant from JB and with lower values in the SCM offshore (**Table 4.2**). In terms of cell abundance, PPE represented 13 ± 0.2% of the total picophytoplankton in both depths in JB, whereas in samples from the NE stations this proportion was greater in the surface with 48 ± 28% and was lowest in SCM with 6 ± 2% (**Table 4.2**). In terms of carbon biomass (C) PPE accounted for 47 to 70 % of picophytoplankton C at both depths of the JB, whereas for the offshore surface stations PPE C varied between 72 to 97 % of picophytoplankton C, and was lower in the SCM of the NE stations (36 to 55 %, **Table 4.2**).

Table 4.2 : Picophytoplankton from flow cytometry data: photosynthetic picoeukaryotes (PPE), *Synechococcus* (*SYN*), and *Prochlorococcus* (*PRO*); and the contribution of PPE to total picophytoplankton (PPP) in term of cell concentrations and carbon biomass expressed as percentage.

		Relative abundance (× 10^3 cells mL^{-1})			**PPE contribution to total PPP**	
Stations	*Depth (m)*	PPE	*SYN*	*PRO*	Cells abundance (%)	Carbon biomass (%)
JB121	2	21.3	93.6	0	18.5	69.8
	14	12.2	62.7	0	16.3	66.4
JB700	2	6.0	69.6	0	7.9	46.7
	14	12.6	91.4	0	12.1	58.3
JB601	2	12.0	77.9	0	13.3	60.9
	12	8.5	76.4	0	10.0	53.0
NEC211	2	8.2	33.1	0	19.9	71.6
	51	7.5	77.3	3.1	8.5	49.4
NEF43	2	42.8	12.6	0	77.2	97.2
	30	6.8	116.8	29.6	4.4	36.4
NEF512	2	20.1	22.4	0	47.3	90.1
	39	3.8	31.8	18.1	7.1	54.7

Maximal PS II photochemistry efficiency (Fv/Fm) is thought to reflect whole PS II function, and its decrease indicates PS II damage or photo-inhibition under environmental stress (Baker, 2008). At the time of sampling small size fraction concentrations of Chl *a* and Fv:Fm were not significantly correlated using a linear regression model (**Figure 4.2up**). The relationship between number of picophytoplankton cells and Fv:Fm (**Figure 4.2middle**) was significant with an R^2= 0.64 and P value = 0.002. Finally, there was no significant correlation between carbon biomass of picophytoplankton and Fv:Fm (**Figure 4.2down**).

4.3.3 Phylogeny and *psbA* diversity

After screening for quality, we retained 622 eukaryote and 158 picocyanobacteria *psbA* sequences since these were from mRNA they represented putatively active phytoplankton. Although *Synechococcus* dominated the FCM analysis, our clone libraries recovered more *psbA* sequences affiliated with *Prochlorococcus*; 71% of total picocyanobacteria sequences compared to *Synechococcus* with 29%. Moreover, we recovered *Prochlorococcus* in the surface NE stations in the clone libraries, while none were detected by FCM (**Figure 4.5**). The number of *psbA* OTUs in each picophytoplankton group depended on the similarity cutoffs tested (**Annex 4_figure A3**). Overall, at the 1% similarity cutoff, OTU number and diversity indexes were 1.5 to 3 fold higher than at the 2% cutoff, while differences between the 2 % and 3 % cutoff were less (**Annex 4_figure A3**). As the *psbA* transcripts showed a high level of diversity at the 1% difference level, we used 99% sequence similarity to determine OTUs for the remaining phylogenetic analyses.

CCA was performed based on small Chl *a* concentrations and abundance of cells from the three categories of picophytoplankton from FCM (**Figure 4.3**). Overall the environmental variables (temperature, salinity, PAR, and differences in the Redfield ratios explained ca. 20% of variability (Axes 1=15%, Axes 2=5%). Distribution patterns from the CCA showed that *Synechococcus* was associated with small Chl *a* concentrations along the axis influenced by higher Si:N ratios, while *Prochlorococcus* abundance clustered apart with offshore SCM stations associated with higher salinity. Finally, PPE abundance trended with temperature and higher N:P ratios, associated with the offshore surface stations (**Figure 4.3**).

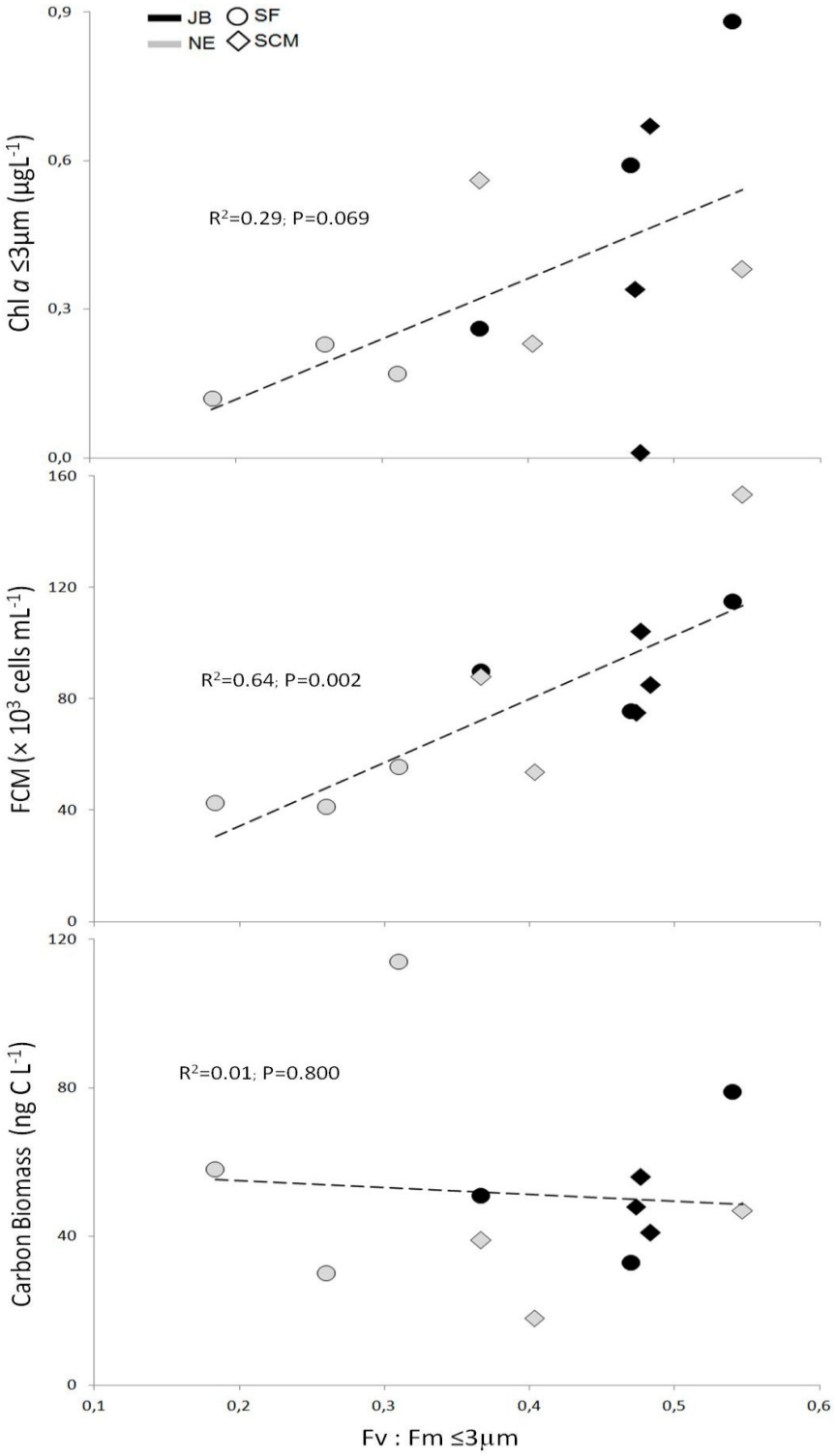


Figure 4.2: Relationship between changes in F_V/F_m ratio measured from small size fraction and biological parameters related to picophytoplankton **a)** Small Chl *a* concentrations, **b)** Abundance of total picophytoplanktonic cells measured using flow cytometry (FCM), and **c)** Carbon biomass of picophytoplanktonic cells(see materials and methods). Legend: stations from Jordan Basin (JB) are in black, stations from North East Channel (NEC) and Fan (NEF) are in grey, surface samples are circles and subsurface chlorophyll maximum (SCM) are diamonds.

4.3.3.1 Picocyanobacteria

The cyanobacteria community phylogenetic structure was visualized using Principal Coordinate Analysis (PCA), as well as cluster analysis using the UniFrac distance metrics (**Figure 4.4**). The PCA indicated that that the P1 axis explained 77.5% of the variance in the data, reflecting substantial differences in the phylogenetic structure of communities from the JB and NE stations. In contrast, only 9.5% of the variance in the data was explained by the P2 axis, which reflected differences between communities from the surface and SCM depths in JB and NE. Finally, samples from JB were relatively close to each other indicating the communities were more similar compared to NE stations with a greater spread (**Figure 4.4**).

The Maximum Likelihood (ML) phylogeny showed that the majority of *psbA* sequences fell into previously identified marine *Prochlorococcus* and *Synechococcus* clades (Scanlan et al., 2009) (**Figure 4.5**). *Synechococcus* were placed into 4 clades of sub-cluster 5.1 (**Figure 4.5a**). Clade I (16 OTUs) were recovered mainly from the surface and accounted for half of the total *Synechococcus* sequences. Other *Synechococcus* OTUs were mainly recovered from the SCM and fell into clade II (2 OTUs), clade III (2 OTUs), clade IV (12 OTUs), and clade IX (2 OTUs). Among *Prochlorococcus* (**Figure 4.5b**), both high-light (HL) and low-light (LL) *Prochlorococcus* clades were recovered falling into 2 previously described ecotypes: HL II and LL I. The majority of *Prochlorococcus* OTUs (42 OTUs) belonged to the HL II ecotype mainly recovered from the SCM and accounting for 84%. The remaining sequences (10 OTUs) clustered with LL I ecotype and were mainly recovered from surface samples.

4.3.3.2 Photosynthetic picoeukaryotes

In contrast to picocyanobacteria, weighted UniFrac cluster analysis indicated no significant relationship between PPE community composition, regions or depths, indicating that picoeukaryotes communities inferred from the *psbA* gene were not distinct from each other (**Figure 4.6**). The PCA, reflecting differences in phylogenetic structure of PPE communities, indicated that only 3 offshore samples (NEF43 surface and NEF512 both

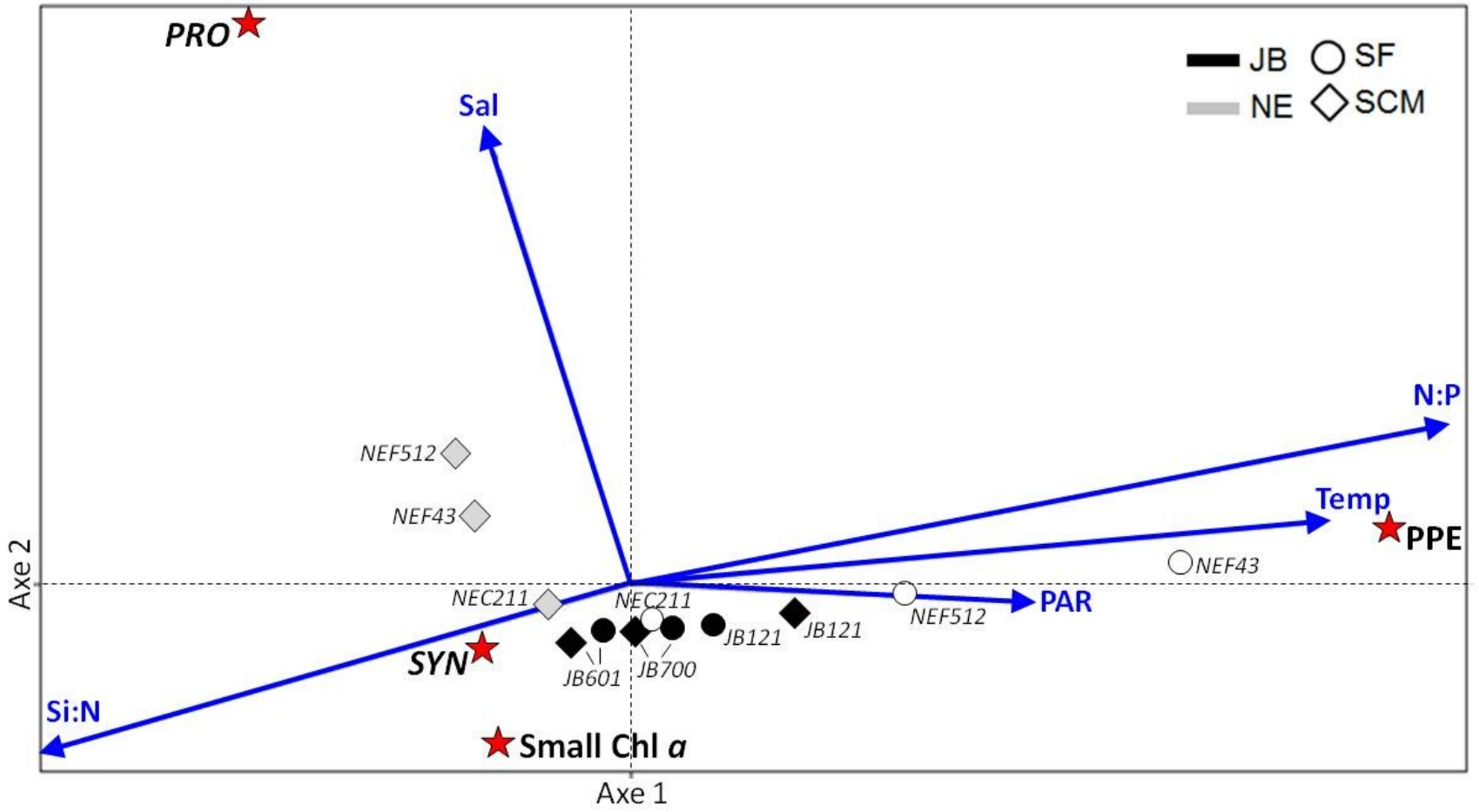


Figure 4.3: Canonical Correspondence Analysis (CCA) plot run on PAST using small Chl *a* concentrations and abundance of picophytoplanktonic cells (*Synechococcus*, *SYN*; *Prochlorococcus*, *PRO*; and picoeukaryotes photosynthetic, PPE) measured by flow cytometry. Variables are salinity (Sal), temperature (Temp, °C), Photosynthetically Available Radiation (PAR, µmol/m^{-2} sec^{-1}), nitrogen:phosphorus ratios (N:P), and silica:nitrogen ratios (Si:N).

depths) clustered well apart from the other samples (**Figure 4.6**) according to the P1 axis which explained 49.5% of the variance in the data.

