



HAL
open science

Réponse à l'échelle génomique des communautés planctoniques au réchauffement climatique et conséquences sur leur biogéographie

Paul Frémont

► **To cite this version:**

Paul Frémont. Réponse à l'échelle génomique des communautés planctoniques au réchauffement climatique et conséquences sur leur biogéographie. Ecologie, Environnement. Université Paris-Saclay, 2022. Français. NNT : 2022UPASL062 . tel-03967591

HAL Id: tel-03967591

<https://theses.hal.science/tel-03967591>

Submitted on 1 Feb 2023

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

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

Réponse à l'échelle génomique des communautés planctoniques au réchauffement climatique et conséquences sur leur biogéographie.

Response at the genomic scale of plankton communities to climate change and consequences on their biogeography.

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577, Structure et dynamique des systèmes vivants (SDSV)

Spécialité de doctorat: Écologie

Graduate School: Life Sciences and Health. Référent: Université d'Évry Val d'Essonne

Thèse préparée dans l'unité de recherche **GM – Génomique Métabolique – DRF/JACOB/CEA**, sous la direction de **Olivier Jaillon**, Directeur de Recherche (CEA, IBFJ) et la codirection de **Marion Gehlen**, Directrice de Recherche (CEA, LSCE)

Thèse soutenue à Paris-Saclay, le 8 Novembre 2022, par

Paul Frémont

Composition du Jury

Hélène Morlon Directrice de Recherche, CNRS	Présidente
Laurent Bopp Professeur, ENS Directeur de Recherche, CNRS	Rapporteur & Examineur
Sakina-Dorothee Ayata Maîtresse de conférences, Sorbonne Université	Rapporteuse & Examinatrice
Christophe Ambroise Professeur des universités, Université d'Évry Val d'Essonne	Examineur
Samuel Chaffron Chercheur, CNRS	Examineur
Olivier Jaillon Directeur de Recherche, CEA	Directeur de thèse
Marion Gehlen Directeur de Recherche, CEA	Co-Directrice de thèse

Titre : Réponse à l'échelle génomique des communautés planctoniques au réchauffement climatique et conséquences sur leur biogéographie.

Mots clés : Plancton, Omiques environnementales, Changement climatique, *Tara* Océans, Biogéographie, Modélisation

Résumé : Le plancton marin est constitué d'organismes microscopiques allant des virus jusqu'aux petit métazoaires en passant par les bactéries et les protistes. Il est transporté passivement par les courants, prospère sur l'ensemble des océans et a plusieurs rôles cruciaux au sein du système Terre. Le phytoplancton permet par la photosynthèse la production primaire de matière organique dans les océans soutenant l'ensemble des réseaux trophiques océaniques. Le plancton participe aussi à la pompe à carbone biologique, mécanisme par lequel la matière organique sédimente vers les fonds marins et y est stockée. Il est aujourd'hui primordial d'évaluer et projeter la réponse du plancton au changement climatique provoqué par la combustion des énergies fossiles. Dans cette thèse, j'étudie cette question sous le prisme de la biogéographie, discipline s'intéressant à la distribution des organismes dans et en interaction avec leur milieu à travers le temps et l'espace.

Dans une première partie, j'étudie à l'aide des données omiques des expéditions *Tara* Océans et des modèles climatiques la distribution du plancton dans les océans et sa réponse au changement climatique. Il est mis en évidence un partitionnement des océans en provinces génomiques dépendant de la taille des organismes. Ces provinces sont mises en relation avec les paramètres physico-chimiques à l'aide de techniques de machine learning et extrapolées à l'ensemble des océans. Un ensemble de « génomes signatures » pour chaque province est aussi mis en évidence. Une importante réorganisation des provinces en réponses au changement climatique est projetée sur environ 50% des océans d'ici la fin du siècle. D'importants changements en composition du plancton sont projetés et mis en lien avec une diminution de 4% de la pompe à carbone biologique ce qui aurait un effet aggravant

sur le changement climatique.

Dans une seconde partie, j'étudie les changements de l'expression des gènes du plancton eucaryote le long de la transition entre l'océan Atlantique Nord et le bassin Arctique. Parmi les variables physiques, la température est la variable expliquant le mieux les changements transcriptionnels. L'analyse fonctionnelle des gènes corrélant avec le gradient fort de température révèle une stratégie commune d'acclimatation des algues eucaryotes. Cette stratégie inclut d'importants changements dans la machinerie d'expression génétique mais aussi la surexpression d'un ensemble de fonctions liées à l'acclimatation au froid. A l'échelle des communautés, des processus liés à la maintenance du pool protéique et de la machinerie transcriptionnelle semblent plus actifs dans l'arctique.

Dans une troisième partie, je développe un modèle mécanistique basé individu de communauté de phytoplancton structurée par la taille et advectée dans un champ de vitesse imitant la circulation océanique Nord Atlantique. Le modèle représente les principaux groupes de phytoplancton: les cyanobactéries, les algues eucaryotes et les diatomées. Les caractéristiques écologiques de bases résultant de simulations sont présentées : α -diversité, biogéographie et biomasse. Une attention est apportée à l'importance du nombre de particules modélisées et son impact sur la diversité maximale représentée ainsi que l'influence des courants sur les niches écologiques des organismes modélisés. Cette thèse aborde des sujets et enjeux fondamentaux de l'écologie contemporaine et son apport s'inscrit dans la révolution omique du début du XXI^{ème} siècle. Une biogéographie génomique du plancton est proposée et sa réponse au réchauffement climatique est explorée.

L'analyse omique *in situ* de la physiologie des algues unicellulaires du plancton à l'échelle de la communauté permet de mettre en évidence leurs capacités d'acclimatation directement dans leur milieu de vie. Enfin, le modèle basé individu du phytoplancton permet d'explorer théoriquement de nombreuses questions écologiques à mettre en lien dans le futur avec l'analyse des données omiques.

Title: Response at the genomic scale of plankton communities to climate change and consequences on their biogeography.

Keywords : Plankton, Environmental omics, Climate change, *Tara* Oceans, Biogeography, Modeling

Abstract: Microscopic marine plankton are organisms ranging from viruses to small metazoans, bacteria and protists. Plankton is transported passively by the currents, thrives in every oceans and plays a crucial role in the Earth system. Through photosynthesis, phytoplankton drives primary production and the resulting flux of organic matter supports the entire oceanic food web. Plankton also participates in the biological carbon pump, a mechanism by which organic matter sediments to the seabed and is stored there. Today it is essential to assess and project the response of plankton to climate change caused by the burning of fossil fuels. In this thesis, I study this issue through the lens of biogeography, a discipline that focuses on the distribution of organisms in and interacting with their environment through time and space.

In a first part, I study the distribution of plankton in the global oceans and its response to climate change using omics data from the *Tara* Oceans expeditions and climate models. A partitioning of the oceans into genomic provinces depending on the size of the organisms is described. These provinces are related to physico-chemical parameters using machine learning techniques and extrapolated to the whole ocean. A set of "signature genomes" for each province is also highlighted. A major reorganization of the provinces in response to climate change is projected over approximately 50% of the oceans by the end of the century. Important changes in plankton composition would result in a decrease by 4% of the global carbon export flux, which would have an aggravating effect on climate change.

In a second part, I study the changes in gene expression of eukaryotic plankton along the transition between the North Atlantic Ocean and the Arctic basin. Among the physical variables, temperature is the variable that best explains

transcriptional changes. Functional analysis of genes correlating with the strong temperature gradient reveals a common acclimation strategy of eukaryotic algae. This strategy includes important changes in the gene expression machinery but also overexpression of a set of functions related to cold acclimation. At the community level, processes related to the maintenance of the protein pool and the transcriptional machinery seem to be more active in the Arctic.

In a third part, I develop a mechanistic individual based model of a size-structured phytoplankton community advected in a velocity field mimicking the North Atlantic Ocean circulation. The model represents the main phytoplankton groups: cyanobacteria, eukaryotic algae and diatoms. Basic ecological characteristics resulting from simulations are presented: α -diversity, biogeography and biomass. Attention is given to the importance of the number of particles modeled and its impact on the maximum diversity represented as well as the influence of currents on the ecological niches of the modeled organisms.

This thesis addresses fundamental topics and challenges in contemporary ecology and it contributes to the ongoing omic revolution of the early 21st century. A genomic biogeography of plankton is proposed and its response to global warming is explored. The in situ omics analysis of the physiology of unicellular planktonic algae at the community scale highlights their acclimatization capacities directly in their living environment. Finally, the individual based model of phytoplankton allows to explore theoretically many ecological questions to be linked in the future with the analysis of omic data.

Remerciements

Je tiens tout d'abord à remercier mes encadrants de thèse, Olivier Jaillon et Marion Gehlen. Vous m'avez apporté votre expertise scientifique (autour de nombreuses conversations passionnantes) mais aussi votre savoir-faire dans le métier de chercheur. Cette formation à la recherche que vous m'avez apportée va me permettre de poursuivre le métier de chercheur et mener à bien mes futurs projets. Vous m'avez aussi permis de développer mon réseau professionnel notamment autour du consortium *Tara Océans*.

Je remercie ensuite Patrick Wincker pour m'avoir accueilli au Genoscope et pour avoir suivi les avancées de mes travaux de thèse.

Je remercie chaleureusement tous mes collègues et anciens collègues du Laboratoire d'Analyse Génomique des Eucaryotes (le LAGE, mon laboratoire d'accueil) que j'ai rencontré de loin ou de près. Je les remercie pour leur bonne humeur au quotidien et aussi pour les différentes collaborations que nous avons pu réaliser en commun. Je tiens ainsi à remercier, avec un clin d'œil particulier, à tous les doctorants et ceux avec qui j'ai partagé le bureau, en oubliant personne je l'espère, Majda Arif, Nicolas Bisson, Laurie Bertrand, Marie Burel, Lucie Cartairade, Quentin Carradec, Margaux Crédeville, Tom Delmont, Evangelia Eleftheriou, Sarah Farhat, Rhita Felali, Morgan Gaia, Nina Guérin, Damien Hinsinger, Soheib Kerbache, Romuald Laso-Jadart, Arnaud Meng, Clément Leboine, Jade Leconte, Julie Lê-Hoang, Amin Madoui, Sophie Mangenot, Céline Orvain, Éric Pelletier, Betina Porcel, Lucas Pavlovic, Julie Poulain, Achal Rastogi, Jana Rigonato, Yoann Seeleuthner, Kevin Sugier, Adrien Thurotte et Alexei Vorobev.

Je remercie l'ensemble des chercheurs, équipes et personnels du Genoscope et particulièrement le Laboratoire de Séquençage (LS) dirigé par Pedro H. Oliveira et l'équipe de Bioinformatique pour la Génomique et la Biodiversité (LBGB) dirigée par Jean-Marc Aury sans lesquels mon travail aurait été impossible. Je remercie aussi le Laboratoire d'Informatique Scientifique (LIS) dirigé par Claude Scarpelli pour son assistance précieuse pour l'utilisation du cluster de calcul, la gestion de l'informatique et du matériel

informatique. Je remercie aussi particulièrement Catherine Sarlande, Nancy Delpech, Marie-Thérèse Corey, Ludivine Douchet et Catherine Contrepois pour leur aide précieuse dans l'organisation et la gestion administrative ainsi que l'organisation des déplacements. Je remercie aussi Mathieu Dubois pour les organisations du fameux « Genopub » ainsi que les autres doctorants du genscope que j'ai pu côtoyer, notamment Jérôme Arnoux, Adelme Bazin, Sophia Belkhef, Oriane Della-Negra, Laurine Ducrot, Guillaume Gautreau, Chloé Lelièvre, Déborah Martin, Humbeline Paupelin, Nazim Sarica, et Marion Schulz.

Je remercie particulièrement Daniele Ludicone qui m'a accueilli six mois dans son laboratoire à Naples à la Stazione Zoologica Anton Dohrn. Je remercie aussi les différents collègues de la Stazione Zoologica, Greta Busseni, Lucia Campese, Luigi Caputi, Bruno Hay Mele, Florian Kokoszka, Maurizio Ribera D'Alcala et Eloïse Trabut.

Je remercie l'ensemble des co-auteurs des articles auxquels j'ai participé sans pour autant les avoir tous rencontrés. La liste est malheureusement trop longue pour tous les citer. Je remercie notamment Matthieu Vrac du Laboratoire des Sciences du Climat et de l'Environnement (LSCE) pour les discussions autour des statistiques et Daniel J Richter pour sa collaboration autour du manuscrit « Biadvection ». Je remercie aussi Noan Le Bescot pour son assistance graphique pour les figures.

Je remercie l'ensemble des membres du consortium *Tara* Océans et des personnes travaillant autour des expéditions de la goélette *Tara*. Je remercie notamment Lee Karp Boss qui m'a accueilli une semaine chez elle et dans son laboratoire à l'université du Maine.

Je remercie les membres de mon comité de thèse pour avoir suivi mes travaux de recherche lors de deux réunions : Laurent Bopp, Mahendra Mariadassou, Meike Vogt et Gwenaël Piganeau.

Je remercie les membres du jury pour avoir évalué et, je l'espère, apprécié ma thèse : Christophe Ambroise, Laurent Bopp, Sakina Dorothee-Ayata, Samuel Chaffron et Hélène Morlon.

Je remercie les membres gestionnaires de l'école doctorale Structure et Dynamique des

Systemes Vivants (SDSV), Florence Gonnet avec qui j'ai réalisé deux rendez-vous pédagogique, Noémie Moutty et Jessica Samé pour la gestion administrative.

Je remercie enfin chaleureusement ma famille et mes amis qui m'ont soutenu tout au long de ma thèse, avec qui j'ai pu en discuter librement. Je les remercie aussi pour tous les bons moments, vacances, sorties, restos, kebabs, fêtes et autres mariages passés ensemble ces dernières années sans lesquelles la vie de thésard serait bien plus compliquée.