We distinguished 4 major algal lineages with *psbA:* Haptophyta, stramenopiles (Heterokonta), Chlorophyta, and Cryptophyta (Figure 4.6). We found a statistically significant difference in distributions for the PPE taxonomic groups among samples (P=0.041) where using a multiple comparison procedure we note that only NEF512SCM differed from the others (Turkey test). Overall, Chlorophyta clones were the most numerous, while haptophytes were more diverse with more OTUs. On average, Haptophyta recovered from the surface accounted for 45% of total OTUs and 55% of the SCM OTUs (**Figure 4.7**). Haptophytes clustered within 3 mains orders; Prymnesiales, Phaeocystales, and Isochrysidales (**Figure 4.7; Annex 4_table A1**). Among Chlorophyta, the majority of *psbA* sequences were affiliated with Mamiellophyceae (**Figure 4.8A**). One clade of *Micromonas* corresponded to clade III (Worden et al., 2009) with best matches to *Micromonas* RCC299, *Micromonas* RCC114 and *Micromonas* RCC434. These sequences were mostly from the surface. We also recovered one sequence related to *Mantoniella squamata* RCC417 from the NEC211 surface. Several *psbA* sequences mainly from the SCM were related to *Ostreococcus sp.* RCC344. Finally, still within the Mamiellophyceae, *Bathycoccus* represented the most abundant genus in our libraries without any particular distribution pattern (**Figure 4.8A; Annex 4_table A1**). We noted two other OTUs within the Chlorophyta, one from the offshore SCM, and the other from both depths of the JB stations, that had best matches to other environmental sequences. The majority of Cryptophyta were recovered from the surface (72%). Taxonomic affiliation at either family or genus level was poorly resolved, and we noted only one match with a cultured species; *Teleaulax amphioxeia*. In addition, we recovered *psbA* sequences related to those of the sequestered chloroplast from the ciliate, *Myrionecta rubra* (**Figure 4.8B**; **Annex 4_table A1**). Among stramenopiles, we identified several classes of diatoms, along with Dictyochophyceae, Chrysophyceae, and Pelagophyceae (in order of the most to least common) (**Figure 4.8C**). Dictyochophyceae were almost always recovered in the SCM from JB, Chrysophyceae were reported only from JB stations, single OTUs of Pelagophyceae were found at the JB601 SCM, and diatoms were in the surface and SCM at all stations (**Figure 4.8C, Annex 4_table A1**).

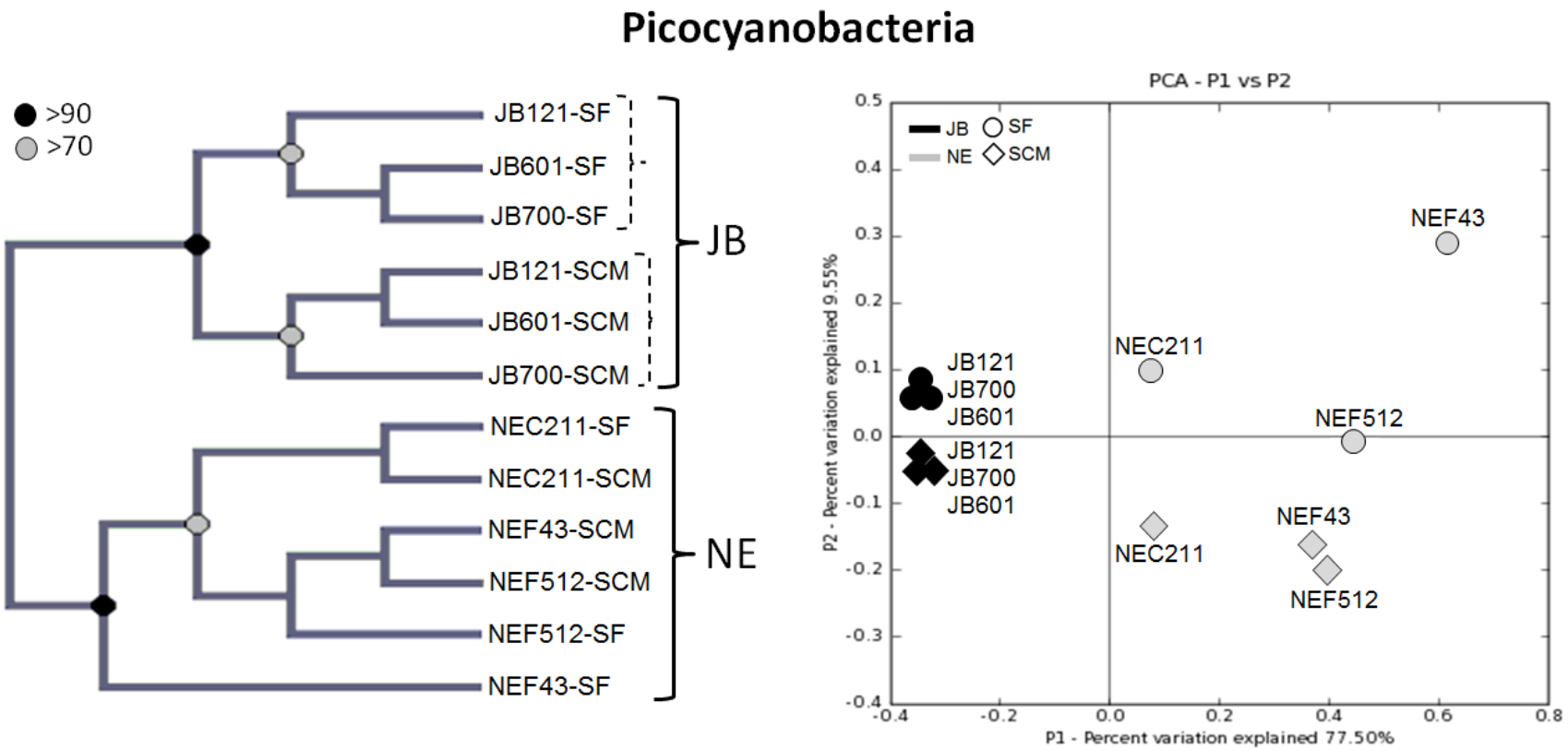


Figure 4.4: UniFrac cluster analysis (left) and Principal Coordinate Analysis (right) using UniFrac weighted distance metric for picocyanobacteria. All *psbA* sequences were normalized across samples. Clustering statistical supports were computed using 100 jackknife replicates and values of ≥50% are shown.

4.4 Discussion

4.4.1 Picophytoplanktonic cells and Fv:Fm

Changes in efficiency of conversion of light energy into chemical energy (photosynthesis) measured by Fv:Fm, indicate some effect picophytoplanktonic cells, either natural (degeneration) or from stress (high temperature or salinity, photoinhibition). JB picophytoplanktonic cells had higher Fv:Fm ratios compared to NE, and we infer seemed to be less stressed (Figure 4.2). Moreover, results showed a positive and significant correlation between number of cells and Fv:Fm values. *Synechococcus* concentrations were lower in the surface samples from the two most offshore stations with higher temperature and salinity, compared to other samples (**Table 4.1; Table 4.2**). In contrast, PPE concentrations were higher at theses stations, suggesting that PPE communities were better able to tolerate the local conditions. Thus, tracking *psbA* transcripts in environmental samples over time could provide information on the species whose PS II repair cycle is most active and need to rapidly adapt to local stressful conditions.

4.4.2 Phylogeny and distribution of picocyanobacteria

FCM analyses showed summer, GoMA waters were dominated by *Synechococcus* in the photic zone. *Synechococcus* are common in many temperate coastal regions and their annual cycle of abundance in the GoMA is well established (Li and Dickie, 2001; Li et al., 2006; Poulton et al., 2009). In contrast, *Prochlorococcus* were only recovered in warm higher salinity waters from SCM NE stations outside the GoMA proper and *Prochlorococcus* spp. have not been reported in the GoMA (Li et al., 2011), likely because the genus requires warmer temperatures (Partensky et al., 1999a; Rocap et al., 2003; Bouman et al., 2006; Malmstrom et al., 2010). The geographical distributions of the two picocyanobacteria genera indicated constraints related to light, temperature, nutrients, and chlorophyll *a* concentrations (Johnson et al., 2006; Flombaum et al., 2013). Our phylogenetic trees of *psbA* also indicated distinct geographical distributions. *Prochlorococcus* is composed of multiple genetically and physiologically differentiated subgroups, including at least 6 closely related subgroups that are associated with different

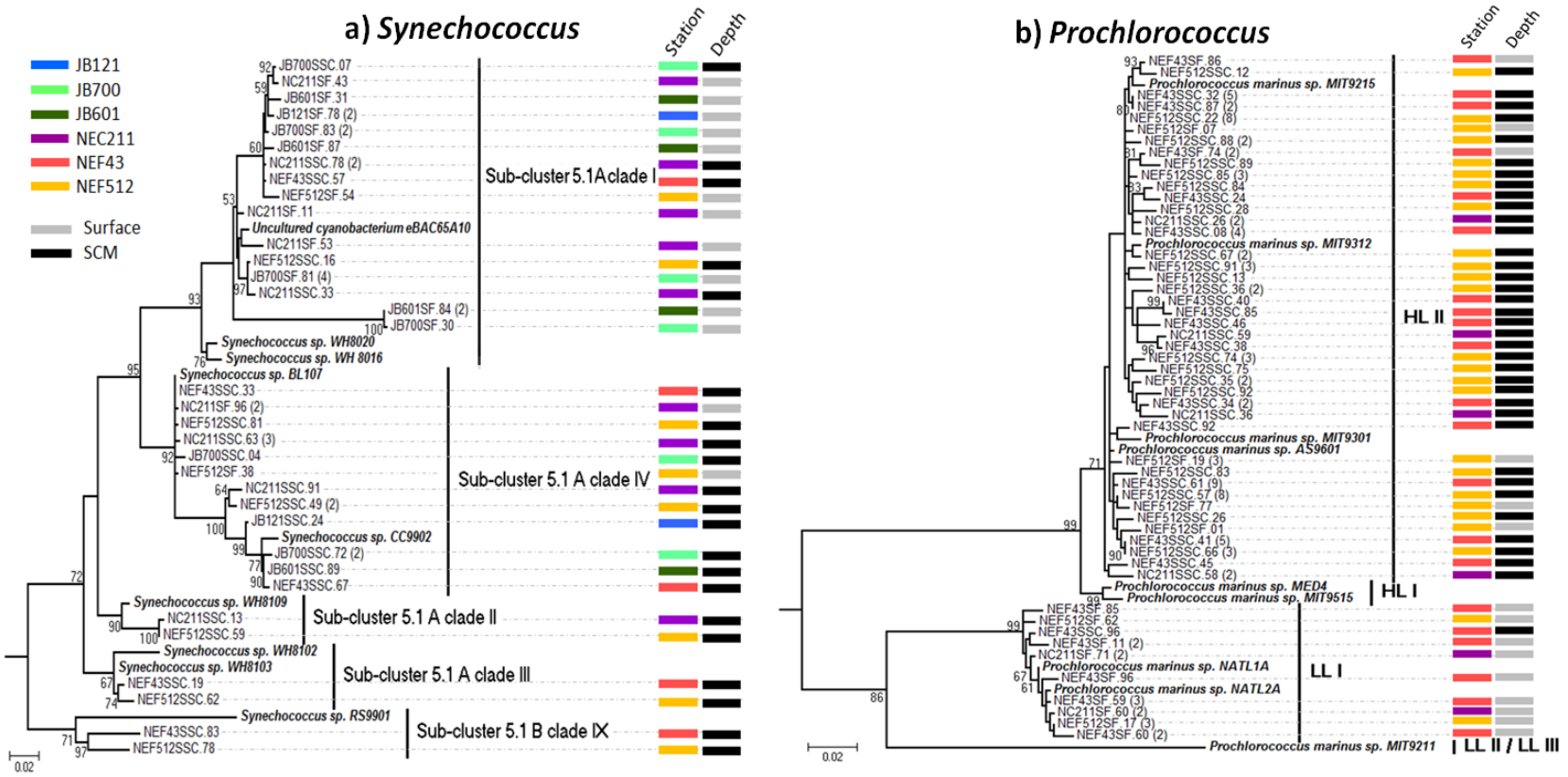


Figure 4.5: ML phylogenetic trees of cyanobacteria based on *psbA* OTUs (≥99% similarity) from the Gulf of Maine; **a)** *Synechococcus,* and **b)** *Prochlorococcus*. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site, and rooted with *Synechococcus elongatus* and *Synechocystis* X13547 not shown. The number of *psbA* sequences for each OTU is given in parenthesis. Bootstrap values >50% from 100 replicates are shown at the nodes.

depths and thought to be adapted to different light regimes (Rocap et al., 2002). Generally, high-light adapted (HL) *Prochlorococcus* dominate in the surface mixed layer and can be further subdivided into HL I and HL II subgroups; while the low-light adapted (LL) *Prochlorococcus* dominate in deeper water of the euphotic zone and comprise at least 4 ecotypes LL I, LL II, LL III and LL IV (Rocap et al., 2002). In our study, HL II *Prochlorococcus* was the predominate group recovered from SCM between 30 m and 50 m, reflecting possible mixing or upwelling. A previous study showed that HL II dominates subtropical and tropical regions while paradoxically HL I (not recovered here) is more abundant at temperate latitudes (Zwirglmaier et al., 2008). Next, using *psbA* clone libraries we cloned transcripts of *Prochlorococcus* from surface (2 m) NE samples. *Prochlorococcus* detected was not detected using FCM demonstrating that our *psbA* gene clone libraries were able to retrieve rare phylotypes from environmental samples. This possibly reflected some primer bias, where *Prochlorococcus* out competed other *psbA* sequences during the PCR reaction or a bias in the conversion of RNA to cDNA. These *psbA* sequences were related to LL I subgroups with the closest match to *Prochlorococcus marinus* sp. NATL1A and NATL2A that tolerate exposure to higher light levels (Coleman and Chisholm, 2007; Zinser et al., 2007; Malstrom et al., 2010). Among *Synechococcus*, at least 10 distinct lineages were initially reported (Rocap et al., 2002; Fuller et al., 2003, Muhling et al., 2005; Lindell et al., 2004; Scalan et al., 2009). Most of our sequences belonged to *Synechococcus* subcluster 5.1, which contains several clades, and is the dominant group within the euphotic zone of both open-ocean and coastal waters (Olson et al., 1990). The most common of these in our samples was within subcluster 5.1, clade IV (Rocap et al., 2002), which dominate in temperate, coastal environments (Zwirglmaier et al., 2008). Other clades recovered (II, III, IX) were less frequent.

4.4.3 Diversity of eukaryotes inferred by *psbA*

Our libraries using RNA as a template recovered eukaryote *psbA* sequences from four taxonomic groups: Haptophyta, Chlorophyta, Cryptophyta and stramenopiles, suggesting that PS II was active in these phyla. The clone library results indicated that eukaryotic *psbA* sequences were phylogenetically diverse, with many sequences distant from cultured and described representative species, which likely reflects limited number of

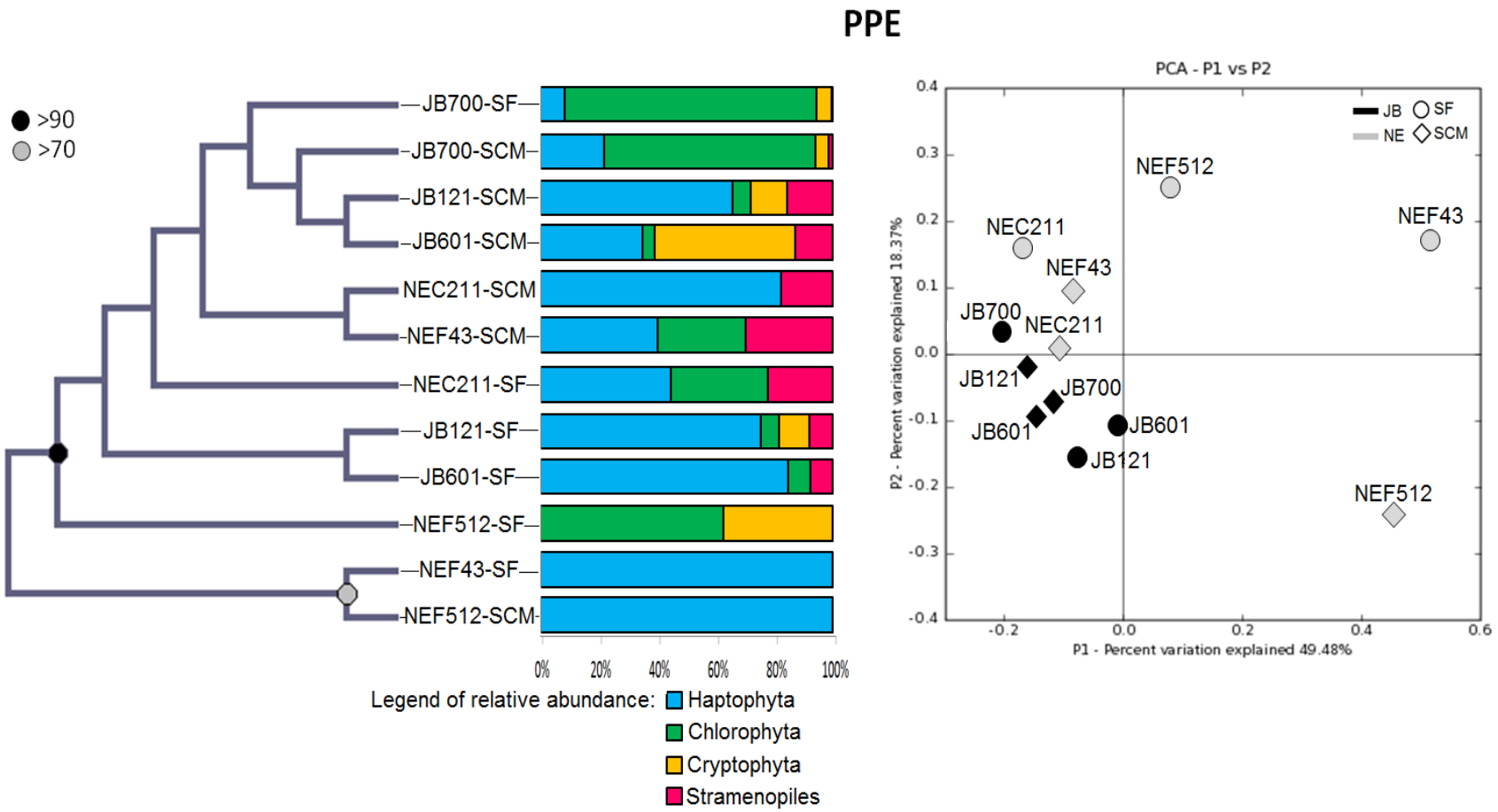


Figure 4.6: Unifrac cluster analysis (left) and Principal Coordinate Analysis (right) using UniFrac weighted distance metric for photosynthetic picoeukaryotes (PPE). All *psbA* sequences were normalized across samples. Clustering statistical supports were computed using 100 jackknife replicates and values of ≥50% are shown.