TABLE DES MATIERES

INTRODUCTION	11
PREAMBULE A L'INTRODUCTION	11
I. LE PLANCTON, ACTEUR MAJEUR DE L'ECOSYSTEME MARIN	11
<i>I.1. Définition du plancton</i>	11
<i>I.2. Position écologique du plancton</i>	13
II. LA PLANCTOLOGIE, LES EXPEDITIONS TARA OCEANS ET LES METHODES METAOMIQUES DE SEQUENÇAGE DE L'ADN	16
<i>II.1. Bref historique de la planctologie</i>	16
<i>II.2. Expéditions et techniques modernes</i>	20
<i>II.3. Les expéditions Tara Océans</i>	24
<i>II.4. Les révolutions omiques et métaomiques: d'importantes avancées scientifiques</i>	30
III. LA BIOGEOGRAPHIE MARINE: LE CONTEXTE DU SEASCAPE	37
<i>III.1. Le contexte du seascape</i>	37
<i>III.2. La circulation océanique : le rôle des courants</i>	42
<i>III.3. Les facteurs limitant la croissance du phytoplancton : température, taille et nutriments, courants et lumière</i>	46
<i>III.4. Etudes biogéographiques actuelles du plancton</i>	53
IV. LES GRANDS CYCLES BIOGEOCHIMIQUES ET LA MODELISATION DU CLIMAT	55
<i>IV.1 La place du plancton dans les grands cycles biogéochimiques et dans le système Terre</i>	55
<i>IV.2. La modélisation biogéochimique au sein des modèles système terre</i>	59
<i>IV.3. Le changement climatique et sa modélisation</i>	62
DEMARCHE SCIENTIFIQUE	66
CHAPITRE 1 : ETUDE DE LA BIOGEOGRAPHIE GENOMIQUE DU PLANCTON : CHANGEMENT CLIMATIQUE ET CYCLES BIOGEOCHIMIQUES	70
I. INTRODUCTION ET RESUME: UNE BIOGEOGRAPHIE GENOMIQUE	70
II. ARTICLE 1: GENOMIC EVIDENCE FOR GLOBAL OCEAN PLANKTON BIOGEOGRAPHY SHAPED BY LARGE-SCALE CURRENT SYSTEMS	82
III. ARTICLE 2: RESTRUCTURING OF PLANKTON GENOMIC BIOGEOGRAPHY IN THE SURFACE OCEAN UNDER CLIMATE CHANGE	85
IV. ARTICLE 3: FUNCTIONAL REPERTOIRE CONVERGENCE OF DISTANTLY RELATED EUKARYOTIC PLANKTON LINEAGES REVEALED BY GENOME-RESOLVED METAGENOMICS	147
CHAPITRE 2 : CHANGEMENT D'EXPRESSION GENETIQUE DU PLANCTON EUCARYOTE DANS LA TRANSITION DE L'ATLANTIQUE NORD A L'OCEAN ARCTIQUE	149
I. INTRODUCTION ET RESUME: LA QUESTION DE L'ACCLIMATATION ET DE L'ATLANTIFICATION	149
II. ARTICLE PRÉLIMINAIRE: GENE EXPRESSION CHANGES OF SURFACE OCEAN EUKARYOTIC PLANKTON IN THE NORTH ATLANTIC TO ARCTIC OCEAN TRANSITION	154
CHAPITRE 3 : MODELISATION DE LA BIOGEOGRAPHIE DU PHYTOPLANCTON, UN MODELE BASE INDIVIDU DANS UN DOUBLE GYRE DE L'ATLANTIQUE NORD IDEALISE	245
I. INTRODUCTION : UNE COMPREHENSION MECANISTIQUE DE LA BIOGEOGRAPHIE ET DE LA BIODIVERSITE?	245
II. RÉSULTATS PRÉLIMINAIRES: ECOLOGICAL CHARACTERISTICS OF A SIZE STRUCTURED INDIVIDUAL BASED MODEL OF A PHYTOPLANKTON COMMUNITY IN AN IDEALIZED NORTH ATLANTIC DOUBLE GYRE FLOW	247
CONCLUSION	280
I. CONTEXTE, ENJEUX ET QUESTIONS SCIENTIFIQUES	280
II. PRINCIPAUX RESULTATS, DISCUSSIONS ET PERSPECTIVES	282
REFERENCES	288
ANNEXE	317

I. CONTRIBUTIONS DIVERSES	317
1. Article 1: Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean.	317
2. Article 2: Genomic adaptation of the picoeukaryote <i>Pelagomonas calceolata</i> to iron-poor oceans revealed by a chromosome-scale genome sequence.	319
II. SUPPLÉMENTS	321
1. Figures de l'article "Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems"	321

INTRODUCTION

PREAMBULE A L'INTRODUCTION

Cette thèse traite de la biogéographie génomique du plancton et de sa réponse au réchauffement climatique. Je présenterai tout d'abord dans l'introduction notre objet d'étude à proprement parler, c'est-à-dire le plancton ainsi que son rôle dans l'écosystème marin. J'aborderai dans une seconde partie l'histoire des recherches et connaissances sur le plancton puis décrirai les recherches actuelles en particulier dans le cadre de la révolution technologique lié à la génomique dans laquelle s'inscrit cette thèse. Dans une troisième partie, j'aborderai les problèmes et connaissances scientifiques liés à la question de la biogéographie marine du plancton. La quatrième partie décrira ensuite la place du plancton dans les grands cycles biogéochimiques des éléments, la modélisation biogéochimique puis introduira la problématique du réchauffement climatique. Cette dernière sera néanmoins évoquée tout au long de l'introduction. Les différents chapitres de la thèse incluant nos résultats seront aussi cités. Enfin, je décrirai la démarche scientifique suivie dans cette thèse, les principaux défis et les questions posées.

I. Le plancton, acteur majeur de l'écosystème marin

I.1. Définition du plancton

Le mot plancton provient de *planktós* en grec ancien et signifie « errant ». En 1887, le zoologiste allemand Victor Hensen donne sa première définition scientifique au plancton en reprenant le terme et le définissant comme « tout ce qui flotte dans l'eau »¹. On peut citer une définition plus moderne du plancton comme « l'ensemble des organismes qui vivent en suspension dans les eaux douces, saumâtres ou salées, et sont inaptes à lutter contre les courants »². D'un point de vue de l'évolution il forme ainsi un groupe dit polyphylétique c'est-à-dire qu'il regroupe un ensemble d'organismes excluant leur ancêtre commun le plus récent. Le plancton est ainsi défini par une contrainte physique auquel est soumis l'ensemble des organismes le composant et non par un lien de parenté évolutif. Il regroupe ainsi un grands

nombre d'espèces et d'une richesse extrêmement variée de par leur taille, leur physiologie, leur niche écologique, leur forme, mais aussi leur position dans l'arbre du vivant. On distingue ainsi les grandes classes de taille de plancton pour la première fois en 1980³, qui n'incluaient pas les virus à l'époque. Les différentes classes de taille se différencient ainsi par leur contenu en type d'organisme aux modes trophiques variées, allant du femtoplankton (<2 µm) regroupant essentiellement les virus jusqu'au mégaplankton (20-200 cm) où l'on trouve majoritairement les méduses ou encore les colonies de salpes. Aux tailles intermédiaires, on retrouve les bactéries (picoplankton, 0.2-2 µm) essentiellement hétérotrophes, souvent parasitaires mais regroupant aussi les cyanobactéries phototrophes. Les protistes ou eucaryotes unicellulaires composés essentiellement d'algues photosynthétiques mais aussi de nombreux parasites recouvrent un plus large spectre de taille allant du nanoplankton (2-20 µm) jusqu'au mésoplankton (20 µm- 2 cm) en passant par le microplankton (20-200 µm). Enfin, le zooplankton, constitué essentiellement de petits métazoaires hétérotrophes tels que les copépodes se retrouve essentiellement dans les fractions de taille du microplankton et du mésoplankton. La figure 1 présente un aperçu de la diversité du plancton en taille, en forme et en type d'organisme. On verra plus tard que la taille du plancton est une caractéristique critique dans la définition de ses différentes niches écologiques, ses modes de vie et dans ses rôles écologiques. Le découpage en fraction de taille d'organismes reste néanmoins une simplification puisqu'une même espèce peut se retrouver sous différentes tailles selon le moment du cycle de vie (par exemple les copépodites, copépodes juvénils). Il se pose tout d'abord la question de l'importance écologique du plancton dans nos écosystèmes et en particulier l'écosystème marin.

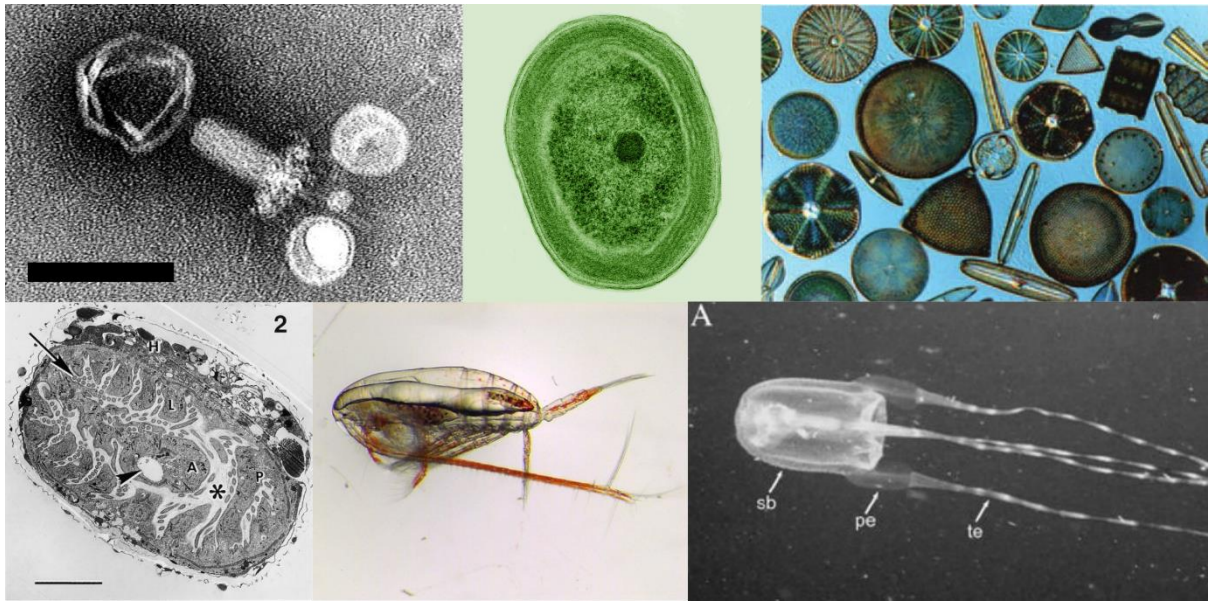


Figure 1 | La diversité du plancton à travers quelques exemples. Haut : un virus de cyanobactérie (bactériophage) (issue de Sullivan *et al.*⁴, © 2005 Sullivan et al.), une cyanobactérie (*Prochlorococcus*) (issue de Alsante *et al.*⁵) et une collection d'algues unicellulaires (diatomées) (issue de Korunic 2013⁶, © 2021 Alsante). Bas : le parasite *Amoebophrya sp* au sein de son hôte *Prorocentrum minimum* (issue de Maranda 2001⁷, Copyright © 1999-2022 John Wiley & Sons, Inc., reproduit avec la permission de John Wiley & Sons), un copépode (*Calanus Finmarchicus*) (issue de Hansen *et al.*⁸, copyright © 2016 reproduit avec la permission de Informa UK Limited, commercialisant sous le nom de Taylor & Francis Group) et une méduse (issue de Bentlage *et al.*⁹, copyright © 2012 reproduit avec la permission de Informa UK Limited, commercialisant sous le nom de Taylor & Francis Group).

I.2. Position écologique du plancton

Une grande partie des organismes planctoniques, appelés phytoplancton, est photosynthétique c'est-à-dire qu'ils ont un mode de vie phototrophe et sont ainsi capable de convertir l'énergie lumineuse en énergie chimique leur permettant de biosynthétiser des molécules organiques constituant leur source énergétique (utilisées via la respiration). Ces organismes sont ainsi appelés « producteurs primaires » aussi dit autotrophes¹⁰ dans le sens où ils convertissent la matière minérale (CO₂ en général, bien qu'il existe aussi des organismes sulforéducteurs) en

matière organique en la réduisant via la chaîne d'oxydoréduction de la photosynthèse. Le phytoplancton est ainsi à l'origine de la matière organique disponible pour le reste du réseau trophique marin, allant du zooplancton, brouteur direct du phytoplancton, en passant par les poissons et jusqu'à l'homme¹¹. Plus précisément la Production Primaire Nette (PPN) définit la quantité d'énergie disponible pour les niveaux trophiques suivants. La PPN est ainsi définie comme la différence entre l'énergie accumulée par la photosynthèse (Production Primaire Brute, PPB) et l'énergie utilisée par la respiration des organismes photosynthétiques :

$$PPN = PPB - \text{respiration}$$

Bien qu'en termes de biomasse le phytoplancton ne représente que 1 à 2% de la biomasse végétale terrestre, il est responsable de 75% de la production primaire océanique¹² (en excluant les macroalgues et herbiers), quasi-équivalente (45 % de la PPN totale) à la production primaire terrestre¹³. Ce faible rapport entre biomasse et production primaire indique un renouvellement très rapide de la biomasse planctonique (cycle de vie de l'ordre de 1 jour pour une cellule) en comparaison du renouvellement de la biomasse terrestre (des centaines d'années pour les forêts). La PPN peut être limitée par de nombreux facteurs comme l'apport en nutriments ou la lumière, ce que je détaille plus loin. La photosynthèse a lieu dans la zone photique de l'océan (0-100m) où le phytoplancton reçoit suffisamment de lumière. En général plus en profondeur, certains bactérioplanctons jouent de manière importante le rôle de reminéralisateur de la matière organique provenant entre autres des cellules mortes de phytoplancton, zooplancton et des fèces du zooplancton. Cette matière reminéralisée peut ainsi revenir dans le réseau trophique en étant remontée vers la zone photique par des remontées d'eaux profondes (upwelling). Par des mécanismes d'agrégation physique, les particules de plancton mort peuvent aussi sombrer vers les fonds océaniques participant ainsi à la pompe à carbone biologique responsable de l'enfouissement de matière organique dans les sédiments. Cette matière est en particulier à l'origine du pétrole et autres gaz naturels (le charbon est lui d'origine végétale terrestre) que les hommes utilisent de manière abusive notamment pour se

fournir en énergie, pour la chimie, la fabrication de nombreux matériaux (plastiques) et à l'origine du changement climatique. Le plancton joue donc un rôle primordial dans le cycle du carbone en particulier via la photosynthèse, responsable de la production primaire, alimentant la chaîne alimentaire marine mais aussi la pompe à carbone biologique à l'origine de la séquestration de carbone sous forme d'hydrocarbures sur des échelles de temps longs (millions d'années). Avant de détailler plus précisément la place du plancton dans les cycles biogéochimiques (Section IV.1 La place du plancton dans les grands cycles biogéochimiques) et le climat, j'expliquerai comment on étudie le plancton aujourd'hui en particulier en décrivant les modes principaux modes d'observations du plancton avec un approfondissement sur les récentes méthodes dites « métaomiques » de séquençage de l'ADN (Acide DésoxyriboNucléique) utilisé dans les campagnes *Tara Océans*.

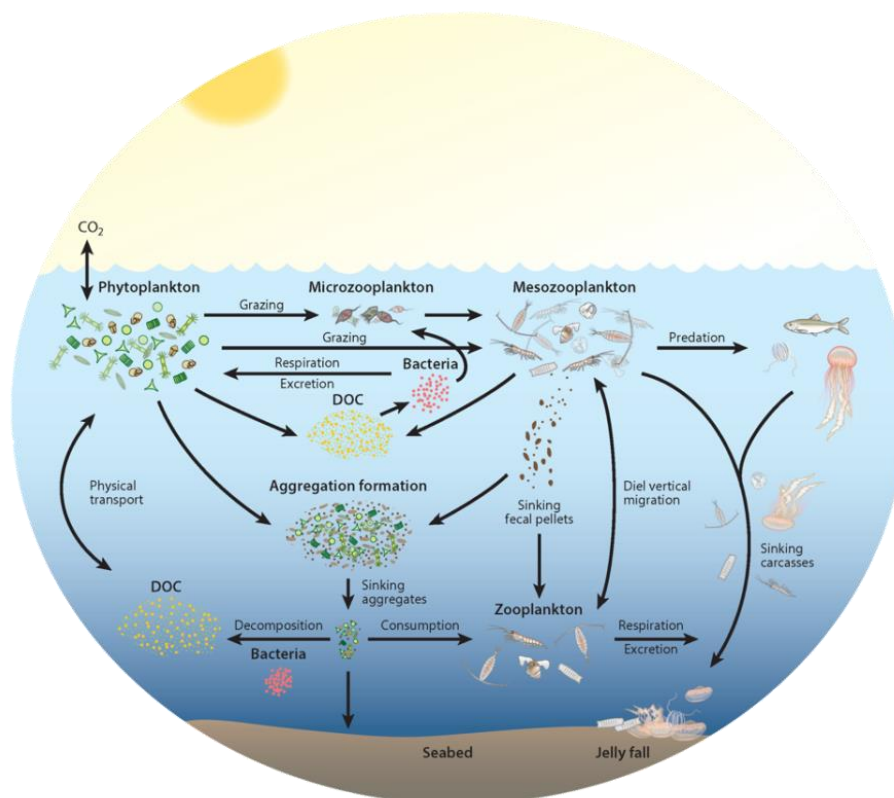


Figure 2 | Le plancton dans l'écosystème marin¹⁴ : importance dans le cycle du carbone. La figure est issue de Steinberg *et al.*¹⁴, Reproduit avec la permission de Annual Reviews.

II. La planctologie, les expéditions *Tara* Océans et les méthodes métagénomiques de séquençage de l'ADN

II.1. Bref historique de la planctologie

On considère que les premiers échantillonnages de plancton datent du XIX^{ème} siècle en particulier par l'échantillonnage de zooplancton à l'aide d'un filet conçu pour prélever des larves de crustacés, inventé en 1828 par John Vaughan Thompson, planctologiste précurseur^{15,16}. Il met notamment en évidence que *Zoea* est en réalité un stade larvaire des décapodes et non un genre à part entière. Dans la suite du siècle, l'usage de filet pour capturer le zooplancton est généralisé. Notamment sur le *Beagle* (1839), Darwin utilise un petit filet et en 1844 Müller utilise un filet conique à petites mailles pour capturer les petits organismes¹⁷. Au cours du siècle les innovations se succèdent (Figure 3) jusqu'à l'innovation du filet à fermeture Nansen permettant le prélèvement à des profondeurs voulues. En fin de siècle (1895), Victor Hensen développe notamment le filet Hensen et révolutionne le domaine en introduisant une méthodologie quantitative à la fois par la méthode de prélèvement via son filet et les questions posées. Hensen introduit ainsi la nécessité de connaître le volume d'eau prélevé afin de connaître le nombre d'organisme prélevé et attache une importance particulière à l'estimation des erreurs de prélèvement (volume d'eau, comptage des organismes)¹⁵. Il postule aussi que la taille des organismes prélevés dépendra de la gaze du filet utilisé¹⁵. En 1888, Hensen, accompagné de 5 scientifiques, parcourt ainsi les mers à bord du bateau à vapeur *National* parcourant 16,000 miles et les principales zones biogéographiques de l'Atlantique Nord¹⁸. Hensen pose ainsi certaines grandes questions de la planctologie moderne telle que la question du contenu en organismes planctoniques, sa quantification mais aussi ses variations saisonnières et annuelles. L'expédition de plus grande ampleur du XIX^{ème} siècle est celle du H.M.S. (Her Majesty Ship) *Challenger* de 1872 à 1876 parcourant 70,000 miles à travers tous les océans de la planète à l'exception des océans polaires et prélevant aux alentours de 13,000 échantillons prélevés à l'aide de filets (Figure 4). Le naturaliste allemand Ernst Haeckel, contemporain de Hensen, réalise notamment

de nombreux dessins des organismes planctoniques prélevés comme des radiolaires, des méduses ou autres siphonophores. Ainsi l'expédition a permis de répertorier plus de 4,700 nouvelles espèces.

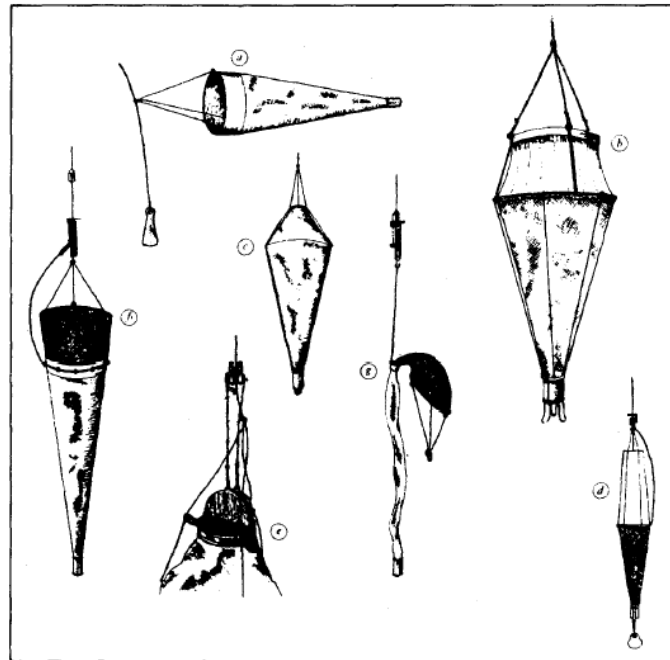


Figure 1

- a Simple conical tow-net.
- b Hensen net.
- c Apstein net.
- d Juday net.
- e Apstein net with semi-circular closing lids.
- f Nansen closing net open.
- g Nansen net closed.

Figure 3 | Les filets à plancton du début du XIX^{ème} siècle au début du XX^{ème} siècle¹⁷. Figure issue de Caspers 1980¹⁷, Copyright © 1980 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

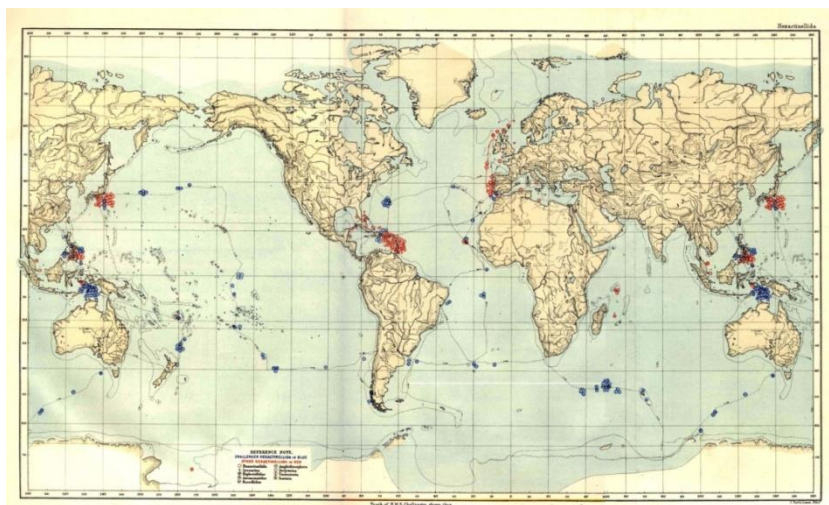


Figure 4 | L'expédition du H.M.S. Challenger de 1872 à 1876¹⁹. Carte issue de C. E. et al.¹⁹.

La fin de siècle est aussi marquée par le prélèvement d'organismes phytoplanctoniques plus petits à l'aide de filets en soie à l'initiative de Per Teodor Cleve²⁰ qui s'intéresse particulièrement aux algues dont les diatomées côtières et des eaux douces²¹. Il participe notamment à une expédition suédoise à Spitsbergen en 1898 suite à laquelle il décrit plusieurs nouvelles espèces de Radiolaires²². Outre les grandes expéditions océanographiques, l'idée d'un réseau international de stations d'observations marines côtières est émise par le zoologiste allemand Anton Dohrn et la création, par ce dernier, de la Stazione Zoologica à Naples en 1872 permet de participer aussi activement à la recherche sur le plancton²³.

Tout au long du XX^{ème} siècle, l'amélioration des filets et des techniques d'observation du plancton se poursuit avec notamment le développement des techniques optiques (caméra-télévision couplé, le Virtual Plankton Recorder (VPR), l'electronic zooplankton counting device (ECZD) ou encore les premières versions de l'Underwater Vision Profiler (UVP)), du Continuous Plankton Recorder (CPR, Figure 5) ou encore de la bouteille Niskin en 1966 (encore utilisée aujourd'hui) amélioration de la bouteille Nansen¹⁵. A noter que les données issues du CPR sont encore utilisées et exploitées aujourd'hui constituant une base de données compilant 60 ans d'échantillonnage²⁴. En parallèle des innovations liées à l'échantillonnage du plancton, les instruments liés à la mesure des paramètres environnementaux se développent allant du simple thermomètre positionné dans une capsule métallique pour le protéger de la pression (*H.M.S. Challenger*)¹⁹ au Conductivity, Temperature, and Depth profiler (CTD) permettant de réaliser des profils sur des milliers de mètres en profondeur de température, pH, densité ou encore salinité (via la mesure de la conductivité)¹⁵. En effet, les avancées technologiques s'accompagnent du développement et de la guerre des idées avec notamment l'émergence de l'importance des paramètres environnementaux dans l'écologie du plancton et de la nécessité de comprendre le plancton non en tant que tel mais dans son écosystème.

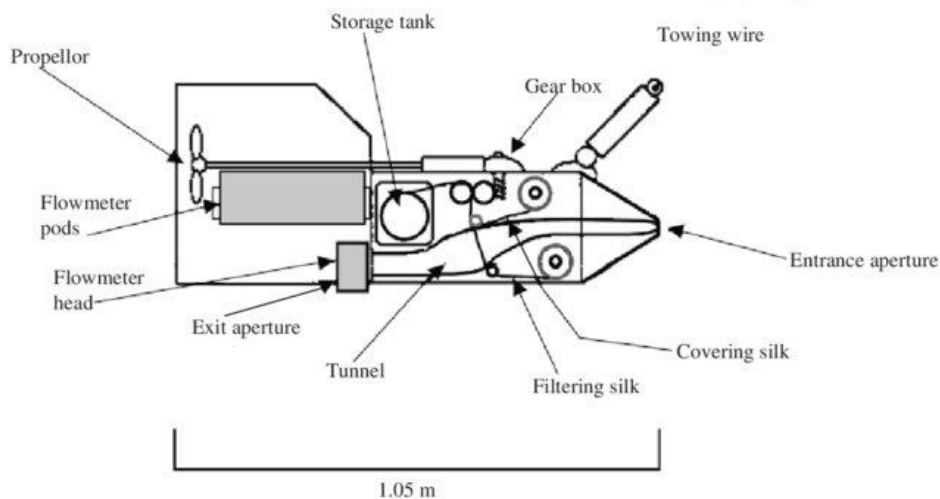


Figure 5 | Schéma du Continuous Plankton Recorder²⁵ (CPR). Figure issue de Jonas *et al.*²⁵. © Oxford University Press 2004, reproduit avec l'autorisation de Oxford University Press.

Du point de vue des interprétations scientifiques, la fin du XIX^{ème} est ainsi marquée par la controverse avec l'affrontement des théories de Hensen et de Haeckel et perdure jusqu'à la moitié du XX^{ème} siècle. Haeckel attaque les travaux de Hensen suite à son expédition sur le *National*¹⁸. Haeckel critique notamment l'approche statistique et quantitative de Hensen. Ce dernier décrit ainsi des océans tropicaux pauvres en plancton par son approche quantitative tandis qu'à l'époque la notion de tropiques riches comparés aux plus hautes latitudes était déjà bien ancrée notamment dans les approches décrivant la biodiversité terrestre. Certaines zones tropicales font aussi exception aux observations de Hensen. Ainsi en 1923, la « présumée carence » de plancton dans les « géniales eaux chaudes des tropiques » est encore reconnue comme controversée par Herdman¹⁵. Il faudra attendre la compréhension du lien entre plancton, nutriments et remontées d'eau froide (upwelling) notamment avec les travaux théoriques d'Albert Defant²⁶ (1936) et de Harald Sverdrup (1938²⁷, 1941²⁸) donnant une explication mécanistique du phénomène. Haeckel croit aussi en une distribution non-uniforme à la fois spatialement (horizontalement et verticalement) et temporellement ; il critique ainsi l'approche quantitative de Hensen¹⁵. Les filets à plancton de l'époque ne permettent ainsi que de conclure à une distribution aléatoire à travers la plupart des régions océaniques du plancton à un même endroit.

L'utilisation de pompes à plancton moins sujettes aux erreurs notamment par Barnes²⁹ (1949) et jusqu'à Hopkins³⁰ (1963) permettent de conclure à une distribution non aléatoire mais inégale (« patchy ») du zooplancton. Cette distribution reste encore largement incomprise aujourd'hui et reste une question importante de l'océanographie contemporaine que je discuterai plus amplement dans des sections ultérieures (III.1. Le contexte du seascape et III.2. La circulation océanique : le rôle des courants).

A l'orée du XXI^{ème} siècle, la biologie moléculaire et le séquençage de l'ADN ont émergé, les techniques optiques ne cessent de s'améliorer, l'imagerie et le suivi satellitaire sont en plein essor mais il naît aussi un nouvel attrait pour les grandes expéditions et la compréhension du climat et du changement climatique. Toutes ces avancées permettent ainsi une nouvelle révolution dans l'étude et la compréhension de la biologie du plancton.

II.3. Expéditions et techniques modernes

La biologie moléculaire consiste en la compréhension des mécanismes de fonctionnement de la cellule à l'échelle des molécules composant celle-ci. Elle apparaît au XX^{ème} siècle et se développe fortement dans sa seconde moitié avec notamment la découverte de la structure de l'ADN, support de l'information génétique, par Rosalind Franklin, Watson et Crick en 1953^{31,32}. L'ADN consiste en une suite de bases nucléiques azotées structurées en double hélice que sont l'adénine (A), la thymine (T), la guanine (G) et la cytosine (C), ATGC ! Ce code à 4 lettres permet le maintien de l'ensemble de l'information contenue dans les organismes vivants par transformation de cette information purement chimique (la séquence nucléique) en de multiples autres types d'information biologiques (nerveuse, métabolique, signalisation cellulaire...). Le génome d'un organisme consiste ainsi simplement en une suite non aléatoire des lettres A, T, G et C, par exemple : AGCTTTTCATTCTGACTGCAACGGC...quelques millions de lettres plus tard...AACGCCTTAGTAAGTATTTTC. Il s'agit ici du début et de la fin du chromosome circulaire d'une souche d'*Escherichia coli* débutant à l'oriT (origine de transfert lors

de la conjugaison). Mais comment peut-on connaître cette suite de lettres ? En 1977, Frédérick Sanger développe une technique de séquençage de l'ADN³³ appelé technique Sanger permettant de « lire » cette information génétique (la séquence des lettres A, T, G et C) et qui devient la principale technique utilisée pendant près de 40 ans. En 1996, le premier échantillon d'ADN, provenant d'eau de mer de Hawaïi³⁴, ciblant les séquences d'un ensemble d'organismes dans leur écosystème et non un organisme donné est séquencé. C'est l'approche dite « méta ». C'est à la croisée de la biologie moléculaire, du séquençage de l'ADN, de l'approche méta et de la planctologie que l'expédition *Global Ocean Sampling* (GOS) débute en 2003 (Figure 6). L'expédition, privée, est menée sur le *Sorcerer II*³⁵, yacht personnel de Craig Venter et les données sont analysées par son laboratoire, le *Venter Institute*. L'expédition a pour but d'explorer la diversité microbienne génomique des océans par approche de metabarcoding (II.4. Les révolutions omiques), ciblant spécifiquement les séquences d'ADN codant les ARNs ribosomiaux permettant de caractériser l'appartenance phylogénétique des organismes présents. Un effort métagénomique de shotgun sequencing consistant à séquencer l'ADN de manière aléatoire par petits fragments est aussi réalisé (II.4. Les révolutions omiques). Une question centrale de l'expédition est le lien entre plancton et contexte environnemental mais aussi la biochimie des organismes présents. Au total, l'expédition réalise des échantillonnages des eaux de surface à travers 41 sites différents allant de la Nouvelle-Écosse à la Polynésie française et couvrant ainsi plusieurs océans et climats. L'effort de séquençage, c'est-à-dire le nombre de bases azotées de l'ADN séquencées (A, T, G et C) par l'expédition s'élève à 6.3 milliards à travers 7.7 millions de lectures (courts fragments séquencés d'ADN). L'exploitation de ces séquences a montré qu'une grande majorité de l'ADN séquencé était alors jusque-là inconnu des bases de données³⁵, issu majoritairement à l'époque d'une poignée d'organismes cultivés. L'effort de séquençage permet notamment d'étendre la collection des familles de protéines connues alors³⁶.