species with *psbA* sequences in the NCBI database. The ML phylogenies indicated that diversity occurred over both regions and depths (**Figure 4.7; Figure 4.8**) and that the haptophytes were the most diverse among these four phyla. Similarly haptophyte diversity has been highlighted by recent studies targeting chloroplast SSU rDNA in pelagic and coastal environments (Fuller et al., 2006; McDonald et al., 2007; Liu et al., 2009). Haptophytes are widespread in Oceanic waters including for example other regions of the Atlantic Ocean, the Indian Ocean and Mediterranean Sea (Not et al., 2008; Man-Aharonovich et al., 2010; Kirkham et al., 2011). These phytoplankton are major contributors to marine CO2 fixation (Cuvellier et al., 2010; Jardillier et al., 2010), and in spite of their ecological and phylogenetic importance the diversity of haptophytes is still likely underestimated in culture collections (Liu et al., 2009; Simon et al., 2013; Unrein et al., 2013). While many of these uncultivated groups are reported to be true picophytoplankton, in our case the size fractionation process was likely not perfect with a number of the sequences belonging to groups normally in the nanoplankton size range (2-20 microns) recovered. For example, we identified presence of calcifying coccolithophores, including from Isochrysidales, (De Vargas et al., 2007; Simon et al., 2013). In addition, there were representatives of the Prymnesiales with many known species considered nanoplankton, these were almost exclusively from coastal and shelf environments. However, several other haptophyte clusters could not be identified because of a lack of representative cultured species with sequenced *psbA* genes available. In contrast to Haptophyta, the majority of potentially active Chlorophyta could be identified at least to genus. The Mamiellophyceae dominated our *psbA* libraries with four genera recovered: *Micromonas*, *Bathycoccus*, *Mantoniella*, and *Ostreococcus*. Mamiellophyceae are frequently abundant in coastal and nutrient-rich areas (Not et al., 2005; Not et al, 2008). Our *Ostreococcus psbA* sequences matched those from *Ostreococcus* RCC344 identified as high-light adapted (Rodriguez et al., 2005; Worden, 2006). However, we recovered these from both the surface and SCM. Cryptophyta seemed to be better adapted to surface conditions. In addition to several free living cryptophytes, we recovered *psbA* sequences that were affiliated to the ciliate *Myrionecta rubra,* which takes up and uses cryptophyte chloroplasts for photosynthesis. Park et al. (2008) reported that the *psbA* sequences from *M. rubra* strain MR-MAL01 (GenBank accession number EF195734); and a cryptophyte strain

Figure 4.7: Phylogenetic tree (ML) of Haptophyta with data from this study are in blue bold (201 *psbA* sequences, OTUs at 99%), cultures sequences are in black bold, and uncultured *psbA* sequences from best BLAST are in black regular (**Annex 4_table A1**). Numbers of clones for each OTU are indicated in bracket. Bootstrap analysis of 100 replicates was conducted, and only values > 50 % are indicated.

CR-MAL01 (GenBank accession number EF195735) were essentially identical and some of out *psbA* transcripts could have been from the cryptophyte, however we note that *M. rubra* is reported from this region from both microscopy and 18S rRNA genes (Chapter 3, this thesis).

To conclude, phylogenetic diversity of picocyanobacteria and their distribution was consistent with different ecotypes occurring dependent on local conditions, while eukaryotes, despite lower abundance, were widely dispersed with little geographic or depth signal from individual phylotype detected from clone libraries. Their PS II seemed to be less affected by environmental conditions and the use of *psbA* transcripts provided a new perspective on metabolically active phytoplankton.

4.5 Acknowledgements

The cruise was in conjunction with the Discovery Corridor Program of the Canadian Department of Fisheries and Oceans, the Centre for Marine Biodiversity (CMB) and the Gulf of Maine Census of Marine Life. We thank the Canadian Coast Guard officers and crew of the *CCGS* Hudson for their assistance and support as well as the chief scientist Dr Peter Lawton. We thank Emmanuelle Medrinal for field assistance, Jeff Anning for nutrients and FCM data, Bob Ryan for CTD data, and Marie Lionard for comments and analytical advice. This work is a contribution to the Canadian Healthy Oceans Network (CHONe) funded by the Natural Science and Engineering Research Council of Canada (NSERC). Addition funds were provided by Fonds de recherche du Québec (FQRNT) to Québec Océan and an NSERC Discovery grant to CL.

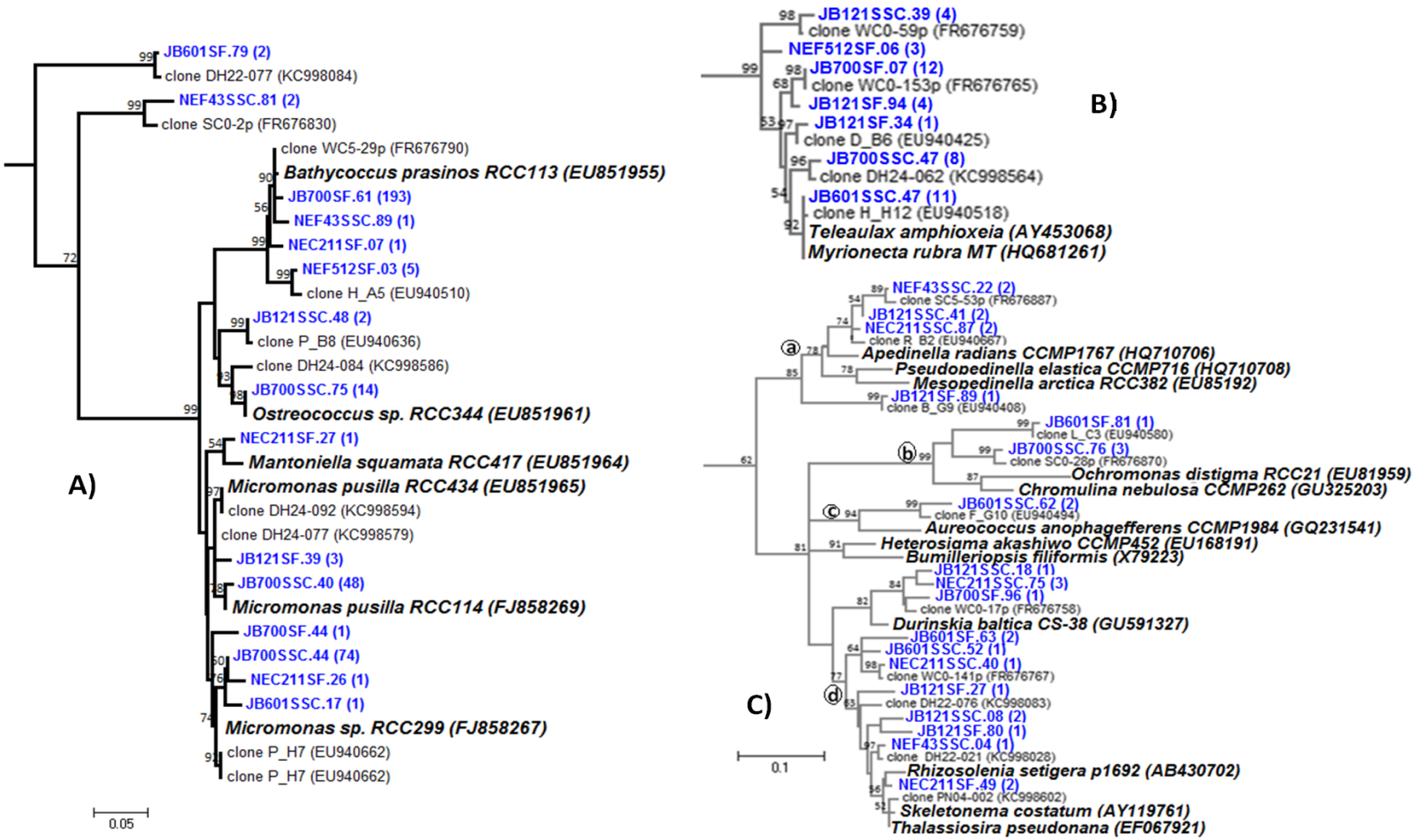
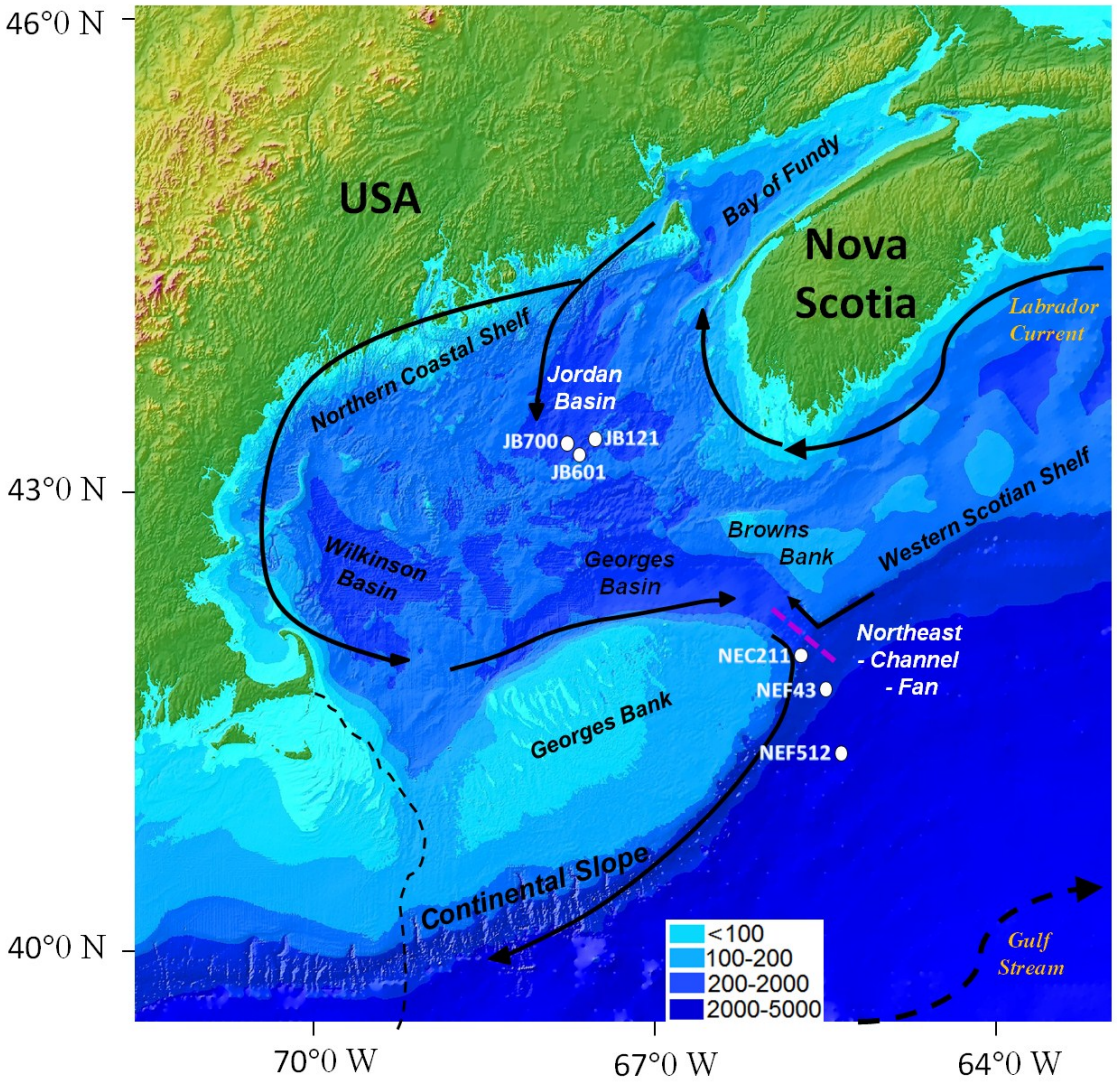


Figure 4.8: Phylogenetic tree (ML) of OTUs at 99% **A)** Chlorophyta (349 *psbA* sequences), **B)** Cryptophyta (43 *psbA* sequences), and **C)** Stramenopiles with classes annoted like as "a" is Dictyochophyceae, "b" is Chrysophyceae, "c" is Pelagophyceae, "d" is diatoms) (29 *psbA* sequences). Data from this study are in blue bold, cultures sequences are in black bold and uncultured *psbA* sequences are in black regular from best BLAST (**Annex 4_table A1**). Numbers of clones for each OTU are indicated in bracket. Bootstrap analysis of 100 replicates was conducted, and only values > 50 % are indicated. .

Annex 4

Annex 4_figure A1: Map of sampling sites from Jordan Basin (JB), Northeast Channel (NEC) and Northeast Fan (NEF) regions in the Gulf of Maine. Arrows represent the major surface currents. Map is modified from image created by Gulf of Maine Area Program, Census of Marine Life 15 arc sec**.** *http://pubs.usgs.gov/of/1998/of98-801/bathy/data.htm*.