L'importante base de données générée par l'expédition a aussi permis de mettre en évidence un lien entre la composition des communautés planctoniques, leur contenu

fonctionnel en gènes et les facteurs environnementaux^{37,38}. A la suite du *Global Ocean Sampling*, de nouvelles grandes expéditions sont lancées telles que les expéditions *Tara Océans* (2009-2013)³⁹ à l'initiative du biologiste cellulaire Eric Karsenti ou encore l'expédition *Malaspina*(2010)⁴⁰ menée par le conseil national de la recherche scientifique espagnole. L'expédition *Tara Oceans*, dont les données sont exploitées dans cette thèse sera plus amplement présentée dans la section suivante (II.3. Les expéditions *Tara Océans*).

En parallèle des grandes expéditions échantillonnant directement le plancton, les techniques d'imagerie satellitaire, les mesures environnementales *in situ* mais aussi la modélisation du plancton dans la machine climatique se sont largement développées. L'imagerie satellitaire permet ainsi de mesurer à grande échelle la température de surface de l'océan⁴¹, mais aussi de quantifier la chlorophylle du phytoplancton et d'autres pigments⁴². Les satellites permettent aussi d'instruire les expéditions océaniques en repérant des zones d'intérêt tel que les blooms planctoniques⁴³ ou encore des structures océaniques intéressantes telles que les larges tourbillons de l'atlantique sud généré par le courant des Aiguilles où les communautés de plancton peuvent être transportées pendant de longues durées⁴⁴. D'autre part le système Argo⁴⁵ dont l'idée remonte à une conférence en 1999 menée par Dean Roemmich possède une flotte d'un peu moins de 4000 « argo floats » permettant la collecte permanente de différentes données environnementales telles que la température ou la salinité à la fois en surface et en profondeur ou encore de données biogéochimiques avec l'extension Biogeochemical Argo⁴⁶. Ces mesures permettent entre autres par des approches statistiques de reconstruire des champs des différentes variables environnementales sur l'ensemble des océans et sur des périodes longues comme par exemple avec les différentes climatologies du World Ocean Atlas⁴⁷⁻⁵⁰ géré par la NOAA (National Oceanic and Atmospheric Administration). Ces données biogéochimiques sont ainsi cruciales pour comprendre les liens entre les communautés de plancton et leur environnement. Enfin, le plancton est aussi étudié via la modélisation et notamment au sein des modèles

systèmes Terre.

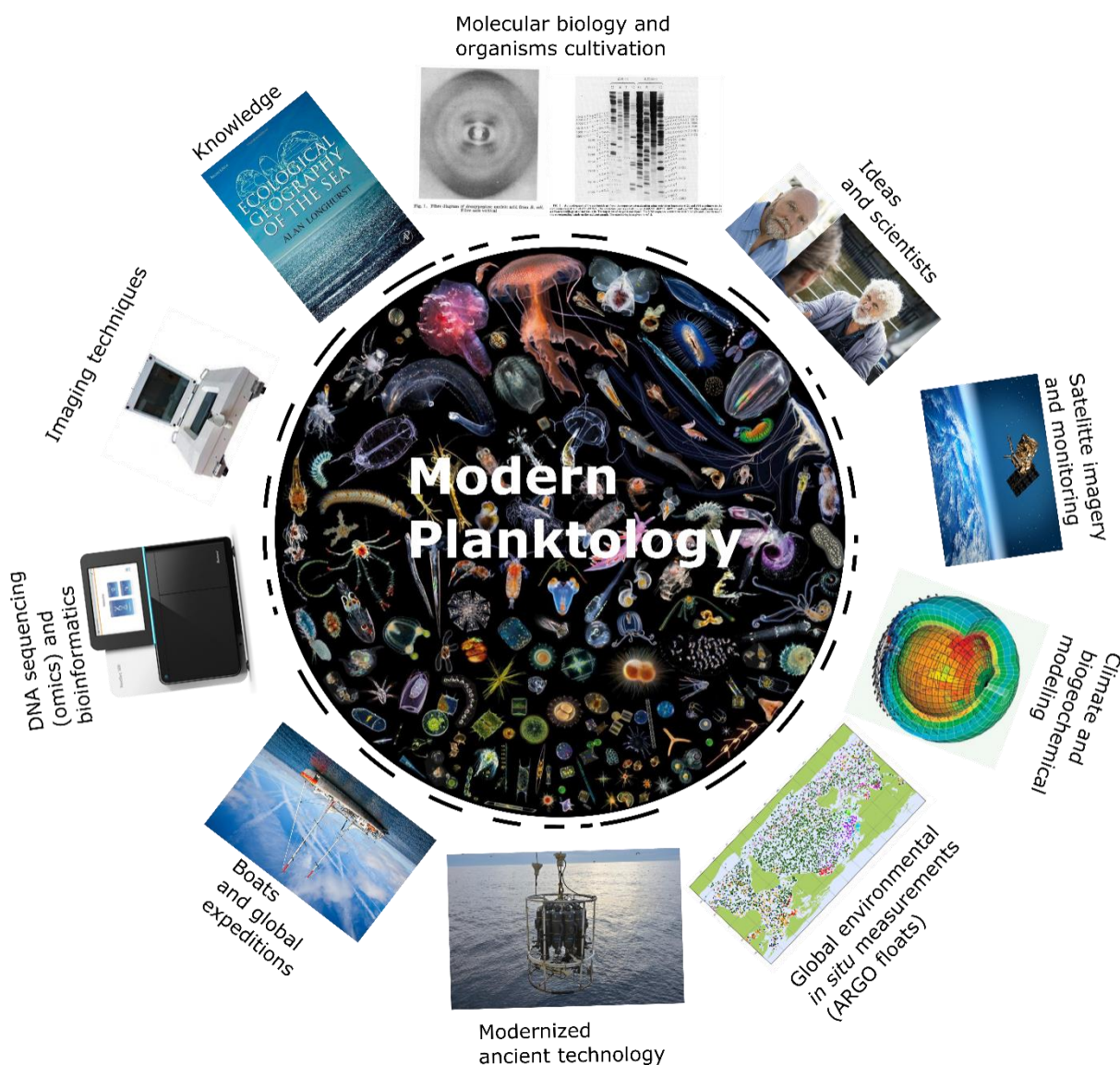


Figure 6 | Les principaux ingrédients de la planctologie moderne. En partant du haut puis dans le sens des aiguilles d’une montre : Images issues de Watson *et al.*³¹, Sanger *et al.*³³, Gross 2007⁵¹, embl.org⁵²(Credit: EMBL/Massimo del Prete), flickr.com⁵³ (Crédit : ESA), ethz.ch (Carbones EU Project, CC⁵⁴), Hjfreeland, CC BY-SA 4.0⁵⁵, ifremer.fr⁵⁶, Yohann.cordelle CC BY 3.0⁵⁷, DataBase Center for Life Science (DBCLS) CC BY 4.0⁵⁸, embrc-france.fr, Longhurst 2007⁵⁹. Image centrale : fournie par Christian Sardet et reproduit avec sa permission.

Ainsi, l’Institut Pierre Simon Laplace développe depuis 1995 la série de modèles système Terre « IPSL-CM » qui comprennent notamment une représentation du plancton via le modèle biogéochimique PISCES⁶⁰ (Pelagic Interactions Scheme for

Carbon and Ecosystems Studies). Ces modèles permettent notamment de projeter les conditions climatiques futures selon différents scénarios d'émission de gaz à effet de serre dans le contexte de la crise climatique⁶¹. Je reviendrai plus amplement sur les modèles climatiques dans la partie IV. Les modèles climatiques sont souvent utilisés en lien avec des modèles de « niche écologique » dite statistique ou réalisée, concept sur lequel je reviendrai dans les différents chapitres, mais aussi de niche écologique dite fondamentale⁶². Ces modèles permettent ainsi de projeter les aires de vie des espèces planctoniques et les conséquences du réchauffement climatique sur celles-ci^{24,63-67}.

II.3. Les expéditions *Tara Océans*

Le 5 septembre 2009, la goélette *Tara*, propriété d'Etienne Bourgois, directeur général de la marque de vêtement *Agnès B.*, s'éloigne de son port d'attache de Lorient en France. Elle part pour une expédition d'envergure qui la mènera dans pratiquement tous les océans du monde entre 2009 et 2012 puis après quelques mois d'arrêt elle repart en 2013 pour un tour du cercle polaire arctique. L'expédition est unique par son envergure à la fois géographique, réalisant des échantillonnages à travers 210 sites océaniques (Figure 7), et par l'effort de séquençage d'ADN réalisé⁶⁸. La goélette *Tara*, qui donne son nom à la fondation *Tara Expéditions* est un navire au passé riche, d'abord appelé *Antarctica* par l'explorateur Jean-Louis Etienne en 1989 puis *Seamaster* par l'explorateur Peter Blake engagé dans la défense de l'environnement qui mourut à son bord en 2001 lors d'une mission sur le fleuve Amazone, attaqué par des pirates⁶⁹. La goélette est un voilier à 2 mats de 36 m de longs et 10 m de large avec une coque en aluminium en forme d'ogive conçue spécialement pour la dérive sur la banquise et résistant ainsi aux basses températures et à la compression par les glaces (Figure 8).

Pour l'échantillonnage et le séquençage du plancton, l'expédition *TaraOceans* adopte une stratégie sans précédent en prélevant l'ensemble des micro-organismes de 0 à 2 mm à travers de multiples fractions de tailles principales englobant les principales familles de plancton. L'expédition a en effet pour objectif d'étudier

l'écosystème planctonique et sa diversité de la manière la plus exhaustive possible^{39,70}. On peut citer Eric Karsenti sur la genèse de ce projet. Il écrit dans son ouvrage *Aux sources de la vie*⁷¹ :

« J'ai passé quelques années d'errance à chercher comment faire une telle expédition, à trouver un voilier — pour finir une goélette de 36 m, Tara — et des financements, et à chercher un thème moderne qui se rattacherait à Darwin et à la théorie de l'évolution. Cette phase a représenté un énorme travail. La première version du projet, « Évolution 2009 », n'a jamais abouti. Tout s'est débloqué au printemps 2007, quand j'ai décidé d'élaborer un vrai projet scientifique associé à l'expédition. Avec Christian Sardet, biologiste cellulaire comme moi, j'ai pensé à étudier le plancton, ce monde de nos origines justement composé de virus, de bactéries,... ».

Pour simplifier, on peut retenir 6 principales fractions de taille échantillonnées par l'expédition. La fraction de taille 0-0.2 μm est enrichie en virus, la 0.22-3 μm est enrichie en procaryotes (bactéries et archées), les fractions 0.8-5 et 5-20 sont enrichies en protistes (algues...) et enfin les fractions de taille 20-180 et 180-2000 μm sont enrichies en petits métazoaires (copépodes) (Figure 9). Les échantillons sont collectés à différentes profondeurs par des bouteilles Niskin placées sur une rosette avec des filets de différentes tailles de mailles. Les profondeurs échantillonnées sont principalement la subsurface (5-10 m), la couche de maximum de chlorophylle (Deep Chlorophyll Maximum en anglais), déterminée en amont (20-100 m) où la concentration en chlorophylle est maximale et les zones mésopélagiques (300-400 m) où la lumière est quasiment absente et constituant souvent des zones minimum d'oxygène⁷² (Oxygen Minimum Zone, OMZ en anglais). Après la récolte des échantillons d'eau, celle-ci est passée dans des tamis de plus en plus fins permettant de séparer les organismes des différentes fractions de taille. Au total les scientifiques à bord de la goélette ont récolté 27 000 échantillons dont la moitié a été utilisée pour les analyses génomiques de composition de la communauté planctonique en gènes, espèces et leurs fonctions métaboliques. Les méthodes génomiques utilisées dans l'expédition seront décrites plus en détail dans la section suivante (II.4. Les révolutions omiques). L'autre moitié des échantillons a servi à des techniques

d'imagerie plus classiques comme le ZooScan⁷³, FlowCam⁷⁴, la cytométrie en flux⁷⁵ (CF) et l'Underwater Vision Profiler (UVP)⁷⁶ (Figure 10). Brièvement, le ZooScan est utilisé pour quantifier et visualiser le zooplancton.

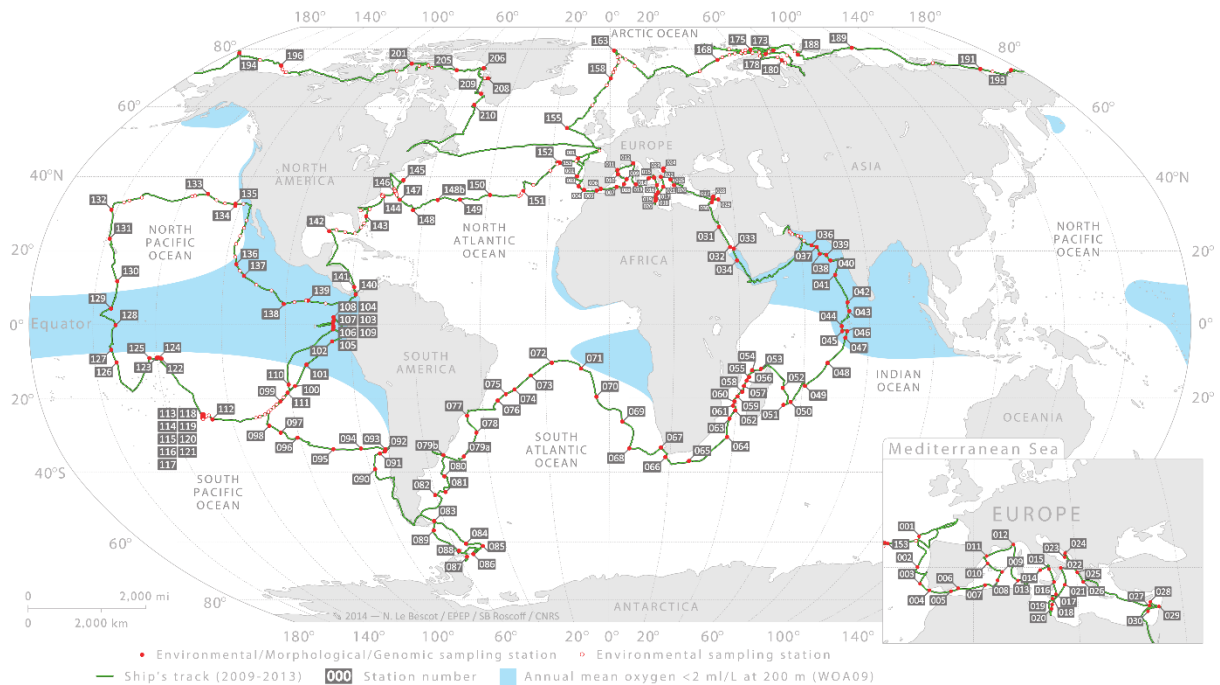


Figure 7 | Le parcours de la goélette *Tara* lors des expéditions *Tara Océans* (2009-2012) et *Tara Polar Circle* (2013). Figure issue de Pesant *et al.*⁷⁷.

Il permet ainsi l'identification taxonomique plus ou moins précise des organismes zooplanctoniques via des techniques de machine-learning supervisés⁷⁸⁻⁸⁰. Le FlowCam permet lui une analyse semi-automatique du microplancton⁸¹ par imagerie à haute résolution. La cytométrie en flux est une méthode de tri des cellules par leur taille et/ou par leur fluorescence naturelle (pigments) ou non (via la biologie moléculaire) (fluorométrie)⁸². Cette méthode permet notamment la quantification des plus petites cellules telle que les bien connues cyanobactéries photosynthétiques *Prochlorococcus* et *Synechococcus*⁸³ caractéristiques des zones tropicales peu riches en nutriments (oligotrophiques). Enfin l'UVP permet de quantifier la distribution des particules de zooplancton de plus de 1 mm⁷⁶. Un des intérêts de l'imagerie est la quantification de la biomasse ou encore du biovolume des espèces présentes. La concentration en ADN des échantillons a aussi été quantifiée. Une limitation de l'imagerie reste la résolution taxonomique à laquelle les espèces sont caractérisées

bien que le problème existe aussi dans tous les types d'approche génomique. Un des challenges important aujourd'hui est notamment la mise en évidence du lien entre les données issues de ces différentes méthodes. Il existe néanmoins quelques études comparant les deux approches et tentant de les réconcilier au niveau de la mesure de la richesse en espèces⁸⁴ ou bien de la bêta-diversité, c'est-à-dire la comparaison des échantillons les uns avec les autres⁸⁵. En revanche la correspondance exacte entre l'imagerie et la génomique est un réel défi bien que l'approche de séquençage dit « single cell » permette de séquencer des organismes précis d'un ou plusieurs échantillons donnés^{86,87}. L'effort de séquençage nécessaire à l'identification des génomes de l'ensemble des organismes d'un échantillon serait ainsi faramineux.



Figure 8 | La goélette *Tara* par temps gris. Crédit : Fondation Tara Océan.

En parallèle de l'échantillonnage du plancton, un certain nombre de données océanographiques environnementales sont aussi mesurées à bord de *Tara* et sont compilées avec certaines données de mesoéchelles (10-100 km) issues de relevés satellitaires, *in situ* ou calculées sont compilés dans des bases de données⁸⁸⁻⁹¹. Ainsi la goélette dispose d'une sonde CTD (Conductivity Temperature Density) permettant de réaliser des profils en profondeur de ces différents paramètres physiques. Ces

paramètres sont cruciaux pour comprendre l'écologie du plancton et permettent notamment de déterminer certaines caractéristiques de leur environnement physique. Le gradient de température peut notamment permettre de déterminer la profondeur de la couche de mélange (Mixed Layer Depth MLD), couche d'eau de surface pour laquelle on peut considérer comme relativement homogène un certain nombre de paramètres et déterminant de manière cruciale la quantité de lumière accessible aux organismes phytoplanctoniques⁹², discutée plus amplement dans la partie III.3. Les facteurs limitant la croissance du phytoplancton : température, taille et nutriments, courants et lumière. Les paramètres géochimiques tels que les concentrations en phosphate, nitrate ou encore silice dissous ont eux aussi été mesurés aux endroits précis (profondeur, latitude, longitude) de prélèvements des échantillons.

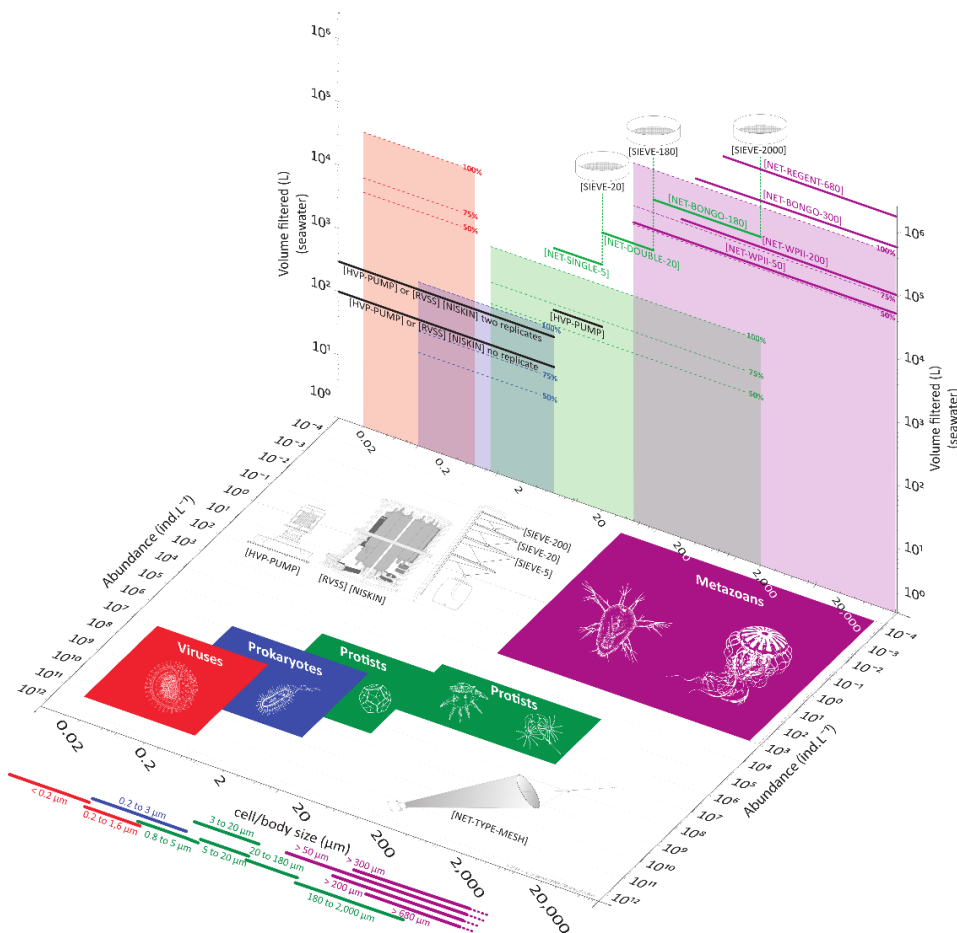


Figure 9 | Approche empirique d'échantillonnage du plancton par fraction de taille et choix des appareils d'échantillonnages⁷⁷. Figure issue de Pesant *et al.*⁷⁷.

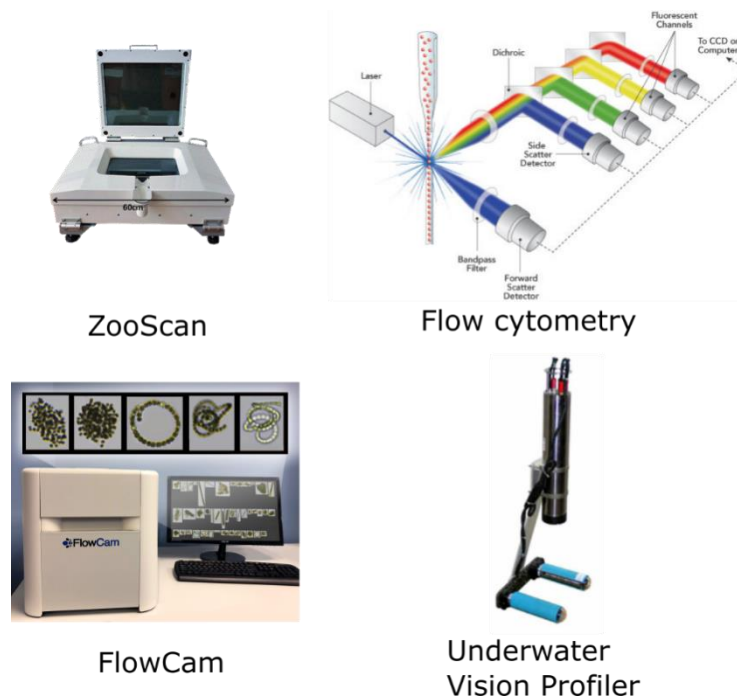


Figure 10 | Les techniques d'imageries utilisés pour les expéditions Tara

Océans.Zooscan : embrc-france.fr, Flow cytometry : IDEX Health & Science (reproduit avec permission), Flow Cam : Kierano, CC BY 3.0 (<https://commons.wikimedia.org/wiki/File:Cytometer.svg>) et Underwater Vision Profiler issu de Picheral *et al.*⁷⁶, © 2010, by the Association for the Sciences of Limnology and Oceanography, Inc..

Les expéditions de la goélette continuent après ces premiers voyages avec notamment les missions *Tara Méditerranée* (2014), *Tara Pacifique* (2016-2018), *Tara Microplastiques* (2014-2019) et aujourd'hui *Tara Microbiome* (2020-2022) aux objectifs scientifiques variés. Outre la collecte et l'expédition en elle-même, les prélèvements de la goélette ont permis d'acquérir un large panel de résultats scientifiques et les données continuent à être analysés aujourd'hui dans l'idée d'une compréhension toujours plus complète de l'écosystème marin⁷⁰. L'exploitation de ces données est réalisée à travers de nombreux laboratoires dans le monde formant des consortiums internationaux pour chaque expédition. Ainsi, le Genoscope à Evry en France (au sud de Paris), où j'ai effectué ma thèse est responsable du séquençage des échantillons des différentes expéditions de la goélette. Ainsi un effort de séquençage de l'ordre de 60 trillions de bases d'ADN a été réalisés pour séquencer

pas moins de 2800 échantillons de communautés planctoniques réalisés par la Goélette⁷⁰ à travers ses différentes expéditions (24 trillions pour *Tara Oceans*), effort sans précédent. Tous les types de données omiques, tant de metabarcoding, visant à décrire l'ensemble de la diversité en organismes que les données de métagénomiques, séquençant au hasard un certain nombre du contenu génomique des organismes présents, ont permis de mettre à jour une diversité inconnue tant en contenu en organismes qu'en contenu génétique. Je m'attacherai maintenant à décrire ces méthodes à travers le prisme des avancées qu'elles permettent tout d'abord dans la recherche en biologie en général puis plus précisément au sein des expéditions *Tara* et le cadre de l'étude de l'écosystème planctonique.

II.4. Les révolutions omiques et métaomiques: d'importantes avancées scientifiques

Comme brièvement abordé dans la partie II.1. Bref historique de la planctologie, les méthodes de séquençage de l'ADN ainsi que les techniques bioinformatiques de curation, d'assemblage puis d'exploitation statistique en aval des premières sont au cœur de l'écologie moderne mais aussi de nombreuses disciplines biologiques. Au début des années 2000, suite au séquençage du génome humain, de nouvelles techniques de séquençage dites de « haut débit » apparaissent et permettent d'accéder à un volume de séquençage bien plus important (tant d'un point de vue économique que temporel) qu'avec la technique de Sanger. L'idée d'utiliser des nucléotides labélisés par leur fluorescence, déjà utilisés dans la technique de Sanger, émerge au milieu des années 90 à l'université de Cambridge à l'initiative de Shankar Balasubramanian et David Klenerman⁹³. Elle donne plus tard naissance aux compagnies Solexa puis Illumina (qui rachète Solexa en 2007) qui devient alors la technologie/compagnie de séquençage monopolisant le marché. C'est ainsi qu'en 2005, un simple « run » de séquençage par biosynthèse (SBS) Solexa permet la génération de 3 millions de bases pour le séquençage du bactériophage phiX-174. La nouvelle technologie se distingue ainsi en parallélisant massivement l'utilisation de sondes ADN, support moléculaire du séquençage⁹³. Aujourd'hui un « run » d'Illumina (15 à 18h) permet de séquencer plusieurs terabases d'ADN. Avant la montée en

puissance des techniques dites NGS (Next Generation Sequencing) tel qu'Illumina (Figure 11), le séquençage d'ADN a notamment été utilisé dans le cadre du *Human Genome Project*(HGP) qui a pour but de séquencer l'ensemble du génome humain et regroupe un consortium international de 20 laboratoires. Le projet coutera 2.7 milliards de dollars et s'achèvera en 2003 avec un génome quasi-complet⁹⁴ mais complété seulement très récemment⁹⁵. Le séquençage du génome humain trouve ainsi des applications dans de nombreux champs de recherche tels que la recherche sur le cancer⁹⁶ avec par exemple le Cancer Genome Project⁹⁷, ou encore les recherches en évolution avec par exemple la publication très récente d'une « *généalogie unifiée de génomes anciens et modernes* » analysant 3601 séquences de génomes humains⁹⁸. Ainsi au cours des années 2000 de nombreux génomes sont séquencés avec notamment les génomes d'*Arabidopsis Thaliana*(2000⁹⁹), de la souris (2002¹⁰⁰) ou encore de la vigne (2007¹⁰¹) avec des applications dans la recherche fondamentale, le domaine médical ou encore l'agriculture.

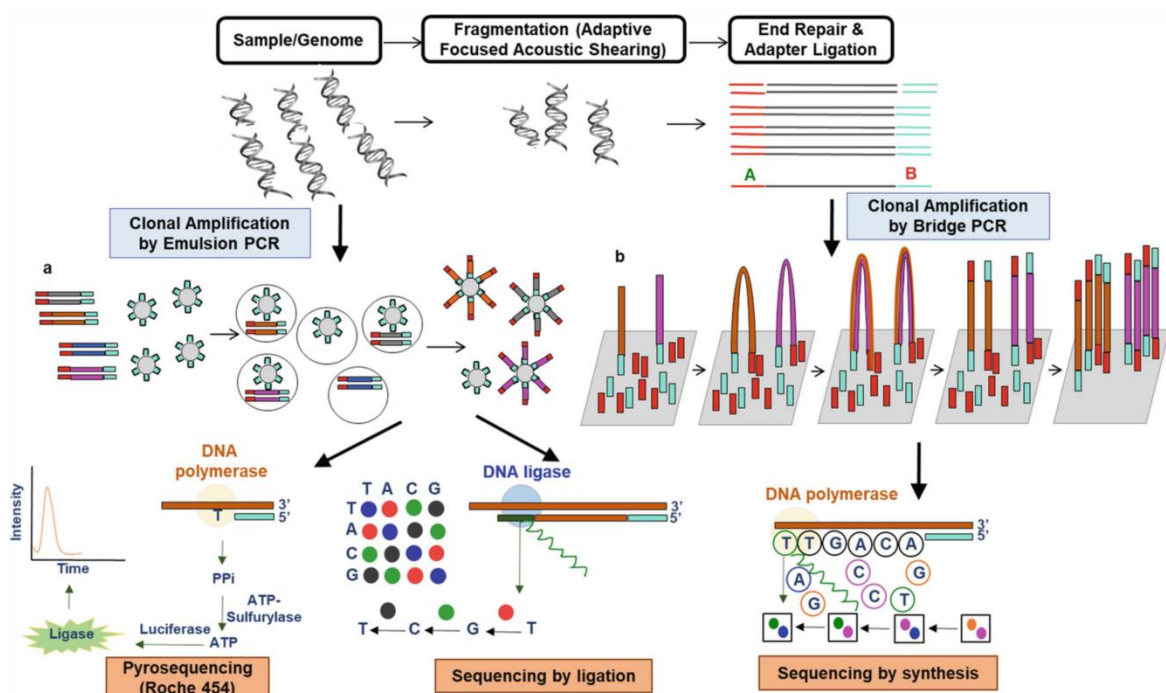


Figure 11 | Exemples de principes de NGS (Next Generation Sequencing) de l'ADN¹⁰². Figure issue de Gupta *et al.*¹⁰², reproduit avec permission Copyright © 2019, Springer Nature Singapore Pte Ltd. La technique utilisée par les séquenceurs Illumina est le séquençage par synthèse (en bas à droite).