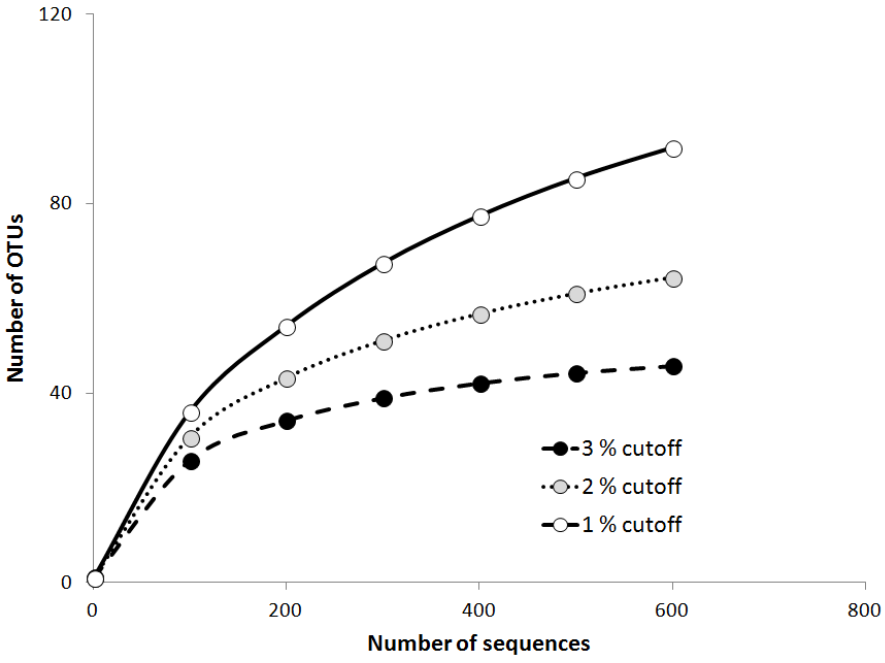
Annex 4_figure A2: Primers information on *psbA* (Wang and Chen, 2008) indicating amplification site. The example is from the nucleotide sequence of *Synechococcus* sp. WH 802

Target Gene	Primer Name	Primer Sequence	Corresponding Amino Acid sequence	PCR Amplicon Size (bp)	Annealing Temperature (°C)
psbA	psbA-93F	5'-TAYCCNATYTGGGAAGC-3'	YPIWEA	745	55
	psbA-341R	5'-TCRAGDGGGAARTTRTG-3'	HNFPLD		

```
1   ATGWSNACNGCNATHMGNWSNGGNMGNCARWSNAAYTGGGARGCNTTYTGYCARTGGGTNYTNGCNGCNA
71  CNATHTGYTTYACNATHGCNTTYATHGCNGCNCCNCCNGTNGAYATHGAYGGNATHMGNGARCCNGTNGC
141 NGGNWSNYTNATHTAYGGNAAYAAYATHATHWSNGGNGCNGTNGTNCCNWSNWSNAAYGCNATHGGNYTN
211 CAYTTYTAYCCNATHTGGGARGCNGCNWSNYTNGAYGARTGGYTNTAYAAYGGNGGNCCNTAYCARYTNG
281 TNTGYTTYCAYTTYYTNATHGGNATHWSNGCNTAYATGGGNMGNCARTGGGARYTNWSNTAYMGNYTNGG
351 NATGMGNCCNTGGATHTGYGTNGCNTAYWSNGCNCCNYTNWSNGCNGCNATGGCNGTNTTYYTNGTNTAY
421 CCNTTYGGNCARGGNWSNTTYWSNGAYGGNATGCCNYTNGGNATHWSNGGNACNTTYAAYTTYATGYTNG
491 TNTTYCARGCNGARCAYAAYATHYTNATGCAYCCNTTYCAYATGYTNGGNGTNGCNGGNGTNTTYGGNGG
561 NWSNYTNTTYWSNGCNATGCAYGGNWSNYTNGTNACNWSNWSNYTNGTNMGNGARACNACNGARACNGAR
631 WSNCARAAYTAYGGNTAYAARTTYGGNCARGARGARGARACNTAYAAYATHGTNGCNGCNCAYGGNTAYT
701 TYGGNMGNYTNATHTTYCARTAYGCNWSNTTYAAYAAYWSNMGNWSNYTNCAYTTYTTYYTNGGNGCNTG
771 GCCNGTNGTNGGNATHTGGTTYACNWSNATGGGNGTNWSNACNATGGCNTTYAAYYTNAAYGGNTTYAAY
841 TTYAAYCARWSNATHYTNGAYGGNCARGGNMGNGTNGTNAAYACNTGGGCNGAYATGGTNAAYMGNGCNG
911 GNYTNGGNATGGARGTNATGCAYGARMGNAAYGCNCAYAAYTTYCCNYTNGAYYTNGCNACNGTNGARWS
981 NACNCCNGTNGCNYTNCARGCNCCNGCNATHGGN
```

Annex 4_figure A3: Table of diversity and predicted richness of the *psbA* gene from picophytoplankton. The Shannon diversity index and Chao1 richness estimator computed using *Mothur* for 1%, 2% and 3 % differences in nucleic acid sequence alignments; and comparison of rarefaction curves of *psbA* transcripts of PPE at different cutoff.

	1 % cutoff			2 % cutoff			3 % cutoff		
	OTUs	Shannon	Chao1	OTUs	Shannon	Chao1	OTUs	Shannon	Chao1
PPE	93	3.16	121	65	2.95	78	46	2.76	51
Synechococcus	16	2.41	25	11	2.07	12	10	2.01	10
Prochlorococcus	45	3.48	62	16	2.16	19	9	1.35	17



Annex 4_table A1: Table of highest BLAST matches, and percent similarity (%) to the non-redundant sequences for 92 OTUs at 99% similarity from *psbA* sequence for eukaryotes

Seq. Rep. OTU (Access. nb)	Nb of clones at 99%	Taxonomic groups	Highest match (access. nb)	Similarity (%)	Highest cultured match (accession number)	Similarity (%)
JB700SF.61	193	Chlorophyta	Uncultured marine eukaryote clone WC5-29p(FR676790)	99	Bathycoccus prasinos RCC113 (EU851955)	99
NEC211SF.07	1	Chlorophyta	Uncultured marine eukaryote clone WC5-29p (FR676790)	97	Bathycoccus prasinos RCC113 (EU851955)	97
NEF43SSC.89	1	Chlorophyta	Uncultured marine eukaryote clone WC5-29p(FR676790)	99	Bathycoccus prasinos RCC113 (EU851955)	99
NEF512SF.03	5	Chlorophyta	Uncultured eukaryote clone H_A5(EU940510)	99	Bathycoccus prasinos RCC113 (EU851955)	97
JB700SSC.40	48	Chlorophyta	Uncultured organism clone DH24-077(KC998579)	98	Micromonas pusilla RCC114 (FJ858269)	99
JB121SF.39	3	Chlorophyta	Uncultured organism clone DH24-092 (KC998594)	99	Micromonas pusilla RCC434 (EU851965)	99
NEF43SSC.81	2	Chlorophyta	Uncultured marine eukaryote clone SC0-2p(FR676830)	97	Micromonas pusilla RCC434(EU851965)	89
JB121SF.68	10	Chlorophyta	Uncultured eukaryote clone P_H7(EU940662)	99	Micromonas sp. RCC299 (FJ858267)	98
JB700SF.44	1	Chlorophyta	Uncultured eukaryote clone J_G4 (EU940558)	98	Micromonas sp. RCC299 (FJ858267)	98
JB700SSC.44	74	Chlorophyta	Uncultured eukaryote clone P_H7(EU940662)	99	Micromonas sp. RCC299 (FJ858267)	99
JB121SSC.48	2	Chlorophyta	Uncultured eukaryote clone P_B8 (EU940636)	99	Micromonas sp. RCC299 (FJ858267)	96
JB601SSC.17	1	Chlorophyta	Uncultured eukaryote clone J_G4 (EU940558)	99	Micromonas sp. RCC299 (FJ858267)	99
NEC211SF.26	1	Chlorophyta	Uncultured eukaryote clone P_H7 (EU940662)	98	Micromonas sp. RCC299 (FJ858267)	98
NEC211SF.27	1	Chlorophyta	Uncultured eukaryote clone J_G4 (EU940558)	98	Micromonas sp. RCC299 (FJ858267)	98
JB700SSC.75	14	Chlorophyta	Uncultured organism clone DH24-084(KC998586)	98	Ostreococcus sp. RCC344 (EU851961)	100
NEF512SF.06	3	Chlorophyta	Uncultured eukaryote clone R_H2 (EU940688)	98	Pedinomonas minor LB1350 (FJ968740)	89
JB601SF.79	2	Chlorophyta	Uncultured organism clone DH22-077(KC998084)	99	Pyramimonas parkeae (FJ493499)	93
JB700SSC.76	3	Chrysophyte	Uncultured marine eukaryote clone SC0-28p(FR676870)	98	Chromulina nebulosa CCMP262(GU325203)	92
JB601SF.81	1	Chrysophyte	Uncultured eukaryote clone L_C3 (EU940580)	99	Ochromonas distigma RCC21 (EU851959)	90
JB121SF.94	4	Ciliophora	Uncultured eukaryote clone D_B6 (EU940425)	99	Myrionecta rubra MR-MAL01 (EU123325)	98
JB601SSC.47	11	Ciliophora	Uncultured eukaryote clone H_H12 (EU940518)	99	Myrionecta rubra MR-MAL01 (EU123325)	99
JB700SF.07	12	Cryptophyta	Uncultured marine eukaryote clone WC0-153p(FR676765)	100	Teleaulax amphioxeia (AB471796)	98
JB700SSC.47	8	Cryptophyta	Uncultured organism clone DH24-062 (KC998564)	99	Teleaulax amphioxeia (AB471796)	98
JB121SF.34	1	Cryptophyta	Uncultured eukaryote clone D_B6(EU940425)	99	Teleaulax amphioxeia (AB471796)	98
JB121SSC.39	4	Cryptophyta	Uncultured marine eukaryote clone WC0-59p (FR676759)	99	Teleaulax amphioxeia (AB471796)	96
JB601SSC.52	1	Diatoms	Uncultured marine eukaryote clone WC0-141p (FR676767)	98	Odontella sinensis (Z67753)	97
NEC211SSC.40	1	Diatoms	Uncultured marine eukaryote clone WC0-141p (FR676767)	99	Odontella sinensis (Z67753)	96
NEF43SSC.04	1	Diatoms	Uncultured organism clone PN04-002 (KC998602)	98	Rhizosolenia setigera p1692 (AB430702)	99
JB121SF.80	1	Diatoms	Uncultured organism clone PN04-002 (KC998602)	97	Skeletonema costatum (AY119761)	97
JB121SF.87	2	Diatoms	Uncultured organism clone PN04-002(KC998602)	98	Thalassiosira oceanica CCMP1005(GU323224)	99
JB121SF.27	1	Diatoms	Uncultured organism clone PN04-002 (KC998602)	97	Thalassiosira pseudonana(EF067921)	96
JB121SSC.08	2	Diatoms	Uncultured organism clone DH22-076(KC998083)	97	Thalassiosira pseudonana(EF067921)	97
JB601SF.63	2	Diatoms	Uncultured organism clone PN04-002 (KC998602)	96	Thalassiosira pseudonana(EF067921)	96
NEC211SF.49	2	Diatoms	Uncultured organism clone PN04-002 (KC998602)	99	Thalassiosira pseudonana(EF067921)	99
NEC211SSC.87	2	Dictyochophyte	Uncultured eukaryote clone R_B2(EU940667)	99	Apedinella radians CCMP:1767(HQ710706)	96
NEF43SSC.22	2	Dictyochophyte	Uncultured marine eukaryote clone SC5-53p(FR676887)	99	Apedinella radians CCMP:1767(HQ710706)	94
JB121SF.89	1	Dictyochophyte	Uncultured eukaryote clone B_G9 (EU940408)	99	Mesopedinella arctica RCC382 (EU851962)	91
JB121SSC.41	2	Dictyochophyte	Uncultured eukaryote clone R_B2 (EU940667)	99	Pseudopedinella elastica CCMP:716(HQ710708)	95
JB700SF.96	1	Dinoflagellates	Uncultured marine eukaryote clone WC0-17p (FR676758)	97	Durinskia baltica CS-38 (GU591327)	95
JB121SSC.18	1	Dinoflagellates	Uncultured marine eukaryote clone WC0-17p (FR676758)	98	Kryptoperidinium foliaceum CCMP1326 (GU591328)	96

NEC211SSC.75	3	Dinoflagellates	Uncultured marine eukaryote clone WC0-17p(FR676758)	97	Kryptoperidinium foliaceum CCMP1326(GU591328)	96
JB700SSC.45	12	Haptophyta	Uncultured eukaryote clone L_H8 (EU940613)	99	Chrysochromulina polylepis (AJ575572)	99
JB121SSC.02	1	Haptophyta	Uncultured eukaryote clone D_F3 (EU940441)	99	Chrysochromulina polylepis (AJ575572)	97
JB601SF.04	4	Haptophyta	Uncultured eukaryote clone R_F5(EU940683)	98	Chrysochromulina polylepis (AJ575572)	97
JB601SSC.77	4	Haptophyta	Uncultured eukaryote clone L_E11 (EU940591)	99	Chrysochromulina polylepis (AJ575572)	99
NEC211SF.91	1	Haptophyta	Uncultured eukaryote clone J_G12(EU940557)	98	Chrysochromulina polylepis (AJ575572)	97
NEF43SSC.50	3	Haptophyta	Uncultured eukaryote clone L_D7 (EU940589)	98	Chrysochromulina polylepis (AJ575572)	97
JB121SSC.76	1	Haptophyta	Uncultured eukaryote clone B_G10 (EU940395)	98	Chrysochromulina polylepis(AJ575572)	96
JB601SSC.88	4	Haptophyta	Uncultured marine eukaryote clone SC0-7p (FR676834)	98	Chrysochromulina polylepis(AJ575572)	98
NEC211SF.08	1	Haptophyta	Uncultured eukaryote clone B_B11 (EU940378)	99	Chrysochromulina polylepis(AJ575572)	97
NEC211SSC.17	18	Haptophyta	Uncultured organism clone DH22-057(KC998064)	99	Chrysochromulina polylepis(AJ575572)	96
JB121SSC.51	1	Haptophyta	Uncultured eukaryote clone L_F8 (EU940601)	99	Emiliania huxleyi CCMP:1516 (JN022705)	96
JB700SF.24	4	Haptophyta	Uncultured eukaryote clone L_B11(EU940573)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
JB700SF.66	2	Haptophyta	Uncultured alga clone HOT2-10 (AY176608)	98	Emiliania huxleyi CCMP:1516 (JN022705)	96
JB700SSC.82	2	Haptophyta	Uncultured eukaryote clone B_A10(EU940375)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
JB700SSC.87	7	Haptophyta	Uncultured eukaryote clone J_F5 (EU940554)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
JB121SF.14	3	Haptophyta	Uncultured eukaryote clone D_F2(EU940440)	99	Emiliania huxleyi CCMP:1516 (JN022705)	96
JB121SSC.29	11	Haptophyta	Uncultured organism clone DH24-075(KC998577)	97	Emiliania huxleyi CCMP:1516 (JN022705)	96
JB601SF.03	2	Haptophyta	Uncultured eukaryote clone P_A6 (EU940632)	98	Emiliania huxleyi CCMP:1516 (JN022705)	97
JB601SF.23	1	Haptophyta	Uncultured marine eukaryote clone SD5-21p (FR676928)	99	Emiliania huxleyi CCMP:1516 (JN022705)	99
JB601SF.44	2	Haptophyta	Uncultured eukaryote clone L_D2 (EU940587)	98	Emiliania huxleyi CCMP:1516 (JN022705)	97
JB601SF.45	1	Haptophyta	Uncultured eukaryote clone F_E5 (EU940483)	97	Emiliania huxleyi CCMP:1516 (JN022705)	96
JB601SF.50	5	Haptophyta	Uncultured eukaryote clone R_F5 (EU940683)	99	Emiliania huxleyi CCMP:1516 (JN022705)	98
JB601SSC.06	2	Haptophyta	Uncultured organism clone DH24-075(KC998577)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
JB601SSC.34	3	Haptophyta	Uncultured eukaryote clone L_F8 (EU940601)	97	Emiliania huxleyi CCMP:1516 (JN022705)	95
JB601SSC.46	2	Haptophyta	Uncultured eukaryote clone J_B7(EU940531)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
NEC211SF.75	2	Haptophyta	Uncultured eukaryote clone R_F4 (EU940682)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
NEC211SSC.05	1	Haptophyta	Uncultured eukaryote clone L_E3 (EU940592)	98	Emiliania huxleyi CCMP:1516 (JN022705)	97
NEF43SSC.78	1	Haptophyta	Uncultured eukaryote clone D_D5 (EU940432)	99	Emiliania huxleyi CCMP:1516 (JN022705)	97
NEF512SSC.50	1	Haptophyta	Uncultured eukaryote clone J_E8 (EU940549)	98	Emiliania huxleyi CCMP:1516 (JN022705)	95
NEC211SSC.66	1	Haptophyta	Uncultured eukaryote clone L_E3 (EU940592)	97	Emiliania huxleyi CCMP:1516 (JN022705)	96
JB700SSC.54	3	Haptophyta	Uncultured marine eukaryote clone SD5-21p (FR676928)	97	Isochrysis sp. SAG 927-2 (AY119753)	97
JB700SF.39	8	Haptophyta	Uncultured organism clone DH22-086(KC998093)	99	Phaeocystis antarctica CCMP1374 (JN117275)	99
JB601SF.26	4	Haptophyta	Uncultured eukaryote clone B_B12(EU940379)	98	Phaeocystis globosa Pg-G(A) (KC900889)	97
JB700SSC.17	1	Haptophyta	Uncultured organism clone PN09-075 (KC998378)	97	Phaeocystis globosa Pg-G(A)(KC900889)	96
JB700SF.40	2	Haptophyta	Uncultured organism clone DH24-075(KC998577)	97	Phaeocystis globosa Pg-G(A)(KC900889)	97
JB700SF.85	3	Haptophyta	Uncultured eukaryote clone J_B8(EU940532)	99	Phaeocystis globosa Pg-G(A)(KC900889)	95
JB700SSC.08	1	Haptophyta	Uncultured organism clone DH24-028(KC998530)	98	Phaeocystis globosa Pg-G(A)(KC900889)	97
JB700SSC.74	9	Haptophyta	Uncultured eukaryote clone B_B12 (EU940379)	99	Phaeocystis globosa Pg-G(A)(KC900889)	98
JB121SF.06	21	Haptophyta	Uncultured marine eukaryote clone SC0-14p (FR676862)	99	Phaeocystis globosa Pg-G(A)(KC900889)	97
JB121SF.15	1	Haptophyta	Uncultured eukaryote clone J_B8 (EU940532)	98	Phaeocystis globosa Pg-G(A)(KC900889)	94
JB121SF.49	2	Haptophyta	Uncultured eukaryote clone J_H5(EU940563)	98	Phaeocystis globosa Pg-G(A)(KC900889)	97
JB121SSC.65	2	Haptophyta	Uncultured eukaryote clone J_F4 (EU940553)	98	Phaeocystis globosa Pg-G(A)(KC900889)	94

NEC211SSC.25	1	Haptophyta	Uncultured eukaryote clone J_F4 (EU940553)	98	Phaeocystis globosa Pg-G(A)(KC900889)	94
NEC211SSC.35	1	Haptophyta	Uncultured marine eukaryote clone WD5-85p (FR676823)	99	Phaeocystis globosa Pg-G(A)(KC900889)	96
NEF43SF.52	3	Haptophyta	Uncultured organism clone DH24-075(KC998577)	97	Pleurochrysis carterae (AY119757)	96
JB601SSC.62	2	Pelagophyte	Uncultured eukaryote clone F_G10 (EU940494)	97	Aureococcus anophageff. CCMP1984 (GQ231541)	92
JB700SSC.38	9	Haptophyta	Uncultured marine eukaryote clone WC5-20p (FR676789)	100	----	
NEC211SSC.03	1	Haptophyta	Uncultured eukaryote clone R_F4(EU940682)	98	----	
JB121SSC.64	5	Haptophyta	Uncultured eukaryote clone J_E9(EU940550)	98	----	
NEC211SSC.64	6	Haptophyta	Uncultured alga clone HOT2-10 (AY176608)	98	----	
JB601SF.94	3	Haptophyta	Uncultured marine eukaryote clone SC5-14p (FR676884)	97	----	

Chapitre 5 - Conclusions générales

5.1 Principales contributions de la thèse

Dans le golfe du Maine et le plateau néo-écossais, la diversité des microorganismes planctoniques a principalement porté sur les cellules de grandes tailles (>5µm), alors que la diversité taxonomique des cellules plus petites restait majoritairement inconnue (Li et al., 2011). Les principaux résultats obtenus au court de cette thèse offrent une vision approfondie de l'état de ces communautés picoplanctoniques dans le temps et l'espace, en particulier pour les picoeucaryotes. De plus, portant sur deux profondeurs dites écologiquement importantes qui sont la surface (3 m) et le maximum de chlorophylle, ils apportent un complément sur ces acteurs de la production primaire qui représentent une source majeure de matière organique pour les écosystèmes aquatiques.

Nos résultats mettent en évidence des communautés picoeucaryotes distinctes dans les différents sites d'échantillonnage, en soulignant la façon dont ces communautés sont associées aux différentes masses d'eau. Le rôle de la stratification sur la diversité taxonomique a été démontré, où il semblerait que la température et/ou de la salinité soient deux facteurs importants dans la structure des assemblages. Les cellules hétérotrophes variaient de façon plus importante en fonction des conditions hydrographiques et semblaient plus sensibles par rapport aux cellules autotrophes. En effet, les principaux groupes autotrophes étaient répertoriés dans tous nos échantillons environnementaux (haptophytes, chlorophytes, stramenopiles et cryptophytes, ces derniers montrant quelques variations), alors que parmi la composante hétérotrophe certains taxa montraient une présence ou absence en fonction du site ou de la période d'échantillonnage (stramenopiles marines heterotrophs et Telonemia). Ce constat peut s'expliquer par le mode de nutrition mixotrophe de certains picoeucaryotes phototrophes qui peuvent être d'important bactérivores (Zubkov and Tarran, 2008; Frias-Lopez et al., 2009; Sanders and Gast, 2012; Hartmann et al., 2012). De plus, certaines études ont mis en évidence que l'augmentation de la température favoriserait leur mode de nutrition hétérotrophe au détriment de la photosynthèse (Rose et Caron, 2007, Hansen, 2011; Wilken, 2013).

Nos résultats confirment la dominance du picoplancton eucaryote par les Chlorophyta, ainsi que l'extrême diversité des Haptophyta. L'ordre des Mamiellales était dominant au sein des Chlorophyta, où plusieurs genres ont été répertoriés : 3 clades de *Micromonas pusilla* (I, III, V), probablement 2 clades distincts de *Bathycoccus*, et *Ostreococcus* pour la première fois mise en évidence dans cette région. Deux autres genres ont été identifiés au sein de l'ordre des Pyramimonadales : *Pterosperma* et *Pyramimonas*. Globalement, *Bathycoccus* et *Micromonas* représentaient les principaux producteurs primaires au sein de cette fraction de taille, cependant la relative abondance de *Pyramimonas* (Jordan Basin) et *Pterosperma* (Northeast Channel) dans le golfe du Maine en juillet a été soulignée. Cette dominance de la lignée verte au sein du picoplancton eucaryote a précédemment était mise en évidence au sein d'autres régions comme la Manche (Not et al., 2005) ou l'Arctique (Lovejoy et al., 2006). L'étude de la taxonomie des cellules picoeucaryotes démontre aussi la relative abondance ainsi que la grande diversité des Haptophyta. Divers ordres ont été reportés : Prymnesiales, Phaeocystales, Syracosphaerales, Coccosphaerales, ou Zygodiscales. Au différent temps d'échantillonnages, Prymnesiales dominait, excepté en avril où la relative abondance de Phaeocystales était supérieure. L'utilisation du gène *psbA* a permis de marquer cette extrême diversité où de nombreuses séquences non puent être classifiées. Ce succès écologique de cette lignée rouge quelques soient les conditions hydrographiques a précédemment était mis en évidence notamment par le caractère mixotrophe de nombreux haptophytes (Liu et al., 2009; Kirkham et al., 2011).

L'identification de nombreux nanoflagellés hétérotrophes largement incultivés et rarement identifiés par microscopie a aussi été répertoriée. Parmi eux étaient les choanoflagellés, les katablepharids, les picozoa, les Telonemia, les cercozoa, et les stramenopiles marines (MAST). Jouant un rôle important dans le broutage des bactéries et reminéralisation de la matière organique, la détection de l'activité de ces hétérotrophes s'avère très intéressante pour ses implications au niveau du flux de carbone et des nutriments. Nos recherches soulignent l'abondance relative de Telonemia et plusieurs clades parmi les MAST (clade-3, -4, et -7). Les marines alvéolées (MALV) infectant de nombreuses espèces planctoniques ont aussi été relativement abondantes dans notre étude comme généralement dans d'autres études environnementales (Guillou et al., 2008). Il a été

souligné que les MALV groupe III sont surreprésentés utilisant ARNr contrairement aux MALV-I et II toujours surreprésentes dans l'ADNr, laissant supposer que les MALV-III étaient potentiellement en mode infectieux lors de notre échantillonnage (juillet, golfe du Maine).

5.2 Synthèses des principales variations taxonomiques dans le temps

Dans le but de cibler des taxa indicateurs de conditions environnementales particulières, le bilan des majeurs changements taxonomiques entre nos trois périodes d'échantillonnages est effectué. On constate que dans des eaux de surface relativement froides et riches en nutriments rencontrées en avril, la biomasse phytoplanctonique totale était élevée, alors que le picophytoplancton (dominé par les picoeucaryotes) indiquait des concentrations relativement basses contrairement aux deux autres périodes d'échantillonnages (**Table 5.1**). Les espèces autotrophes majoritairement répertoriées étaient des diatomées, *Bathycoccus* (Chlorophyta), l'ordre des phaeocystales parmi les haptophytes, et le complexe Bolido-Parmales caractéristique des maximums cellulaires des picoeucaryotes. Parmi les hétérotrophes, les Cercozoa (Rhizaria) avec les Picozoa et les choanoflagellés étaient dominants. Les stramenopiles marines s (MAST) clade -1 et -7 étaient aussi caractéristiques de ces eaux. À l'inverse, en juillet les eaux de la couche supérieure étaient plus chaudes et pauvres en nutriments, et en contraste à la situation d'avril, la biomasse phytoplanctonique totale était plus faible avec une nette augmentation du nombre de cellules phytoplanctoniques (dominé par *Synechococcus*) (**Table 5.1**). Les picoeucaryotes autotrophes identifiés ont montré un changement dans la dominance relative de certains taxa avec l'ordre des Prymnesiales parmi les haptophytes, *Bathycoccus* et *Micromonas* (chlorophytes), conjointement à la relative importance de *Pyramimonas* et de *Pterosperma*. Au sein des hétérotrophes, l'importance de Telonemia dans Jordan Basin (golfe du Maine) a été mise en évidence, et comme en avril les MAST-1c et-7, ainsi que MAST-4 étaient dominant parmi les stramenopiles marines . Finalement, en octobre les eaux de la surface présentaient une situation intermédiaire dans les paramètres physico-chimiques, avec une concentration en chlorophylle *a* totale plus ou moins stable à celle du

mois du juillet, et un maximum dans les concentrations cellulaires picophytoplanctonique (**Table 5.1**). Au sein de la composante photosynthétique, l'ordre des Prymnesiales (Haptophyta) et *Micromonas* (Chlorophyta) étaient comme en juillet relativement dominants. Parmi les hétérotrophes, les katablepharids (reportés en juillet, mais pas en avril), et MAST-3 (phylogénétiquement différente des MAST-3 d'avril) étaient caractéristiques des eaux de la surface en automne.

5.3 Discussion sur la méthodologie

L'utilisation de plusieurs méthodes moléculaires pour l'étude de la diversité des picoeucaryotes a apporté de la solidité dans nos conclusions. Cela a permis premièrement, l'obtention d'un grand nombre de séquences de référence presque complètes du gène 18S rRNA provenant de cette région via les banques de clones (**Chapitre 2**). Ces séquences pourront être utilisées pour de prochaines études phylogénétiques notamment celles utilisant le séquençage à haut débit fournissant de courtes séquences pas très efficaces pour des études phylogénétiques poussées. Deuxièmement, notre étude comparative de l'ADNr et de l'ARNr utilisant le séquençage à haut débit apporte des éclaircissements sur l'estimation de la diversité et l'activité métabolique des picoeucaryotes (**Chapitre 3**). Elle souligne que même si l'ARN est souvent utilisé pour réduire les biais liés à l'ADN, l'utilisation de l'un ou de l'autre dépend de la physiologie des organismes, et ce malgré un effort de séquençage. Troisièmement, l'utilisation des transcrits du gène *psbA* s'avère une approche efficace pour cibler et estimer la diversité des producteurs primaires potentiellement actifs eucaryotes (**Chapitre 4**), où la comparaison de leur distribution par rapport aux picocyanobactéries nous éclaire un peu plus sur les potentialités écologiques des différentes composantes du picophytoplancton.

Le fractionnement des échantillons en fonction de leur taille est une pratique commune dans les études environnementales et constitue une étape importante pour l'étude du picophytoplancton. Néanmoins, des séquences correspondantes à des cellules supérieures à 3 µm sont régulièrement reportées dans les échantillons de la petite fraction de taille comme les dinoflagellés, les ciliés, les diatomées, voire même des métazoaires (Vaulot et al. 2008). Pour limiter cet effet, Sorensen et al. (2013) suggère d'utiliser des

filtres de 0.8 µm (limite de la taille cellulaire du plus petit eucaryote connu) au lieu de filtre 0.2 µm, ce qui évite d'amplifier par PCR l'ADN/ARN des larges cellules qui généralement se brisent lors de la filtration. L'utilisation du tri des populations par cytométrie en flux s'est aussi avérée une technique très efficace pour isoler et enrichir les échantillons environnementaux de cellules picoeucaryotiques (Shi et al., 2011; Balzano et al., 2012).

De même, les approches moléculaires souffrent de plusieurs biais, notamment lors de l'étape de PCR (Von Wintzingerode et al., 1997). Les amorces utilisées pour l'amplification par PCR peuvent fixer avec une efficacité variable parmi les différentes espèces ou groupes dans un seul échantillon et ainsi conduire à une amplification biaisée qui dépend aussi du nombre de copies du gène 18S de ARNr (Zhu et al, 2005; Potvin et Lovejoy, 2009). L'étape de PCR induit aussi plusieurs autres biais lors du séquençage tels que la présence de séquences chimériques, des insertions, des délétions, ou des mauvaises identifications des nucléotides. Le séquençage à haut débit n'empêche pas ces erreurs rencontrées lors de la PCR, de même pendant le séquençage plusieurs séquences peuvent se retrouvées sur la même bille, et il peut y avoir extension des séquences homopolymères ou des erreurs d'interprétation peuvent surgirent lors de l'interprétation du flux du travail (Huse et al., 2010).

Finalement, l'identification des taxa répertoriés en avril et octobre par l'utilisation des banques de clones est à prendre avec précaution sachant que cette méthode a tendance à favoriser les espèces les plus abondantes (**Chapitre 2**). De plus, le ciblage du gène 18S ARNr n'indique en rien que ces espèces soient métaboliquement actives lors de nos échantillonnages. L'analyse de l'expression du gène ribosomique (**Chapitre 3 et 4**) s'est donc avérée une comparaison intéressante pour associer la présence des petits eucaryotes à leur potentielle activité métabolique au moment de l'échantillonnage. De plus, le séquençage à haut débit apportant une profondeur d'échantillonnage satisfaisante (**Chapitre 3**), a permis d'obtenir une meilleure estimation de l'état des communautés. Souligons que l'abondance relative des espèces répertoriées soit par banque clone, soit par pyroséquençage n'est pas proportionnelle à leur biomasse. Cela nous permet seulement d'indiquer des tendances dans les changements au sein de la communauté des picoeucaryotes.

5.4 Perspectives

À la vue de nos résultats, il nous apparait important d'approfondir les recherches sur les mixotrophes qui tiennent un rôle important dans le transfert d'énergie et de nutriments du réseau trophique microbien marins. De même, il faudrait isoler certains représentants des OTUs répertoriés dans nos recherches dont beaucoup sont encore incultivés afin de mieux comprendre leur écologie. L'apport du séquençage d'un plus grand nombre de génomes provenant de cellules picoeucaryotiques apporterait des informations cruciales sur les gènes impliqués dans certaines conditions environnementales, et permettrait peut- être de mettre en évidence la présence d'îlots génomiques spécialisés dans l'adaptation à des niches écologiques particulières. Et bien que l'impact des facteurs biotique ne soit pris en compte dans nos recherches, le broutage ou la liste exerçant des pressions sélectives importantes en dehors des facteurs abiotiques.