En parallèle du séquençage d'organismes spécifiques générant des génomes dits de « références », le séquençage de l'ADN issu d'échantillons environnementaux se développe aussi fortement dans les années 2000. En principe les méthodes de séquençages (Illumina...) sont les mêmes mais il existe plusieurs approches pour séquencer ces échantillons. Les principaux échantillons environnementaux sont le microbiote humain¹⁰³, les océans (*Tara, Malaspina, GOS...*), fleuves et rivières¹⁰⁴, les sols¹⁰⁵ et plus récemment les eaux usées¹⁰⁶. On distingue principalement l'approche dite de metabarcoding¹⁰⁷ de l'approche de métagénomique¹⁰⁸. La première permet de cibler des parties spécifiques des génomes et permet de réaliser un inventaire des espèces présentes dans un échantillon. De manière complémentaire l'approche métagénomique ou « shotgun sequencing » séquence au hasard une partie du contenu génomique d'un échantillon environnemental. Je décrirai maintenant plus précisément ces deux techniques et leurs utilisations principalement dans les expéditions *Tara* puis l'approche métatranscriptomique permettant d'évaluer l'expression des gènes contenus dans un échantillon sous forme d'ARN (Acide Ribonucléique), produit de la transcription de l'ADN.

Le metabarcoding est une méthode largement utilisée dans le domaine de la génomique environnementale. A l'aide de petites séquences d'ADN appelées « amorces » (primers en anglais) dites « universelles » car relativement bien conservées (c'est à dire identiques) dans une catégorie d'organismes, les séquences d'ADN dites « variables » situées entre les amorces sont amplifiées puis séquencées (Figure 12). Ces séquences variables sont ainsi spécifiques d'un groupe taxonomique à un niveau de résolution phylogénétique variable. Il s'agit en général de l'ADN codant les ARN ribosomiques (rRNA) 18S (régions V4 et V9) pour les organismes eucaryotes et 16S pour les organismes procaryotes. Il s'agit donc d'une approche phylogénétique de la diversité planctonique que l'on peut trouver au sein d'un échantillon. L'abondance relative, notion sur laquelle je reviendrai, de différents groupes taxonomiques dans un échantillon peut ainsi être calculée. La résolution phylogénétique à laquelle une séquence est assignée correspond ainsi grossièrement

(il existe des sous rangs) au rang de la nomenclature linnéenne (1753)¹⁰⁹ auquel la séquence est assignée : royaume, classe, ordre, génus, espèce ou inférieur. Malgré une certaine variabilité dans cette résolution notamment selon les différents groupes (particulièrement les arthropodes dont font partis les copépodes¹¹⁰) l'approche de metabarcoding a pour mérite de pouvoir détecter les espèces rares et peu abondantes⁷⁰. Il est important de garder à l'esprit dans l'approche de metabarcoding que celle-ci peut être aussi biaisée par le set de primers utilisé privilégiant ainsi le séquençage de certains organismes et biaisant l'estimation de leur abondance¹¹¹ bien que certaines propositions aient été faites pour corriger ces biais¹¹².

Ainsi dans les expéditions *Tara*, l'approche de metabarcoding a mis en évidence une biodiversité planctonique insoupçonnée auparavant. En particulier, en cumulant l'ensemble des sites de prélèvement de l'expédition et les différentes fractions de taille et donc l'ensemble des lectures d'ADN issu du metabarcoding, le nombre d'OTU (Operational Taxonomic Unit) eucaryotes recensés s'élève à 150,000 (non saturé) dont environ un tiers environ n'ont pas pu être assignés taxonomiquement à un groupe connu¹¹³ (Figure 12).

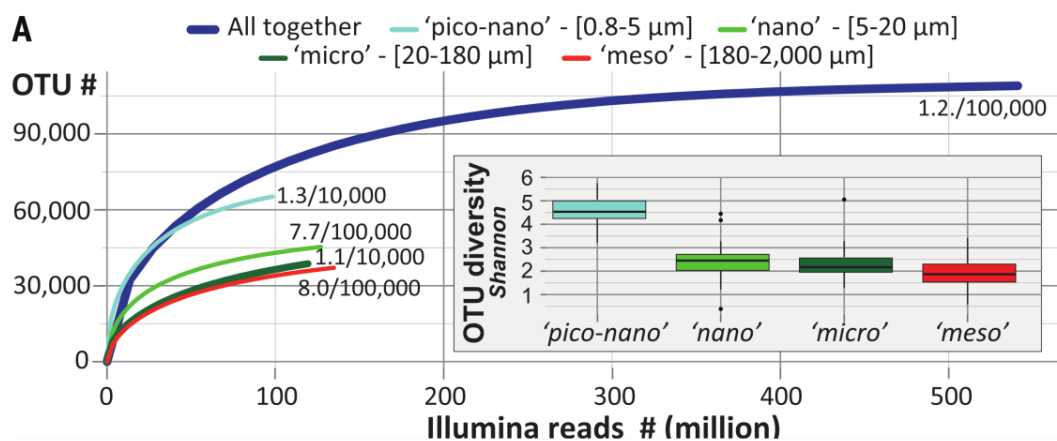


Figure 12 | Diversité ribosomique du plancton eucaryote de la zone photique¹¹³.

Figure issue de Vargas *et al.*¹¹³, Copyright © 2015, The American Association for the Advancement of Science.

Une telle diversité de plancton était alors largement insoupçonnée avant l'analyse de ces séquences. Les données de metabarcoding ont notamment ensuite été utilisées

pour étudier et modéliser la variabilité latitudinale de la biodiversité à travers les différents règnes du plancton. La température et la chlorophylle se sont ainsi avérées suffisantes pour expliquer statistiquement cette biodiversité¹¹⁴. Les projections sous réchauffement climatique montrent une potentielle diminution de cette biodiversité dans les différents groupes planctoniques^{84,114}. Les approches réseaux ont aussi largement été utilisées pour étudier les interactions entre planctons et facteur environnementaux¹¹⁵ et aussi leur vulnérabilité face à des changements environnementaux¹¹⁶. J'y reviendrai respectivement dans les parties III.2. La circulation océanique : le rôle des courants et IV.1 La place du plancton dans les grands cycles biogéochimiques et dans le système Terresur les contributions à des résultats de biogéographie et les liens entre plancton et pompe à carbone biologique.

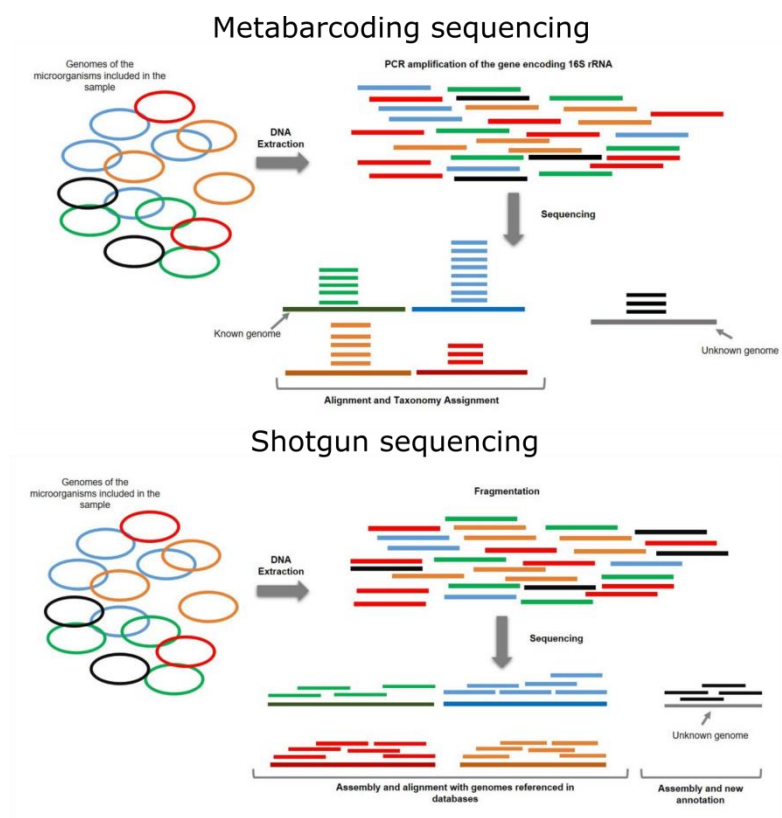


Figure 13 | La séquençage de l'ADN environnemental par metabarcoding¹¹⁷ (16S rRNA) (haut) et par shotgun¹¹⁸ (bas). Figures issues de france-genomique.org^{117,118}, reproduit avec la permission de France génomique.

Outre l'approche de metabarcoding, l'approche purement métagénomique de séquençage des communautés environnemental consiste à séquençer selon une

profondeur de séquençage (le nombre moyen qu'une même position de la séquence d'ADN a été séquencé) donnée une partie du contenu génomique de la communauté. Dans cette technique l'ADN environnemental est extrait puis coupé en petits brins et enfin séquencé par séquençage Illumina (taille des lectures : 150 paires de bases). Elle se distingue ainsi de l'approche metabarcoding par son côté aléatoire de séquençage de brin d'ADN d'où son nom de « shotgun sequencing » (Figure 12) et c'est cette même méthode qui est aujourd'hui utilisée la plus communément pour le séquençage de génomes classiques (c'est-à-dire de génomes uniques). Ainsi, elle vise l'ensemble des séquences de l'ADN de l'échantillon ce qui est colossal si l'on considère le nombre de génomes qu'un échantillon peut contenir (égal au nombre de cellules dans l'échantillon modulo les cellules purement clonales). Selon le budget et l'objectif de l'étude, la profondeur de séquençage c'est-à-dire le nombre de lectures totales d'ADN séquencé par échantillon peut être adapté^{119,120}. Ainsi, l'ADN séquencé sera principalement celui des organismes les plus abondants et des organismes aux grands génomes¹²¹. Il existe un effort récent regroupant les gènes de métagénomes provenant de divers habitats cumulant 303 millions de gènes bactériens non redondants¹²².

La métatranscriptomique, par approche de RNA-seq¹²³ permet de quantifier l'expression génétique de l'ADN d'une communauté, c'est-à-dire de quantifier les ARNs issus de la transcription de l'ADN et qui permettent ensuite la synthèse des protéines des organismes présents dans l'échantillon. Elle se distingue donc de la métagénomique par le fait qu'elle apporte une information sur l'état physiologique et métabolique de la communauté de plancton dans l'échantillon donné. L'exploitation des données métatranscriptomiques permet aussi de cibler plus facilement les gènes (car les données métagénomiques contiennent aussi les séquences non codantes) et a notamment permis l'assemblage d'important catalogues de gènes, séquences consensus au travers des différents échantillons prélevés. Ainsi, un catalogue de 40 millions de gènes procaryotes appelé Ocean Microbiome Reference Gene Catalog (OMR-GC)¹²⁴ puis de 47 millions¹²⁵ et un

catalogue de 116 millions de gènes eucaryotes¹²⁶ appelé *Marine Atlas of Tara Oceans Unigenes* sont publiés à la suite des expéditions *Tara Océans* (basé sur la métatranscriptomique, décrite ci-dessous). De manière importante, ces catalogues contiennent un grand nombre de gènes aux fonctions inconnues, ~60% de gènes aux fonctions inconnues¹²⁶ pour le catalogue eucaryote et 39% pour le catalogue procaryote¹²⁵.

Les collections de gènes ont ainsi par exemple été exploitées pour mieux comprendre la réponse des communautés aux limitations en fer^{126,127} ou encore de différencier certains métabolismes clés comme la photosynthèse, les métabolismes de l'azote¹²⁸ ou de la captation du carbone entre océans polaires et les autres océans¹²⁵. En utilisant la co-abondance des gènes à travers les stations, les gènes peuvent aussi être regroupés en « MetaGenomic-based Transcriptomes »¹²⁹ formant des unités plus ou moins cohérentes taxonomiquement permettant leur analyse fonctionnelle. Ainsi cette approche a permis l'identification d'organismes clés dans le métabolisme du dimethylsulfoniopropionate (DMSP), un aérosol (refroidissant l'atmosphère) ou encore de symbiose entre algues diazotrophes (fixant l'azote de l'atmosphère) et cyanobactérie¹²⁹.

Ces dernières années, les avancées en capacité de calculs et en technique d'assemblage^{130,131} ont aussi permis des analyses résolues au niveau du génome des communautés microbiennes des échantillons environnementaux. Ainsi de nombreuses collections de Metagenome-Assembled Genomes (MAGs) eucaryotes¹²¹ mais majoritairement procaryotes¹³²⁻¹³⁴ ont été générées. Ils ont notamment permis de mettre à jour une nouvelle classe de bactéries diazotrophes dis Heterotroph Bacterial Diazotrophs (HBDs) capable de capter l'azote de l'atmosphère bien qu'étant hétérotrophes^{132,135}. Les collections de MAG ont aussi permis de montrer un important potentiel biosynthétique dans leur collection de gènes, bactérien notamment¹³⁴, des convergences fonctionnelles évolutives¹²¹, la distribution géographique des fonctions¹³⁶ mais aussi d'explorer leur niches et caractéristiques écologiques en lien avec leur biogéographie^{121,133,137}. Il est important de noter que

les méthodes d'assemblages de données métagénomiques en MAG souffrent de certains biais. Il est ainsi difficile de reconstruire les grands génomes, ce qui est le cas pour les eucaryotes (dinoflagellés et diatomées notamment bien qu'ils soient des organismes très abondants¹²¹). Il est aussi difficile d'assembler les génomes peu abondants car peu représentés dans les métagénomes. De plus les algorithmes d'assemblage sont loin d'assembler l'intégralité des lectures d'ADN séquencés : pour les génomes eucaryotes de *Tara Océans*, on atteint seulement 20% de lectures assemblées dans des MAGs pour la meilleur fraction de taille (20-180µm)¹²¹. Les assembleurs souffrent aussi de limitations dans l'assemblage des séquences répétées, des séquences très variables, des régions codant pour les ARNs ribosomiques, les ARNs de transfert, les éléments mobiles ou encore les gènes de fonction inconnues¹³⁸.

Bien qu'elles souffrent de biais dans l'estimation des abondances des différentes espèces présentes notamment par le protocole d'extraction¹³⁹, les techniques de séquençage métagénomique (metabarcoding et shotgun sequencing) ont révolutionné notre compréhension des communautés microbiennes (environnementales aussi bien procaryote que eucaryotes). Les récentes avancées en technique de séquençage longue lecture (Oxford Nanopore Technologies¹⁴⁰ et Pacific Biosciences¹⁴¹ entre autres) et la diminution du taux d'erreur de ces nouvelles méthodes^{142,143} sont très prometteuses pour l'avenir de la métagénomique.