La cote atlantique du Canada constitue une importante zone de pêche. Par exemple, la pêche sur le plateau néo-écossais débute en janvier, mais les crevettiers du Nouveau-Brunswick débuteraient leurs activités vers la mi-mars. De même, la saison de la pêche au crabe des neiges commence au mois d'avril et se termine en juillet. Durant ces périodes, nos échantillons ont montré la présence de potentielles algues toxiques comme *Aureococcus* ou *Pseudochattonella* pouvant si elles se développent à des concentrations importantes, rendre ces produits impropres à la consommation. Cela peut engendrer des impacts économiques importants sur l'activité de pêche. À côté de cela, nos résultats ont montré que la composante picoplanctonique avait une contribution plus importante au sein de la biomasse totale au mois d'octobre comparé à avril ou juillet. Ceci implique une biomasse notable pour les organismes filtreurs, le zooplancton et ainsi une quantité de nourriture plus importante pour les poissons de tout genre. Tout ce réseau sous-entend des stocks de pêche plus élevée au cours de cette période. Finalement, dans un contexte plus large, l'importante quantité de nos données peuvent servir dans des modèles de modélisations notamment ceux cherchant à éclaircir le cycle du carbone, tout comme les données apportées aux satellites pour l'estimation de la production primaire.

Tableau 5.2 : Tableau résumant les données physiques, chimiques, et biologiques des trois pèriodes d'échantillonnages dans le golfe du Maine en Juillet 2010 (été), et sur le plateau Néo-Écossais en Avril 2009 (printemps) et Octobre 2009 (automne).

Variables	**PRINTEMPS**	**ÉTÉ**	**AUTOMNE**
Physiques			
Température (°C)	1 – 5	10 – 22	9 – 17
Salinité	31.2 – 33.1	32.1 – 33.5	30.3 – 33.3
Chimiques			
Nitrate (µM)	0.6 – 3.9	0.2 – 2.3	0.1 – 4.6
Phosphate (µM)	0.4 – 0.9	0.1 – 0.7	0.1 – 0.6
Silicate (µM)	0.2 – 3.0	0.3 – 3.6	1.2 – 4.6
Biologiques			
Total Chl *a* (µg L^{-1})	0.7 – 16.3	0.11 – 1.33	0.3 – 1.2
Picoeucaryotes (× 10^3 cells mL^{-1})	0.6 – 6.5	4 – 43	4 – 73
Synechococcus (× 10^3 cells mL^{-1})	0.2 – 1.2	13 – 117	33 – 133
Prochlorococcus (× 10^3 cells mL^{-1})	n.d.	3 – 31	9 – 18

Bibliographie

Adl SM, Simpson AGB, Lane CE, Lukes J, Bass D, Bowser SS, *et al.* (2012). The revised classification of eukaryotes. *J Euk Microbiol*, **59:** 429–493.

Agawin NSR, Duarte CM, Agustí S (2000). Nutrient and temperature control of the contribution of picoplankton to total phytoplankton biomass and production. *Limnol Oceanogr*, **45**: 591–600.

Aljanabi SM, Martinez I (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.*, **25:** 469–4693.

Allakhverdiev SI, Nishiyama Y, Miyairi S, Yamamoto H, Inagaki N, et al.(2002). Salt stress inhibits the repair of photodamaged photosystem II by suppressing the transcription and translation of psbA genes in *Synechocystis*. *Plant Physiol*, **130:** 1443–1453.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25:** 3389–3402.

Andersen RA, Bidigare RR, Keller MD, Latasa M (1996). A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep Sea Res II,* **43:** 517–537.

Anderson R, Wylezich C, Glaubitz S, Labrenz M, Jürgens K (2013). Impact of protist grazing on a key bacterial group for biogeochemical cycling in Baltic Sea pelagic oxic / anoxic interfaces. *Environ Microbiol*, **15**: 1580–1594.

Aro EM, Mc Caffery S, Anderson JM (1993a). Photoinhibition and D1 degradation in peas acclimated to different growth irradiances. *Plant Physiol,* **103:** 835–843.

Aro EM, Virgin I, Andersson B (1993b). Photoinhibition of photosystem II, inactivation, protein damage and turnover. *Biochim Biophys Acta*, **1143:** 113–134.

Arrigo KR (2005). Marine microorganisms and global nutrient cycles. *Nature*, **437:** 349–355.

Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F (1983). The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser*, **10:** 257–263.

Bagley MJ, Franson SE, Christ SA, Waits ER, Toth GP (2002). Genetic Diversity as an Indicator of Ecosystem Condition and Sustainability: Utility for Regional Assessment of Streams in the Eastern United States. In: Agency EP (ed), Cincinnati, OH, USA

Balzano S, Gourvil P, Siano R, Chanoine M, Marie D, Lessard S, Sarno D, Vaulot D (2012). Diversity of cultured photosynthetic flagellates in the northeast Pacific and Arctic Oceans in summer. *Biogeosciences*, **9**: 4553–4571.

Barber J, Andersson B (1992). Too much of a good thing: light can be bad for photosynthesis. *Trends Biochem Sci,* **17:** 61–66.

Barton AD, Dutkiewicz S, Flierl G, Bragg J, Follows MJ (2010). Patterns of diversity in marine phytoplankton. *Science*, **327**: 1509–1511.

Bérard-Therriault L, Poulin M, Bosse L (1999). Guide d'identification du phytoplankton marin de l'estuaire et du golfe du Saint-Laurent incluant également certains protozoires. *Can Spec Publ Fish Aquat Sci,* **128**, NRC Research Press Ottaway 387pp.

Berry J, Bjorkman O (1980). Photosynthetic response and adaptation to temperature in higher-plants. *Annu Rev Plant Physiol,* **31:** 491–543.

Blazewicz S, Barnard RL, Daly RD, Firestone MK (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J,* **7:** 2061–2068.

Booth BC, Larouche P, Bélanger S, Klein B, Amiel D, Mei ZP (2002). Dynamics of *Chaetoceros socialis* blooms in the North Water. *Deep-Sea Res II*, **49**: 5003–5025.

Bougis P (1974). Écologie du plancton marin, Le phytoplankton. Masson et Cie, Paris: 196 pp.

Bouman HA, Ulloa O, Scanlan DJ. *et al.* (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science*, **312:** 918–921.

Brate J, Krabberod AK, Dolven JK, Ose RF, Kristensen T, Bjorklund KR, Shalchian-Tabrizi K (2012). Radiolaria associated with large diversity of marine alveolates. *Protist*, **163:** 767–777.

Brate J, Klaveness D, Rygh T, Jakobsen KS, Shalchian- Tabrizi K (2010). Telonemia-specific environmental 18S rDNA PCR reveals unknown diversity and multiple marine- freshwater colonizations. *BMC Microbiol*, **10**: 168.

Breton E, Brunet C, Sautour B, Brylinski JM (2000). Annual variations of phytoplankton biomass in the Eastern English Channel: comparison by pigment signatures and microscopic counts. *J Plankton Res,* **22**: 1423–1440.

Brown MV, Philip GK, Bunge JA, Smith MC, Bissett A, Lauro FM *et al.* (2009). Microbial community structure in the North Pacific Ocean. *ISME J*, **3:** 1374–1386.

Buitenhuis ET, Li WKW, Vaulot D, Lomas MW, Landry MR, Partensky F, *et al.* (2012) Picophytoplankton biomass distribution in the global ocean. *Earth Syste Sci Data*, **4:** 37–46.

Calbet A, Landry MR (2004). Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnol Oceanogr,* **49:** 51–57.

Campbell L, Nolla HA, Vaulot D (1994). The importance of Prochlorococcus to community structure in the central North Pacific Ocean. *Limnol Oceanogr,* **39:** 954–961.

Canter HM, Lund JWG (1953). Studies on plankton parasites II, the parasitism of diatoms with special reference to lakes in the English Lake District. *Trans Br Mycol Soc*, **36:** 13–37.

Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A (2012). Marine protistan diversity. *Annu Rev Mar Sci*, **4**: 467–493.

Caron DA, Peele ER, Lim EL, Dennett MR (1999). Picoplankton and nanoplankton and their trophic coupling in surface waters of the Sargasso Sea south of Bermuda. *Limnol Oceanogr,* **44:** 259–272.

Caron D (1994). Inorganic nutrients, bacteria, and the microbial loop. *Microb Ecol,* **28:** 295–298.

Cavalier-Smith T (2005). Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann Bot-London,* **95:**147–175.

Chambouvet A, Morin P, Marie D, Guillou L (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science,* **322:** 1254–57.

Charpy L, Blanchot J (1998). Photosynthetic picoplankton in French Polynesian atoll lagoons: estimation of taxa contribution to biomass and production by flow cytometry. *Mar Ecol Prog Ser,* **162:** 57–70.

Charvet S, Vincent WF, Comeau A, Lovejoy C (2012). Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. *Front Microbiol*, **3:** 422.

Church MJ, Ducklow HW, Letelier RM, Karl DM (2006). Temporal and vertical dynamics in picoplankton photoheterotrophic production in the subtropical North Pacific Ocean. *Aquat Microb Ecol*, **45:** 41–53.

Clarke KR (1993) .Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol*, **18**: 117–143.

Coats DW, Park MG (2002). Parasitism of photosynthetic dinoflagellates by three strains of *Amoebophrya* (Dinophyta): parasite survival, infectivity, generation time, and host specificity. *J Phycol*, **38:** 520–28.

Comeau A, Li WKW, Tremblay JE, Carmack E, Lovejoy C (2011). Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLOS One, **6:** 0027492.

Courties C, Perasso R, Chrétiennot-Dinet MJ, Gouy M, Troussellier M (1998). Phylogenetic analysis and genome size of *Ostreococcus tauri* (Chlorophyta, Prasinophyceae). *J Phycol*, **34:** 844–849.

Courties C, Vaquer A, Trousselier M, Lautier J, Chrétiennot-Dinet MJ, Neveux J, *et al.* (1994). Smallest eukaryotic organisms. *Nature*, **370:** 255

Cuvelier ML, Allen AE, Monier A, Mc Crow JP, Messié M, Tringe SG, *et al.* (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *PNAS*, **107:** 14679–14684.

Cuvelier M, Ortiz A, Kim E, Moehlig H, Richardson DE, *et al.* (2008). Widespread distribution of a unique marine protistan lineage. *Environ Microbiol,* **10:**1621–34.

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

Degerlund M, Huseby S, Zingone A, Sarno D, Landfald B (2012). Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). *J Plankton Res*, **34**: 416–431.

Del Campo J, Massana R (2011). Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist*, **162:** 435–448.

DeLong E, Karl D (2005). Genomic perspectives in microbial oceanography. *Nature,* 437: 336–342.

De Vargas C, Aubry MP, Probert I, Young J (2007). Origin and evolution of coccolithophores: from coastal hunters to oceanic farmers. In *Evolution of Primary Producers in the Sea*. Falkowski, P.G., and Knoll, A.H. (eds). London, UK: Elsevier, 251–285.

Díez B, Pedrós-Alió C, Massana R (2001). Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA Gene cloning and sequencing. *Appl Environ Microbiol,* **67:** 2932–2941.

Durham W, Kessler J, Stocker R (2009). Disruption of vertical motility by shear triggers formation of thin phytoplankton layers. *Science,* **323:** 1067–1070.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**: 2194–2200.

Edvardsen B, Eikrem W, Shalchian-Tabrizi K, Riisberg I, Johnsen G, Naustvoll L, Throndsen J (2007). *Verrucophora farcimen* gen. et sp. nov. (Dictyochophyceae, Heterokonta), a bloom-forming ichthyotoxic flagellate from the Skagerrak, Norway. *J Phycol*, **43**: 1054–1070.

Eikrem W, Throndsen J (1990). The ultrastructure of *Bathycoccus* gen. nov. and *B. Prasinos* sp. nov., a non-motile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. *Phycologia,* **29**: 344–350.

Ellegaard M, Godhe A, Harnstrom K, Mc Quoid M (2008). The species concept in a marine diatom: LSU rDNA-based phylogenetic differentiation in *Skeletonema marinoi/dohrnii* (Bacillariophyceae) is not reflected in morphology. *Phycologia*, **47**: 156–167.

Fenchel T (2008). The microbial loop-25 years later. *J Exp Mar Biol Ecol,* **366:** 99–103.

Ferris MJ, Palenik B (1998). Niche adaptation in ocean cyanobacteria. *Nature,* **396:** 226-228.

Fierer N (2008). Microbial biogeography: patterns in microbial diversity across space and time. In *Accessing Uncultivated Microorganisms: From the Environment to Organisms and Genomes and Back*. Zengler, K. (ed.). Washington, DC, USA: ASM Press, 95–115.

Flombaum P, Gallegos JL, Gordillo LA, Rincon J, *et al.* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA*, **110:** 9824–9829.

Flynn KJ, Stoecker DK, Mitra A, Raven JA, Glibert PM, Hansen PJ, Graneli E, Burkholder JM (2012). Misuse of the phytoplankton-zooplankton dichotomy: the need to assign organisms as mixotrophs within plankton functional types. *J Plankton Res*, **35:** 10.1093/plankt/fbs062.

Fogg GE (1991). Transley Review No. 30, the phytoplanktonic ways of life. *New Phytologist*, **118:** 191–232.

Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, Holling CS (2004). Regime shifts, resilience, and biodiversity in ecosystem management. *Annu Rev Ecol Evol Syst,* **35:** 557–581.

Follows MJ, Dutkiewicz S (2011) Modeling diverse communities of marine microbes. *In* C. A. Carlson and S. J. Giovannoni [eds.]. *Ann Rev Mar Sci*, **3**: 427–451.

Frankignoul C, de Coetlogon G, Joyce TM, Dong SF (2001). Gulf Stream variability and ocean-atmosphere interactions. *J Phys Oceanogr*, **31**: 3516–3529.

Frias-Lopez J, Thompson A, Waldbauer J, Chisholm S (2009). Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environ Microbiol,* **11:** 512–525.

Friedland KD, Hare JA, Wood GB, Col LA, Buckley LJ, Mountain DG, Kane J, *et al.* (2008). Does the fall phytoplankton bloom control recruitment of Georges Bank haddock, *Melanogrammus aeglefinus*, through parental condition?. *Can J Fish Aquat Sci*, **65**: 1076–1086.

Fuller NJ, et al. (2006). Molecular analysis of photosynthetic picoeukaryote communitystructure along an Arabian Sea transect. *Limnol Oceanogr,* **51:** 2502–2514.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol*, **69:** 2430– 2443.

Garrett CJR, Keeley JR, Greenberg DA (1978). Tidal mixing versus stratification in the Bay of Fundy and Gulf of Maine. *Atmos Ocean*, **16:** 403–423.

Gibb SW, Cummings DG, Irigoien X, Barlow RG, Fauzi R, Mantoura C (2001). Phytoplankton pigment chemotaxonomy of the northeastern Atlantic. *Deep Sea Res Part II*, **48**: 795–823.

Giovannoni S, Stingl U (2005). Molecular diversity and ecology of microbial plankton. *Nature*, **437:** 343–348.

Gobler CJ, Sunda WG (2012). Ecosystem disruptive algal blooms of the brown tide species, *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. *Harmful Algae*, **14**: 36–45.

Golden SS (1994). Light-responsive gene expression and the biochemistry of the photosystem II reaction center. *Mol Biol Cyano*, 693–714.

Gómez F, Moreira D, Benzerara K, Lopez-Garcia P (2011). *Solenicola setigera* is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. *Environ Microbiol*, **13**: 193–202.

Gómez F, Moreira D, López-García P (2009). Life cycle and molecular phylogeny of the dinoflagellates *Chytriodinium* and *Dissodinium*, ectoparasites of copepod eggs. *Eur J Protistol,* **45**: 260–270.

Gong J, Dong J, Liu X, Massana R (2013). Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist*, **164:** 369–379.

Greene CH, Pershing AJ (2007). Climate drives sea change. *Science*, **315:** 1084–1085.

Greenan BJW, Petrie BD, Harrison WG, Strain PM (2008). The onset and evolution of a spring bloom on the Scotian Shelf. *Limnol Oceanogr*, **53**: 1759–1775.

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R, *et al.* (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (Alveolata). *Environ Microbiol*, **10:** 3349–3365.

Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A, Saenger W (2009). Cyanobacterial photosystem II at 2.9-angstrom resolution and therole of quinones, lipids, channels and chloride. *Nat Stru Mol Biol,* **16:** 334–342.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/Nt. *Nucleic Acids Symp Ser*, **41**: 95–98.

Hamady M, Lozupone C, Knight R (2010). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J*, **4:** 17–27.

Hammer A, Pitchford J (2005). The role of mixotrophy in plankton bloom dynamics, and the consequences for system productivity. *J Mar Sci*, **62:** 833–840.

Hammer A, Schumann R, Schubert H (2002). Light and temperature acclimation of *Rhodomonas salina* (Cryptophyceae): photosynthetic performance. *Aquat Microb Ecol,* **29:** 287–296.

Hammer Ø, Harper DAT, Ryan PD (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electron*, **4**: 9.

Hansen PJ (2011). The role of photosynthesis and food uptake for the growth of marine mixotrophic dinoflagellates. *J Euk Microbiol*, **58:** 203–214.

Hansen G, Botes L, De Salas M (2007). Ultrastructure and LSU rDNA sequences of *Lepidodinium viride* reveal a close relationship to *Lepidodinium chlorophorum* comb. nov. (=*Gymnodinium chlorophorum*). *Phycol Res*, **55**: 25–41.

Hansen G (2001). Ultrastructure of *Gymnodinium aureolum* (Dinophyceae): toward a further redefinition of *Gymnodinium sensu stricto*. *J Phycol*, **37**: 612–23.

Harding T, Jungblut AD, Lovejoy C, Vincent WF (2011). Microbes in high Arctic snow and implications for the cold biosphere. *Apply Environ Microbiol*, **77:** 3234–3243.

Hartmann M, Zubkov MV, Scanlan DJ, Lepère C (2013). *In situ* interactions between photosynthetic picoeukaryotes and bacterioplankton in the Atlantic Ocean: evidence for mixotrophy. *Environ Microbiol Rep,* DOI: 10.1111/1758-2229.12084.

Hartmann M, Grob C, Tarran GA, Martin AP, Burkill PH, Scanlan DJ, Zubkov MV (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc Natl Acad Sci*, **109**: 5756–5760.

Head EJH, Pepin P (2010). Spatial and inter-decadal variability in plankton abundance and composition in the Northwest Atlantic (1958-2006). *J Plankton Res*, **32**: 1633–1648.

Hendriks L, Goris A, Neefs JM, Van De Peer Y, Hennebert G, De Wachter R (1989). The nucleotide-sequence of the small ribosomal-subunit RNA of the yeast *Candida albicans* and the evolutionary position of the fungi among the Eukaryotes. *Syst Appl Microbiol*, **12:** 223–229.

Houliez E, Lizon F, Lefebvre S, Artigas LF, Schmitt FG (2013). Short-term variability and control of phytoplankton photosynthetic activity in a macrotidal ecosystem (the Strait of Dover, eastern English Channel). *Mar Biol,* **160:** 1661–1679.

Howeth JG, Leibold MA (2010). Species dispersal rates alter diversity and ecosystem stability in pond metacommunities. *Ecology*, **91:** 2727–2741.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol*, **8:** R143.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol*, **12:** 1889–1898.

Ichinomiya M, Yoshikawa S, Kamiya M, Ohki K, Takaichi S, Kuwata A (2011). Isolation and charcterization of Parmales (Heterokonta /Heterokontophyta /Stramenopiles) from the Oyahio region, western North Pacific. *J Phycol*, **47**: 144–151.

Jardillier L, Zubkov MV, Pearman J, Scanlan DJ (2010). Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J,* **4:** 1180–1192.

Jeffrey SW (1976). A report of green algal pigments in the central North Pacific Ocean. *Mar Biol,* **37:** 33–37.

Jeong HJ, Yoo YD, Kang NS, Rho JR, Seong KA, Park JW, Nam GS, Yih WH (2010). Ecology of *Gymnodinium aureolum*. I. Feeding in western Korean water. *Aquat Microb Ecol*, **59**: 239–255.

Johnson C, Harrison G, Head E, Spry J, Pauley K, Maass H, Kennedy M, *et al.* (2012). Optical, chemical, and biological oceanographic conditions in the Maritimes Region in 2009 and 2010. *DFO Can Sci Advis Sec Res Doc*, 2012/012: iv + 64p.

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

Joint I, Pomroy A, Savidge G, Boyd P (1993). Size- fractionated primary productivity in the northeast Atlantic in May-July 1989. *Deep Sea Res* Part B, **40:** 423–440.

Josephson KL, Gerba CP, Pepper IL (1993). Polymerase chain-reaction detection of nonviable bacterial pathogens. *Appl Environ Microbiol*, **59**: 3513–3515.

Jossi JW, Goulet JR (1993). Zooplankton trends: US north-east shelf ecosystem and adjacent regions differ from north-east Atlantic and North Sea. *ICES Journal of Marine Science*, **50:** 303–313.

Kane J (2011). Multiyear variability of phytoplankton abundance in the Gulf of Maine. *ICES J Mar Sci*, **68:** 1833−1841.

Katoh K, Toh H (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, **26**: 1899–1900.

Kelly KA, Gille ST (1990). Gulf Stream surface transport and statistics at 69°W from the Geosat altimeter. *J Geophys Res*, **95**: 3149–3161.

Kerouel R, Aminot A (1997). Fluorometric determination of ammonia in sea and estuarine waters by direct segmented flow analysis. *Mar Chemistry*, **57:** 265–275.

King N, Young SL, Abedin M, Carr M, Leadbeater BSC (2009). The choanoflagellates: heterotrophic nanoflagellates and sister group of the metazoa, *Cold Spring Harb Protoc*, doi:10.1101/pdb.emo116.

Kirkham AR, Lepère C, Jardillier LE, Not F, Bouman H, Scanlan DJ (2013). A global perspective on marine photosynthetic picoeukaryote community structure. *ISME J,* **7:** 922–936.

Kirkham AR, Jardillier LE, Tiganescu A, Pearman J, Zubkov MV, Scanlan DJ (2011). Basin-scale distribution patterns of photosynthetic picoeukaryotes along an Atlantic Meridional Transect. *Environ Microbiol,* **13**: 975–990.

Kirchman DL (2008). Introduction and Overview. In: Kirchman, D. L. [Ed.] *Microbial Ecology of the Oceans*. 2nd ed. John Wiley & Sons, Inc., New Jersey, 1–26.

Koeppel AF, Wu M (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Res,* **41:** 5175–5188.

Kruskal WH, Wallis WA (1952). Use of ranks in one-criterion variance analysis *J Am Stat Assoc,* **47:** 583.

Lepère C, Vaulot D, Scanlan DJ (2009). Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. *Environ Microbiol*, **11:** 3105–3117.

Leterme SC, Pingree RD (2007). Structure of phytoplankton (Continuous Plankton Recorder and SeaWiFS) and impact of climate in the northwest Atlantic shelves. *Ocean Science*, **3**: 105–116.

Levasseur M, Therriault JC, Legendre L (1984). Hierarchical control of phytoplankton succession by physical factors. *Mar Ecol Prog Ser*, **19**: 211–222.

Li WKW, Andersen RA, Gifford DJ, Incze LS, Martin JL, Pilskaln CH, Rooney-Varga, *et al.* (2011). Planktonic microbes in the Gulf of Maine Area. *PLOS One*, **6**: e20981.

Li WKW (2009). From cytometry to macroecology: a quarter century quest in microbial oceanography. *Aquat Microb Ecol*, **57**: 239–251.

Li WKW, Harrison WG, Head EJH (2006). Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. *Proc R Soc B*, **273**: 1953–1960.

Li WKW, Dickie PM (2001). Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry*, **44**: 236–246.

Li WKW (1998). Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnol Oceanogr*, **43**: 1746–1753.

Li WKW (1995). Composition of ultraphytoplankton in the central North Atlantic. *Mar Ecol Prog Ser,* **122**: 1–8.

Li WKW (1994). Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr,* **39:** 169–175.

Lin YC, Campbell T, Chung CC, Gong GC, Chiang KP, Worden AZ (2012). Distribution patterns and phylogeny of marine stramenopiles in the north Pacific ocean. *Appl Environ Microbiol*, **78**: 3387–3399.

Lindell D, Sullivan MB, Johnson ZI, Tolenen AC, Rohwer F, Chisholm SW (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA,* **155:** 11013–11018.

Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, *et al.* (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *PNAS*, **106:** 12803–12808.

Lochte K, Ducklow HW, Fasham MJR, Stienens C (1993). Plankton succession and carbon cycling at 47°N 20°W during the JGOFS North Atlantic bloom experiment. *Deep Sea Res II*, **40**: 91–14.

Loder JW, Petrie B, Gawarkiewicz G (1998). The coastal ocean off northwestern North America: A large-scale view. pp. 105–133. *In: A.R. Robinson and K.H. Brink (eds).* The Sea, Vol 11. John Wiley and Sons.

Logares R, Audic S, Santini S, Pernice MC, de Vargas C, Massana R (2012). Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J,* **6:** 1823–1833.

Longhurst A (1998). Ecological geography of the sea. *Academic Press Ltd*, London.

Longhurst A (1995). Seasonal cycles of pelagic production and consumption. *Prog Oceanogr*, **36**: 77–167.

López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature,* **409:** 603–607.

Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R, Potvin M, *et al.* (2007). Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic seas. *J Phycol*, **43**: 78–89.

Lovejoy C, Massana R, Pedrós-Alió C (2006). Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl Environ Microbiol*, **72:** 3085–3095.

Lozupone C, Hamady M, Knight R (2006). UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*, **7**: 371.

Lozupone C, Knight R (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microb,* **71**: 8228–8235.

Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, Chisholm SW (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J*, **4**: 1252–1264.

Malmstrom RR, Kiene RP, Vila M, Kirchman DL (2005). Dimethylsulfoniopropionate (DMSP) assimilation by *Synechococcus* in the Gulf of Mexico and northwest Atlantic Ocean. *Limnol Oceanogr*, **50:**1924–1931.

Man-Aharonovich D, Philosof A, Kirkup BC, Le Gall F, Yogev T, *et al.* (2010). Diversity of active marine picoeukaryotes in the Eastern Mediterranean Sea unveiled using photosystem-II *psbA* transcripts. *ISME J,* **4:**1044–1052.

Mann NH, Cook A, Millard A, Bailey S, Clokie MRJ (2003). Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature*, **424:** 741.

Margulies M, Egholm M, Altman W E, *et al.* 2005. Genome sequencing in microfabricated highdensity picolitre reactors. *Nature*, **437:** 376–380.

Marie D, Simon N, Vaulot D (2005). Phytoplankton cell counting by flow cytometry. pp 253–267. *In Andersen, R. A. [ed.]*, Algal Culturing Techniques. Academic Press.

Martinez RR, Rocap G, Salazar G, Massana R (2013). Biogeography of the uncultured marine picoeukaryote MAST-4: temperature-driven distribution patterns. *ISME J*, **7:** 1531–1543.

Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW (2009). Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus. Environ Microbiol,* **11:** 823–832.

Massana R (2011). Eukaryotic picoplankton in surface oceans. *Ann Rev Microbiol,* **65**: 91–110.

Massana R, Unrein F, Rodríguez-Martínez R, Forn I, Lefort T, Pinhassi J, Not F (2009). Grazing rates and functional diversity of uncultured heterotrophic flagellates. *ISME J*, **3:** 588–596.

Massana R, Terrado R, Forn I, Lovejoy C, Pedros-Alio C (2006). Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ Microbiol*, **8**: 1515–1522.

Massana R, Castresana J, Balague V, Guillou L, Romari K, Groisillier A, Valentin K, Pedrós-Alió C (2004). Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol*, **70**: 3528–3534.

Massana R, Guillou L, Diez B, Pedros-Alio C (2002). Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Appl Environ Microbiol*, **68**: 4554–4558.

Mather L, Mac Intosh K, Kaczmarska G, Klein G, Martin JL (2010). A checklist of diatom species reported (and presumed native) from Canadian coastal waters. *Can Tech Rep Fish Aquat Sci*, **2881**, iii+78.

McDonald S, Sarno D, Scanlan D, Zingone A (2007). Genetic diversity of eukaryotic ultraphytoplankton in the Gulf of Naples during an annual cycle. *Aquat Microb Ecol*, **50:** 75–89.

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988). The characterization of enzymatically amplifie eukaryotic 16S-Like rRNA-coding regions. *Gene*, **71**: 491–499.

Metfies K, Gescher C, Frickenhaus S, Niestroy R, Wichels A, Gerdts G, Knefelkamp B, Wiltshire KH, Medlin L (2010). Contribution of the class Cryptophyceae to phytoplankton structure in the German Bight. *J Phycol*, **46**: 1152–1160.

Mills EL (1989). Biological oceanography: an early history. *New York: Cornell University Press,* 1870–1960.

Mitchell MR, Harrison G, Pauley K, Gagné A, Maillet G, Strain P (2002). Atlantic zonal monitoring program, sampling protocol. *Can Tech Rep Hydrogr Ocean Sci*, **223**, iv + 23.

Monier A, Terrado R, Thaler M, Comeau AM, Medrinal E, Lovejoy C (2013). Upper Arctic Ocean water masses harbor distinct communities of heterotrophic flagellates. *Biogeosciences*, **10**: 3397–3430.

Monier A, Welsh RM, Gentemann C, Weinstock G, Sodergren E, Armbrust EV, Eisen JA, Worden AZ (2012). Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. *Environ Microbiol*, **14**: 162– 76.

Moon van der Staay SY, Wachter RD, Vaulot D (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature,* **409:** 607–610.

Moore LR, Chisholm SW **(**1999). Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr*, **44:** 628–638.

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

Moore LR, Goericke R, Chisholm SW (1995). Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar Ecol Prog Ser,* **116:** 259–275.

Morden CW, Delwiche CF, Kuhsel M, Palmer JD (1992). Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems*, **28 :** 75–90.

Moreau H, Verhelst B, Couloux A, Derelle E, *et al.* (2012). Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol,* **13:** R74.

Mousseau L, Legendre L, Fortier L (1996). Dynamics of size-fractionated phytoplankton and trophic pathways on the Scotian Shelf and at the shelf break, northwest Atlantic. *Aquat Microb Ecol*, **10**:149–163.

Muhling M, Fuller NJ, Somerfield PJ, Post AF, Wilson WH, Scanlan DJ, Joint I, Mann NH (2006). High resolution genetic diversity studies of marine *Synechococcus* isolates using *rpoC1*-based restriction fragment length polymorphism. *Aquat Microb Ecol,* **45:**2 63–275.

Muhling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, Scanlan DJ, et al. (2005). Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environ Microbiol*, **7:** 499–508.

Mulo P, Sakurai I, Aro EM (2012). Strategies for *psbA* gene expression in cyanobacteria, green algae and higher plants: From transcription to PS II repair. *Biochim Biophys Acta-Bioenerg,* **1817:** 247–257.

Murphy LS, Haugen EM (1985). The distribution and abundance of phototrophic ultraplankton in the North Atlantic. *Limnol Oceanogr*, **30**: 47–58.

Naeem S, Li SB (1997). Biodiversity enhances ecosystem reliability. *Nature,* **390:** 507–509.

Nei M, Kumar S (2000). Molecular evolution and phylogenetics. *Oxford University Press*, New York.

Nickelsen J, Rengstl B (2013). Assembly of photosystem II: from cyanobacteria to plants. *Annu Rev Plant Biol,* **64:** 609–635.