Je m'attacherai maintenant à décrire plus précisément une des applications permises par les avancées en métagénomique qui consiste à étudier la biogéographie du plancton et d'évaluer l'impact du réchauffement climatique sur celle-ci, c'est-à-dire la répartition biogéographique présente et future des espèces en lien avec leur niche écologique, le sujet principal de cette thèse.

III. La biogéographie marine: le contexte du seascape

III.1. Le contexte du seascape

La biogéographie est l'étude de la répartition spatiale et temporelle des espèces aussi bien actuelle, à travers les temps géologiques et plus récemment en réponse au

réchauffement climatique (Chapitre I). Cette notion est due aux travaux précurseurs du scientifique et explorateur Alexander Von Humboldt au XIX^{ème} siècle notamment à travers son voyage en Amérique du Sud où il décrit de nombreuses espèces végétales et il est le premier à faire le lien entre leur répartition et le climat¹⁴⁴. Comme brièvement énoncé dans la partie II.1. Bref historique de la planctologie, une question majeure soulevée par l'observation de structures biogéographiques dans l'océan est la compréhension des mécanismes à leur origine ainsi que leur variabilité temporelle et spatiale. En 1989, Steele introduit pour la première fois la notion de « paysage » océanique¹⁴⁵ connue aujourd'hui sous le nom de « seascape » soit « paysage marin »¹⁴⁶ et très différente du paysage terrestre. En particulier, dans les océans, les échelles de variabilité du milieu physique sont très différentes de celles des milieux terrestres. Généralement dans l'océan, la variabilité des paramètres environnementaux (par exemple la température, la salinité) augmente avec les échelles de temps de ces phénomènes alors que dans l'écosystème terrestre, la majeure partie de la variabilité provient de cycles saisonniers avec une faible variance sous-saisonnière¹⁴⁵⁻¹⁴⁷.

Dans le paysage marin, une grande partie de la variabilité environnementale provient ainsi de phénomènes purement physiques subsaisonniers tels que les vagues internes, les tourbillons et les gyres dans un ordre croissant d'échelles temporelles et spatiales. Les ondes internes et les tourbillons de petite à moyenne échelle entraînent des variations à petite échelle et entraînent des irrégularités (patchiness en anglais) locales (~0-100 km) correspondant aux mésocosmes à l'échelle biologique (échelle moyenne)⁴⁴. À plus grande échelle, les gyres (~1000-10000 km) peuvent former des fronts environnementaux avec une richesse et une agrégation d'espèces élevées^{148,149}. Ils distinguent souvent des communautés biologiques très différentes sur les côtés opposés de ces gyres avec des gradients environnementaux abrupts¹⁵⁰. Les petits organismes comme le plancton doivent s'adapter à cette variabilité physique. Cette dernière détermine ainsi une grande partie de la variabilité à petite échelle des communautés planctoniques et l'échelle à laquelle la sélection naturelle

agit potentiellement dans les océans^{151,152}. Les interactions biologiques, la taille, le comportement et le statut trophique des organismes jouent également un rôle majeur. Le phytoplancton, c'est-à-dire les producteurs primaires qui sont passivement transportés par les courants, sont soumis à une dispersion pure (nombre de Reynolds¹⁵³ < 1) et dépendent des apports en nutriments (principalement par les courants verticaux). Ils peuvent avoir des taux de croissance très rapides (temps de doublement < 1 jour) auxquels les consommateurs secondaires doivent être réactifs. En particulier, ces réponses temporelles biologiques rapides peuvent entraîner un découplage entre la biologie et la physique, en particulier à petite échelle (< 1 km)^{146,154,155}.

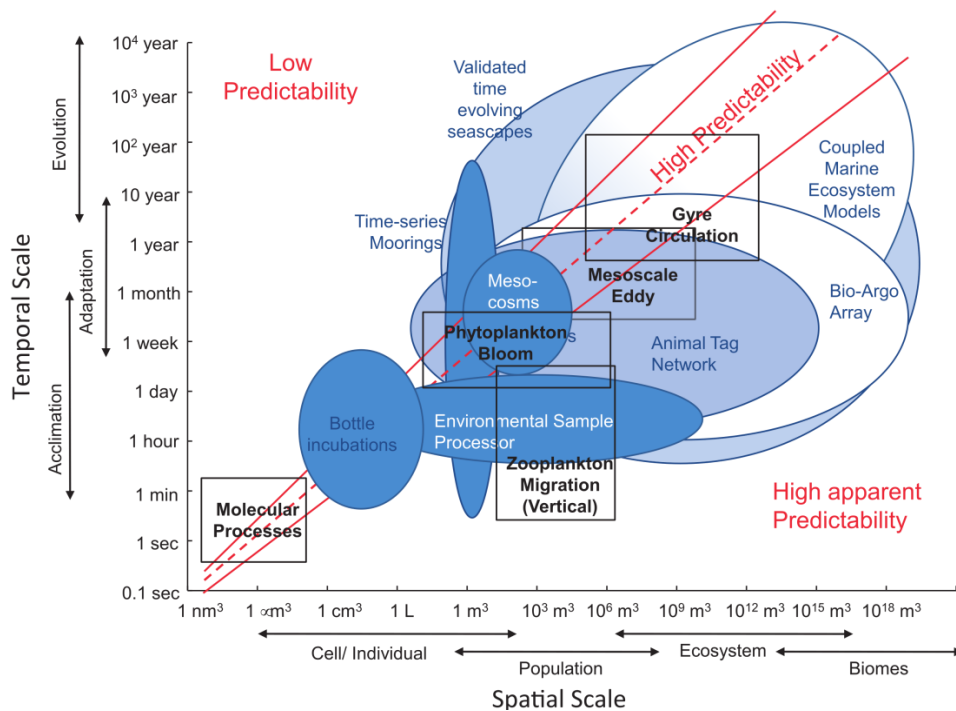


Figure 14 | Diagramme de Stommel montrant les échelles de temps et d'espace des phénomènes biogéophysicochimiques du « seascape »¹⁴⁶. Figure issue de Kavanaugh *et al.*¹⁴⁶, Copyright © 2016, Oxford University Press.

À des tailles plus intermédiaires, bien qu'ils ne puissent pas nager contre les courants horizontaux, les zooplanctons tels que les copépodes utilisent la migration verticale diurne pour attraper leurs proies et se cacher des prédateurs car ils ont une vitesse de nage supérieure aux vitesses des courants verticaux^{156,157}. Ils peuvent également

utiliser des rafales de motilité rapide pour mieux éviter leurs prédateurs¹⁵⁸. Ces interactions entre le transport physique et les comportements biologiques à plusieurs échelles se traduisent dans les océans par une relation croissante continue entre la durée de vie des organismes et des tailles d'organismes congruentes avec leur statut trophique et une forte dépendance aux courants physiques. En d'autres termes, plus un organisme est grand, mieux il pourra utiliser les courants à son avantage (vitesses de nage plus élevées) sans être simplement soumis à une dispersion pure (phytoplancton) et couvrir une gamme spatiale plus large au cours de sa vie. D'autre part, les organismes plus grands sont néanmoins plus limités en termes de dispersion que les plus petits organismes, car l'abondance des organismes est inversement proportionnelle à leur taille¹⁵⁹. Plus récemment, une étude a également démontré que la dérive génétique neutre (mutations aléatoires du génome), combinée aux courants de la circulation globale, est suffisante à l'émergence d'une biogéographie planctonique¹⁶⁰. Enfin la niche écologique dite « fondamentale » d'un organisme, concept introduit et formalisé par Hutchinson en 1957⁶² (Figure 15), c'est-à-dire l'hypervolume des paramètres physico-chimiques permettant la vie d'un organisme ou d'un ensemble d'organismes sera critique pour la détermination de sa biogéographie.

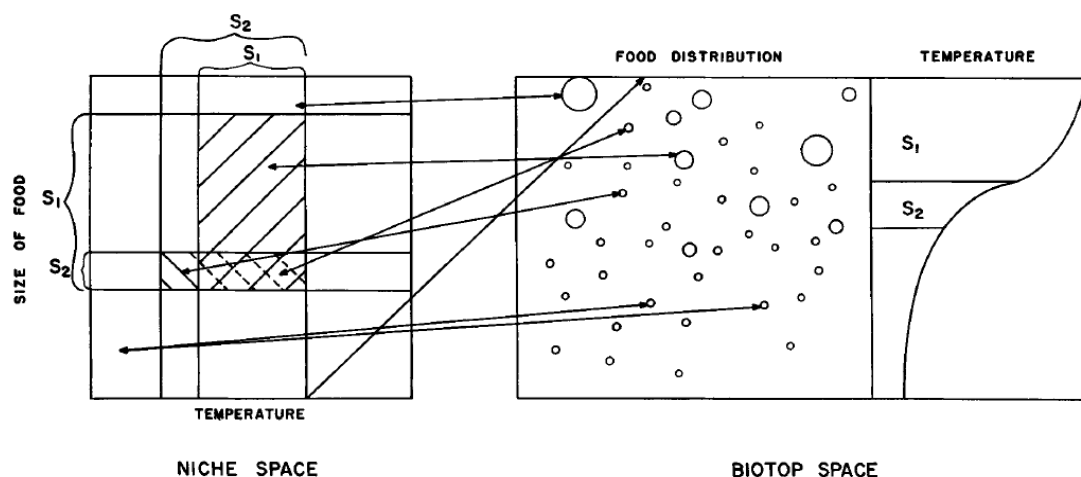


FIGURE 1. Two fundamental niches defined by a pair of variables in a two-dimensional niche space. Only one species is supposed to be able to persist in the intersection subset region. The lines joining equivalent points in the niche space and biotop space indicate the relationship of the two spaces. The distribution of the two species involved is shown on the right hand panel with a temperature depth curve of the kind usual in a lake in summer.

Figure 15 | Le concept de niche écologique « fondamentale » et de niche écologique « réalisée » formalisé en 1957 par le zoologiste Hutchinson⁶². Figure

issue de Hutchinson 1957⁶², Copyright © 2022 by Cold Spring Harbor Laboratory Press, reproduit avec la permission de Cold Spring Harbor Laboratory Press.

Pour un organisme unicellulaire type phytoplancton, elle sera principalement déterminée par sa taille (voir partie III.3. Les facteurs limitant la croissance du phytoplancton : température, taille et nutriments, courants et lumière), sa niche métabolique¹⁶¹⁻¹⁶³ et sa niche thermique (voir partie III.3. Les facteurs limitant la croissance du phytoplancton : température, taille et nutriments, courants et lumière).

Pour les plus gros organismes elle dépendra de son stade de développement, de sa phénologie, et la physiologie globale de l'organisme. La niche « fondamentale » se distingue ainsi de la niche « observée » ou statistique d'un organisme qui est à proprement parler sa biogéographie c'est-à-dire la zone de vie où il est observé. En plus de la niche fondamentale, elle inclut ainsi les interactions biotiques et dans les océans toutes les contraintes physiques, environnementales et trophiques/biotiques (prédation, compétition, symbiose, commensalisme...) liées au seascape que j'ai décrites. Paradoxalement, les organismes planctoniques dans le contexte du seascape sont en permanence transportés par les courants vers de nouveaux environnements ce qui interroge nécessairement sur la stabilité de leur niche écologique notamment la différence entre leur niche « fondamentale » et leur niche « réalisée ». Ainsi certaines études s'intéressent particulièrement à l'impact de la variabilité de température subie par les organismes planctoniques le long des courants notamment sur leur taux de croissance^{164,165}.

La compréhension de l'équilibre entre tous ces mécanismes façonnant les biogéographies microbiennes et les compositions des communautés (taille, sélection naturelle et par l'environnement (niche écologique), dispersion, courants et dérive génétique, Figure 14) reste ainsi une question fortement débattue en écologie marine en raison de la grande complexité de ces processus qui forment le paysage marin. Je m'attacherai maintenant plus particulièrement à décrire le rôle des courants marins.

III.2. La circulation océanique : le rôle des courants

Comme décrit brièvement dans la partie précédente, les organismes planctoniques sont soumis aux courants marins et donc à des changements environnementaux (apports en nutriments, température...) constants. Il se pose alors la question de la description de ces courants marins aussi bien à petite qu'à large échelle et leur impact sur la biogéographie du plancton. Les principaux moteurs des courants marins sont les vents, la densité des eaux (circulation thermo-haline), la force de Coriolis issue de la rotation de la terre, les différences de pression atmosphérique au niveau de la mer ou encore les forces gravitationnelles de la lune et du soleil (marées). Les vents sont responsables des courants superficiels de surface (couche de mélange ~10m¹⁶⁶) à petite et large échelle (Figure 16). La combinaison des vents et de la force de Coriolis génère ce qui est appelé le transport de Ekman¹⁶⁷. Ce transport peut être considéré comme quasi-géostrophique, c'est-à-dire que la force de Coriolis et la force de gradient de pression sont presque à l'équilibre. A large échelles, les alizés déviés vers la droite dans l'hémisphère nord et vers la gauche dans l'hémisphère sud par la force de Coriolis créent une remontée des eaux équatoriales vers les pôles et ce dans les deux hémisphères. Ce transfert d'eau vers les pôles crée une zone de divergence et permet des remontées d'eau froide au niveau de l'Équateur (upwelling), très importantes notamment dans le Pacifique. A l'inverse, la ventilation d'Ekman est le phénomène inverse : dans un anticyclone, le transport d'eau se fait vers le centre du système ce qui crée un excédent de masse au centre de celui-ci. L'excédent plonge alors vers les profondeurs (downwelling) créant des zones peu riches en nutriments typiques des gyres tropicaux oligotrophiques. La combinaison à large échelle de ces deux types de transport crée de larges structures océaniques appelées gyres océaniques visibles sur la figure 16, tournant dans le sens des aiguilles d'une montre dans l'hémisphère nord et dans le sens inverse dans l'hémisphère sud. Ces gyres sont ainsi la combinaison de différents courants de surfaces plus ou moins chauds ou froids. Un courant notable est le Gulf Stream, courant chaud de l'Atlantique Nord prenant son origine dans le golfe du Mexique,

suivant la côte est des Etats-Unis, puis prenant le nom de dérive Nord Atlantique en traversant l'Atlantique du sud-ouest vers le nord-est et ce jusqu'à l'Océan Arctique. Le Gulf Stream et la dérive nord atlantique sont notables par les volumes d'eau qu'ils transportent avec des débits respectifs de 30 Sverdrup (Sv, million $m^3.s^{-1}$) jusqu'à 150 Sv pour le premier (au niveau du Détroit de Floride puis à 55°W vers Newfoundland)^{168,169} et de 40 Sv pour le second (vers son origine sud)¹⁷⁰. Un autre système de courant important que l'on peut citer est le courant antarctique circumpolaire qui crée d'important fronts polaires à partir de 50°S dont les eaux sont très peu mélangées et clairement distinguées bien qu'ayant une certaine variabilité naturelle¹⁷¹. De manière intéressante, le transport d'Ekman peut avoir lieu à toutes les échelles créant des tourbillons océaniques de toutes tailles¹⁷² (de quelques centimètres à l'échelle de dizaines de milliers de kilomètres (gyres)) et dont l'influence sur l'écologie du plancton est largement étudiée notamment sur les irrégularités locales de distribution du zooplancton¹⁷² ou encore l'apport en nutriments^{173,174}. Un système notable de larges tourbillons que j'ai déjà cité résulte de la rencontre du courant des Aiguilles avec les courants plus froids de l'Atlantique Sud. Cette rencontre crée de larges tourbillons océaniques appelés anneaux des Aiguilles.