Not F, Siano R, Kooistra WHCF, Simon N, Vaulot D, Probert I (2012). Diversity and ecology of eukaryotic marine phytoplankton. *Adv Bot Res*, **64**: 1–53.

Not F, Del Campo J, Balagué V, De Vargas C, Massana R (2009). New insights into the diversity of marine picoeukaryotes. PLOS One, **4:** e7143.

Not F, Latasa M, Scharek R, Viprey M, Karleskind P, Balagué V, *et al.* (2008). Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep Sea Res I,* **55:** 1456–1473.

Not F, Gausling R, Azam F, Heidelberg J, Worden A (2007). Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ Microbiol,* **9:** 1233–1252.

Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N (2004). A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. *Appl Environ Microbiol,* **70:** 4064–4072.

Nusch EA (1980). Comparison on different methods for chlorophyll and phaeopigment determination. *Arch Hydrobiol Beih Ergebn Limnol*, **14**: 14–36.

Olson RJ, Zettler ER, Altabet MA, Dusenberry JA, Chisholm SW (1990). Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep Sea Res I,* **37:** 1033–1051.

Park MG, Park JS, Kim M, Yih W (2008) Plastid dynamics during survival of Dinophysis caudata without its ciliate prey. *J Phycol,* **44:**1154-1163.

Parker MS, Mock T, Armbrust EV. 2008. Genomic insights into marine microalgae. *Annu. Rev. Genet.* 42:619–645.

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

Partensky F, Blanchot J, Vaulot D (1999b). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review, p. 457–475. *In* L. Charpy and A. Larkum (ed.), Marine cyanobacteria, vol. 19. Musée Océanographique, Monaco.

Partensky F, Hoepffner N, Li WKW, Ulloa O, Vaulot D (1993). Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and the Mediterranean Sea. *Plant Physiol*, **101:** 285–296.

Pedrós-Alió C (2006). Marine microbial diversity: can it be determined? *Trends Microbiol,* **14:** 257.

Piwosz K, Pernthaler J (2010). Seasonal population dynamics and trophic role of planktonic nanoflagellates in coastal surface waters of the southern Baltic Sea. *Environ Microbiol*, **12**: 364–377.

Poulton N, Thompson B, Tupper B, Cucci T, Their E, et al. (2009). Monitoring microbial patterns in plankton during a multiyear study in Boothbay Harbour, Maine. Gulf of Maine Symposium: Advancing ecosystem research for the future of the Gulf.

Pomeroy L (1974). The ocean's food web: a changing paradigm. *Bioscience*, **24:** 409–504.

Potvin M, Lovejoy C (2009). PCR-based diversity estimates of artificial and environmental 18S rRNA gene libraries. *J Euk Microbiol*, **56**: 174–181.

Price MN, Dehal PS, Arkin AP (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLOS One*, **5**: e9490.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Gloeckner FO (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*, **35**: 7188–7196.

Rasconi S, Niquil N, Sime-Ngando T (2012). Phytoplankton chytridiomycosis: community structure and infectivity of fungal parasites in aquatic systems. *Environ Microbiol*, **14**: 2151–2170.

Raven JA (1998). Essay Review - The twelfth Transley lecture. Small is beautiful: the Picophytoplankton. *Functional Ecology*, **12:** 503–513.

Rocap G, Larimer FW, Lamerdin J. *et al.* (2003). Genome divergence in two *Prochlorococus* ecotypes reflects oceanic niche differentiation. *Nature*, **424**: 1042–1047.

Rocap G, Distel D, Waterbury JB, Chislom SW (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S rDNA internal transcribed spacer sequences. Appl Env Microbiol, **68:** 1180–1191.

Rodríguez-Martínez R, Rocap G, Salazar G, Massana R (2013). Biogeography of the uncultured marine picoeukaryote MAST-4: temperature driven distribution patterns. *ISME J,* 7: 1531–1543.

Rodriguez F, Derelle E, Guillou L, LeGall F, VaulotD, Moreau H (2005). Ecotype diversity in themarine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol,* **7:** 853–859.

Romari K, Vaulot D (2004). Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr*, **49**: 784–798.

Rose JM, Caron DA (2007). "Does low temperature constrain the growth rates of heterotrophic protists? Evidence and implications for algal blooms in cold water." *Limnol Oceanogr,* **52:** 886–895.

Roy S, Chanut JP, Gosselin M, Sime-Ngando T (1996). Characterization of phytoplankton communities in the lower St. Lawrence estuary using HPLC-detected pigments and cell microscopy. *Mar Ecol Prog Ser*, **142**: 55–73.

Saldarriaga JF, Taylor F, Cavalier-Smith T, Menden-Deuer S, Keeling PJ (2004). Molecular data and the evolutionary history of dinoflagellates. *Eur J Protistol*, **40**: 85–111.

Sanders RW, Gast RJ (2012) Bacterivory by phototrophic picoplankton and nanoplankton in Arctic waters. *FEMS Microbiol Ecol*, **82:** 242–253.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci*, **74:** 5463–5467.

Sarmento H, Gasol JM (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ Microbiol*, **14:** 2348–2360.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, *et al.* (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev*, **73:** 249–299.

Scanlan DJ **(**2003). Physiological diversity and niche adaptation in marine *Synechococcus*. *Adv Microb Physiol,* **47:** 1–64.

Scanlan DJ, West NJ (2002). Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol Ecol,* **40:** 1–12.

Scherer S, Herrmann G, Hirschberg J, Böger P (1991). Evidence for multiple xenogenous origins of plastids: comparisons of psbA genes with a xanthophyte sequence. *Curr Genet,* **19:** 503–507.

Schlitzer R (2002) Interactive analysis and visualization of geoscience data with Ocean Data View. *Comput Geosci*, **28**: 1211–1218.

Schloss PD (2009) A high-throughput DNA sequence aligner for microbial ecology studies. *PLOS One*, **4**: e8230.

Seenivasan R, Sausen N, Medlin LK, Melkonian M (2013) *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the Picozoa Phylum nov., a widespread group of picoeukaryotes, formerly known as 'Picobiliphytes'. *PLOS One*, **8**: e59565.

Selosse MA, Loiseaux-de-Goer S (1997). La saga de l'endosymbiose-Les mitochondries et les plastes temoins et acteurs de l'evolution. La recherche. **296:** 36–41.

Shalchian-Tabrizi K, Kauserud H, Massana R, Klaveness D, Jakobsen KS (2007). Analysis of environmental 18S ribosomal RNA sequences reveals unknown diversity of the cosmopolitan phylum Telonemia. *Protist*, **158**: 173–180.

Shapiro LP, Guillard RRL (1986). Physiology and ecology of the marine eukaryotic ultraplankton. *Can Bull Fish Aquat Sci*, **214**: 371–390.

Sherman K, Jaworski NA, Smayda TJ (1996). The Northeast Shelf ecosystem. *Blackwell Scientific*, Oxford. 564 p.

Sherr EB, Sherr BF (2009). Capacity of herbivorous protists to control initiation and development of mass phytoplankton blooms. *Aquat Microb Ecol*, **57:** 253–262.

Sherr EB, Sherr BF (2002). Significance of predation by protists in aquatic microbial food webs. *A Van Leeuw J Microb*, **81**: 293–308.

Shi XL, Lepere C, Scanlan DJ, Vaulot D (2011). Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLOS One*, **6:** e18979.

Siano R, Alves-De-Souza C, Foulon E, Bendif EM, Simon N, Guillou L, Not F (2011). Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. *Biogeosciences,* **8**: 267–278.

Sieburth JMcN, Smetacek V, Lenz J (1978). Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationships to plankton size fractions. Limnol.Oceanogr. **23:** 1256–1263.

Simon M, López-García P, Moreira D, Jardillier L (2013). New haptophyte lineages and multiple independent colonizations of freshwater ecosystems. *Environ Microbiol Rep,* **5:** 322–332.

Skjelbred B, Horsberg TE, Tollefsen KE, Andersen T, Edvardsen B (2011). Toxicity of the ichthyotoxic marine flagellate *Pseudochattonella* (Dictyochophyceae, Heterokonta) assessed by six bioassays. *Harmful Algae*, **10**: 144–154.

Slapeta J, Lopez-Garcia P, Moreira D (2006). Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol*, **23**: 23–29.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *PNAS*, **103:** 12115–12120.

Song HJ, Ji RB, Stock C, Kearney K, Wang ZL (2011). Interannual variability in phytoplankton blooms and plankton productivity over the Nova Scotian Shelf and in the Gulf of Maine. *Mar Ecol Prog Ser*, **426**: 105–118.

Sørensen N, Daugbjerg N, Richardson K (2013). Choice of pore size can introduce artefacts when filtering picoeukaryotes for molecular biodiversity studies. *Microbial Ecology*, **65:** 964–968.

Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow CE, Sachdeva R, *et al.* (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J*, **5**: 1414–1425.

Stiller JW, Hall BD (1997). The origin of red algae: Implication for plasmid evolution. *Proc Nat Acad Sci USA*, **94:** 4520–4525.

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology,* **19:** 21–31.

Stoeck T, Zuendorf A, Breiner H-W, Behnke A (2007). A molecular approach to identify active microbes in environmental eukaryote clone libraries. *Microbial Ecology*, **53:** 328–339.

Stoupin D, Kiss AK, Arndt H, Shatilovich AV, Gilichinsky DA, Nitsche F (2012). Cryptic diversity within the choanoflagellate morphospecies complex *Codosiga botrytis* – Phylogeny and morphology of ancient and modern isolates. *Eur J Protistol*, **48**: 263–273.

Suttle CA, Chan AM, Cottrell MT (1990). Infection of phytoplankton by viruses and reduction of primary productivity. *Nature*, **387**: 467–469.

Takishita K, Yubuki N, Kakizoe N, Inagaki Y, Maruyama T (2007). Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extremophiles*, **11**: 563–576.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, **28**: 2731–2739.

Tamura K (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol Biol Evol*, **9**: 678–687.

Taylor FJR, Hoppenrath M, Saldarriaga JF (2008). Dinoflagellate diversity and distribution. *Biodivers Conserv*, **17**: 407–418.

Ter Braak CJF, Smilauer P (2002). CANOCO reference manual and canodraw for windows user's guide: software for canonical community ordination (version 4.5). *Microcomputer Power*, Ithaca, New York, 500 pp.

Terrado R, Medrinal E, Dasilva C, Thaler M, Vincent W, Lovejoy C (2011). Protist community composition during spring in an Arctic flaw lead polynya. *Polar Biol*, **34**: 1901–1914.

Thaler M (2013). Diversité et répartition des flagellés hétérotrophes dans l'océan Arctique. Thèse université Laval.

Thaler M, Lovejoy C (2012). Distribution and diversity of a protist predator Cryothecomonas (Cercozoa) in arctic marine waters. *J Eukaryot Microbiol*, **59**: 291–299.

Therriault JC, Petrie B, Pepin P, Gagnon J, Gregory D, Helbig J, Herman A, *et al.* (1998). Proposal for a Northwest Atlantic Zonal Monitoring Program. *Can Tech Rep Hydrogr Ocean Sci*, **194**, vii+57 p.

Throndsen J (1976). Occurence and productivity of small marine flagellates. *Norw J Bot,* **23:** 269–293.

Townsend DT, Thomas AC, Mayer LM, Thomas M, Quinlan JW (2006). Oceanography of the northwest Atlantic continental shelf. *In* Robinson AR, Brink KH, eds. *The Sea*, Cambridge: Harvard University Press, MA **14**, 119–168.

Tzahor S, Man-Aharonovich D, Kirkup BC, Yogev T, Berman-Frank I, et al. (2009). A supervised learning approach for taxonomic classification of core-photosystem-II genes and transcripts n the marine environment. *BMC Genomics*, **10:** 229.

Unrein F, Gasol JM, Not F, Forn I, Massana R (2013). Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *ISME J*, in press.

Vaulot D, Lepère C, Toulza E, De la Iglesia R, Poulain J, Gaboyer F, Moreau H, Vandepoele, *et al.* (2012). Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. *PLOS One*, **7:** e39648.

Vaulot D, Eikrem W, Viprey M, Moreau H. (2008). The diversity of small eukaryotic phytoplankton (< 3 µm) in marine ecosystems. *FEMS Microbiol Rev*, **32:** 795–820.

Vaulot D, Partensky F, Neveux J, Mantoura R, Llewellyn C (1990). Wintertime presence of prochlorophytes in surface waters of the North-Western Mediterranean Sea. *Limnol Oceanogr,* **35:** 1156–1164.

Veldhuis MJW, Timmermans KR, Croot P, *et al*. (2005) Picophytoplankton; a comparative study of their biochemical composition and photosynthetic properties. *J Sea Res*, **53**: 7–24.

Von Wintzingerode F, Gobel U, Stackebrandt E (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev,* **21:** 213–229.

Wang G, Wang X, Liu X, Li Q (2012). Diversity and biogeochemical function of planktonic fungi in the ocean. In: Raghukumar C. (ed) Biology of marine fungi, Vol 53. Springer, Berlin, 71−88.

Wang K, Chen F (2008). Prevalence of highly hostspecific cyanophages in the estuarine environment. *Environ Microbiol*, **10:** 300–312.

Waterbury JB, Watson SW, Valois FW, Franks DG (1986). Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus. Can Bull Fish Aquat Sci*, **214**: 71–120.

Wauthy B, Desrosières R, Le Bourhis J (1967). Importance présumée de l'ultraplancton dans les eaux tropicales oligotrophes du Pacifique central sud. *Cah. O.R.S.T.O.M*., *Ser Oceanogr,* **5:** 109–116.

Wilken S, Huisman J, Naus-Wiezer S, Donk E (2013). Mixotrophic organisms become more heterotrophic with rising temperature. *Ecol Lett*, **16**: 225–233.

Woese C, Kandler O, Wheelis M (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS*, **87:** 4576–4579.

Wommack KE, Colwell RR (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev*, **64:** 69–114.

Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL, *et al. (*2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, **324:** 268–272.

Worden AZ, Not F (2008). Ecology and Diversity of Picoeukaryotes. Book Chapter in: *Microbial Ecology of the Ocean, 2nd Edition.* Ed. D. Kirchman. Wiley.

Worden AZ (2006). Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol,* **43**: 165–175.

Worden AZ, Nolan JK, Palenik B (2004). Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnol Oceanogr,* **49:** 168–179.

Xue H, Chai F, Pettigrew NR (2000). A model study of the seasonal circulation in the Gulf of Maine. *J Phys Oceanogr,* **30:** 1111–1135.

Yeats P, Ryan S, Harrison G (2010). Temporal Trends in Nutrient and Oxygen Concentrations in the Labrador Sea and on the Scotian Shelf. *AZMP,* Bulletin **n9**.

Yoon HS, Price DC, Stepanauskas R, Sieracki ME, Wilson WH, Yang EC, *et al.* (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science*, **332**: 714–717.

Zeidner G, Beja O (2005). Community level analysis of phototrophy: psbA gene diversity. In Methods in enzymolgy, **397:** 372–380.

Zeidner G, Preston CM, Delong EF, Massana R, Post AF, Scanlan DJ, Béjà O (2003). Molecular diversity among marine picophytoplankton as revealed by *psbA* analysis. *Env Microbiol,* **5:** 212–216.

Zhang Z, Green BR, Cavalier-Smith T (2000). Phylogeny of ultra rapidly evolving dinoflagellate chloroplast genes. A possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol,* **51:** 26–40.

Zhu F, Massana R, Not F, Marie D, Vaulot D (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol*, **52:** 79–92.

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

Zubkov MV, Tarran GA (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature,* **455:** 224–226.

Zubkov MV, Tarran GA (2005). Amino acid uptake of *Prochlorococcus* spp. in surface waters across the South Atlantic Subtropical Front. *Aquat Microb Ecol*, **40:** 241–249.

Zwirglmaier K, Spence E, Zubkov MV, Scanlan DJ, Mann NH (2009).Differential grazing of two heterotrophic nanoflagellates on marine *Synechococcus* strains. *Environ Microbiol,* **11:**1767–1776.