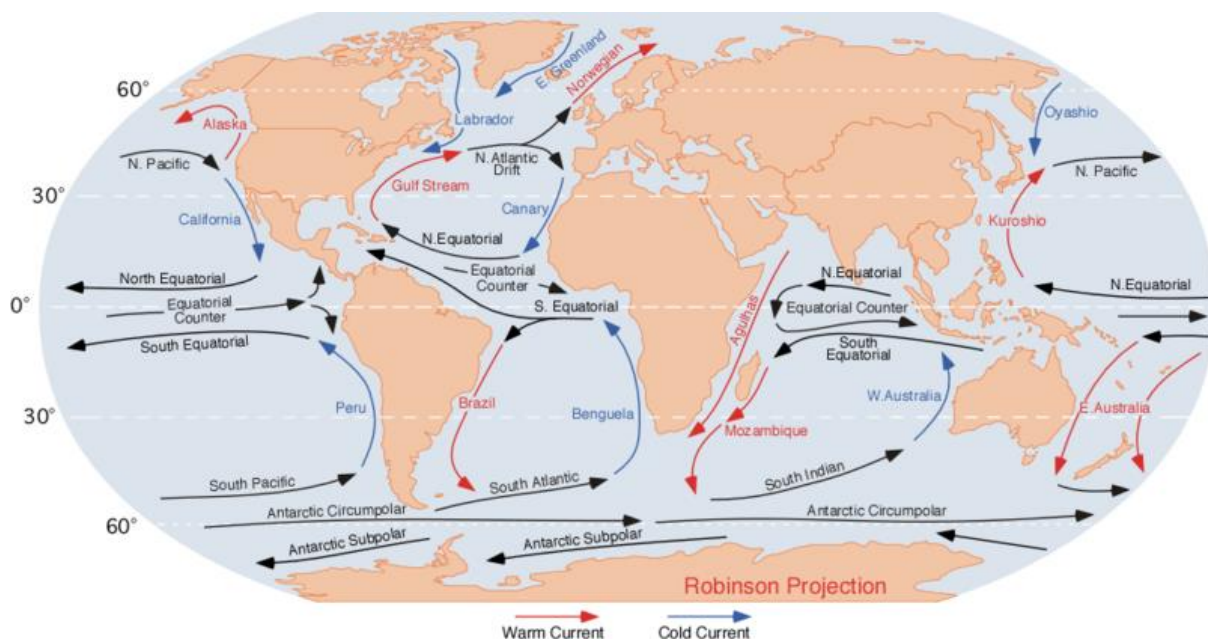


Figure 16 | Les principaux courants marins de surface. Image du domaine public.

Ces tourbillons prennent origine vers le Cap de Bonne-Espérance, remontent ensuite le gyre subtropical de l'Atlantique sud le long des côtes africaines, traversent l'Atlantique puis repiquent vers les côtes de l'Amérique du sud. Ces tourbillons transportent ainsi des communautés de plancton pendant des années sur des milliers de kilomètres^{44,175,176}. Enfin on peut citer un troisième type de transport d'Ekman entraînant les upwellings côtiers, importantes zones de pêche car très riches en plancton. Il a lieu lorsque les vents sont perpendiculaires à la côte et suivant sa direction créent un pompage d'Ekman (upwelling) ou une ventilation. En effet, par la force de Coriolis les eaux sont repoussées avec un angle de 90° par rapport à la direction du vent. Il existe sur Terre quatre grands systèmes d'upwelling côtiers qui sont l'upwelling du Benguela au sud de l'Angola, l'upwelling de Humboldt au large du Pérou, l'upwelling de Californie et enfin l'upwelling du Sénégal et des Canaries. Ils représentent 30% des captures de pêches mondiales¹⁷⁷.

Outre la circulation de surface, la circulation thermo-haline est la circulation océanique résultant des différences de densité des eaux, liée à leur température et salinité. Les eaux chaudes sont peu denses tandis que l'augmentation de la teneur en sel augmente la densité des eaux. Elle est principalement à l'origine de la circulation des eaux en profondeur. Malgré son nom, les vents et les marées semblent être aussi un important moteur de la circulation des eaux en profondeur notamment dans les remontées d'eaux profondes^{178,179}. Ainsi par exemple lorsque les eaux chaudes et salines tropicales du Gulf Stream remontent vers les pôles et se refroidissent elles sont subduites vers le fond lorsqu'elles rencontrent les eaux polaires moins denses car peu salines créant une sorte de tapis roulant géant à l'échelle du globe. Cette circulation est ainsi un important moteur de recirculation et de distribution de la chaleur sur Terre mais aussi de circulation des organismes planctoniques. Dans le nord de l'Atlantique, cette circulation est appelée Circulation méridienne de retournement Atlantique (AMOC) et PMOC dans l'Océan Pacifique. Avec le réchauffement climatique, les eaux tropicales se réchauffent et les glaces arctiques fondent entraînant une diminution du gradient de densité entre les eaux polaires et

les eaux tropicales ce qui pourrait ralentir l'AMOC et avoir de graves conséquences sur la régulation du climat mondial et l'écosystème planctonique bien qu'un point de bascule reste difficile à identifier¹⁸⁰. La diminution future de l'AMOC et lors du siècle dernier reste encore débattue¹⁸¹ bien que le déclin semble l'emporter¹⁸²⁻¹⁸⁴. D'autre part, l'augmentation du forçage des courants par l'augmentation des vents (dûe à l'apport d'énergie à l'atmosphère) avec le réchauffement climatique semble elle avoir des effets complexes sur la circulation océanique globale¹⁸⁵. Un défi majeur dans la compréhension des courants est leur modélisation permettant notamment de projeter les potentiels changements dans le système de courant avec le réchauffement climatique¹⁸⁶. Le mouvement des fluides visqueux est ainsi gouverné par les équations de Navier-Stokes¹⁸⁷. En particulier pour les océans, c'est l'approximation de Boussinesq de ces équations qui est utilisée notamment dans le modèle de circulation développé par le MIT¹⁸⁸ (Figure 17). Cette approximation considère le fluide incompressible à surface libre et avec gradient verticale de masse volumique. Grossièrement, en intégrant numériquement ces équations dans une grille représentant la topographie de la Terre, elles permettent d'obtenir un champ de vitesse à une certaine résolution spatiale et temporelle approximant la réalité de l'océan physique. Ces modèles sont ainsi partie intégrante des modèles système Terre que je présenterai dans une partie ultérieure (IV.2. La modélisation biogéochimique au sein des modèles système terre). Ces modèles permettent aussi d'étudier la connectivité océanique par l'approche dite « Lagrangienne » consistant à suivre des « traceurs » ou particules qui suivent le champ de vitesse du modèle. Ainsi, certaines études mettent en évidence des régions de forte connectivité ainsi que leur variabilité temporelle au sein des océans^{189,190}.

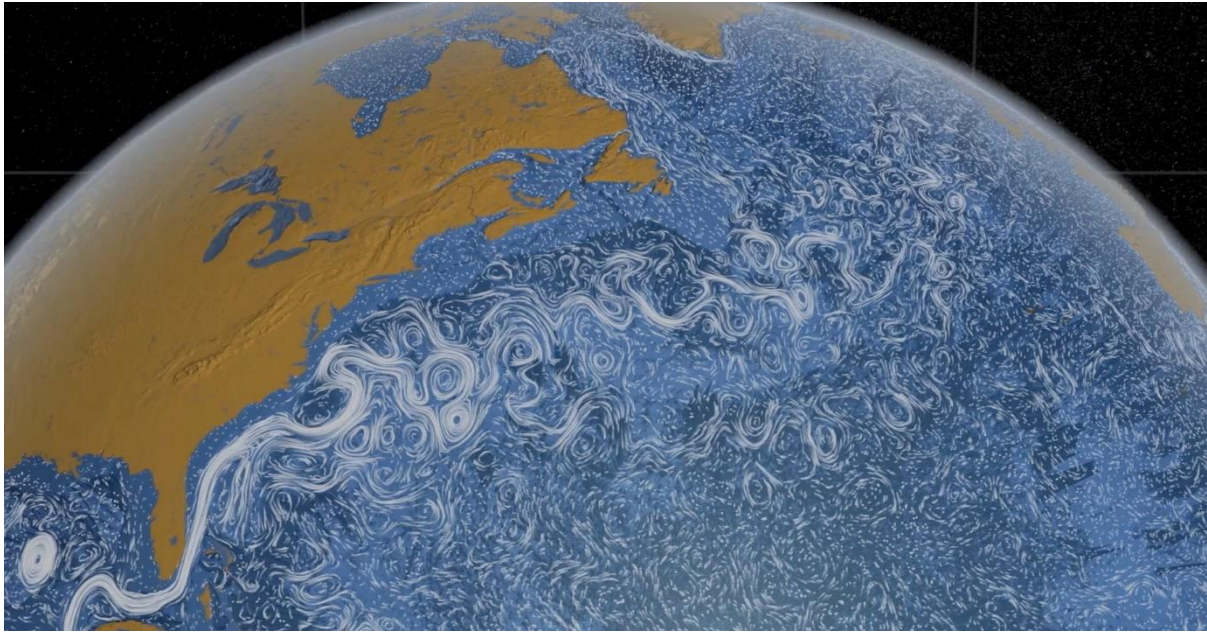


Figure 17 | Les courants marins modélisés par le modèle de circulation global MITgcm^{188,191} : zoom sur le Gulf Stream. Image issue de nasa.gov (Crédit : NASA's Goddard Space Flight Center).

D'autre part, ces traceurs peuvent être considérés comme des particules de plancton. Ainsi, la connectivité océanique entre les stations des expéditions *Tara Océans* a été déterminée par calcul du temps minimal de transport entre chaque paire de stations. Ces temps de transport sont ainsi fortement corrélés avec les patterns biogéographiques obtenus par analyse métagénomiques⁸⁵ ou de metabarcoding^{85,192}. De manière intéressante, en se basant sur la métagénomique des patterns biogéographiques clairs et à l'échelle globale sont apparus pour les différentes classes de taille du plancton (voir Chapitre I, Article 1 et 2).

Outre les courants marins, de nombreux paramètres abiotiques peuvent influencer la biogéographie du plancton. Je les décrirai maintenant avec un focus sur les paramètres influençant la croissance du phytoplancton.

III.3. Les facteurs limitant la croissance du phytoplancton : température, taille et nutriments, courants et lumière

La croissance du plancton dépend avant tout de la production primaire par le phytoplancton photosynthétique comme je l'ai déjà brièvement évoqué dans la partie I.2. Position écologique du plancton. Je décrirai ici les facteurs limitant cette

production primaire sachant que celle-ci limite la croissance du reste de la communauté pélagique (zooplancton, reminéralisateur, poissons...).

Le premier facteur que l'on peut citer comme limitant la croissance du phytoplancton est la température. Bien que dans les océans son rôle soit souvent considéré comme insignifiant puisque ce sont souvent d'autres facteurs qui limitent cette croissance¹⁹³, la température semble néanmoins limiter la croissance maximale du phytoplancton¹⁹⁴. La courbe d'Eppley¹⁹⁴, une exponentielle croissante de la température impose ainsi une limite supérieure à la croissance du phytoplancton dans des conditions de lumière et de nutriments non limitant :

$$\log_{10}\mu = 0.0275.T - 0.070$$

Où μ est le taux de croissance maximal et T est la température. Dans le chapitre 3 je coderai cette dépendance comme un facteur limitant le taux de croissance maximale. Cette relation est attendue du fait que la croissance d'un organisme unicellulaire dépend de son activité métabolique. Selon la loi d'Arrhenius la vitesse d'une réaction chimique est en effet une fonction croissante de la température et dépendante de l'énergie d'activation de la réaction en question¹⁹⁵. En plus de cette loi, chaque organisme possède aussi une enveloppe de température dans laquelle il peut croître. Il s'agit de sa niche thermique qui dépendra de sa capacité à s'adapter à différentes températures. Celle-ci dépend de nombreux facteurs génétiques liés à la plasticité phénotypique (acclimatation, voir chapitre II) ou la capacité adaptative des organismes¹⁹⁶⁻¹⁹⁸. En réponse au réchauffement climatique ou un changement d'environnement, le plancton peut ainsi suivre (passivement) différentes stratégies : migration passive, acclimatation (modulation de l'expression des gènes), adaptation (au niveau de sa séquence génétique) ou extinction. Je reviendrai sur ces stratégies lorsque je décrirai la démarche scientifique suivie dans cette thèse.

Un second facteur critique pour la croissance du phytoplancton est sa taille, c'est-à-dire la taille de la cellule-organisme. Ce facteur est directement lié à la capacité de la cellule à absorber des nutriments. En effet, d'un point de vue physique le taux

maximal d'absorption d'un nutriment est proportionnel à la surface de sa cellule. Ainsi les plus petites cellules auront une plus forte affinité pour les nutriments et atteindront ainsi leur maximum de croissance à de faibles concentrations en nutriments (voir chapitre III). On retrouvera en général de plus petites cellules de type cyanobactéries (*Synechococcus*, *Prochlorococcus*) dans les gyres tropicaux oligotrophiques (bien qu'il y fasse chaud, ils sont peu riches en nutriments) et des grandes cellules (diatomées, dinoflagellés) près des côtes et en s'écartant vers les pôles¹⁴⁹. En revanche la quantité de nutriments nécessaires à la vie de la cellule est proportionnelle à son volume et sa capacité à croître l'est aussi. Selon la théorie métabolique de l'écologie¹⁹⁹ (MTE) c'est ce dernier paramètre qui influence le taux maximal de croissance de la cellule :

$$\mu_{max} = aV^b$$

Où μ_{max} est le taux de croissance maximale de la population, V le volume de la cellule, a et b des paramètres. En revanche selon le modèle de quota de Droop²⁰⁰, il a été montré que pour des cellules de très petite taille, c'est la capacité à absorber les nutriments qui peut être déterminante ce qui résulte en une balance entre la taille de la cellule et le taux de croissance maximal²⁰¹. Dans le chapitre 3, je développerai un modèle de croissance du phytoplancton basé sur les équations de Monod²⁰² pour la dépendance aux nutriments, considérant une croissance logistique du plancton dépendant uniquement de la concentration ambiante en nutriments (nitrate, voir chapitre 3), modèle plus simple que le modèle de Droop²⁰³. Bien évidemment, la croissance du phytoplancton dépend de la disponibilité de nombreux nutriments et non d'un seul dont les principaux sont le nitrate²⁰⁴, le phosphate²⁰⁴ et le fer^{204,205} et d'autres micronutriments comme le manganèse²⁰⁶. Pour les diatomées à la frustule siliceuse, leur croissance peut aussi dépendre de la disponibilité en silice dissoute²⁰⁷. Une fois caractérisée la dépendance aux nutriments, on peut maintenant se poser la question de leur apport aux cellules phytoplanctoniques dans les océans. Comme je l'ai expliqué précédemment, le transport d'Ekman par ventilation a tendance à diminuer l'apport en nutriments au sein des grandes structures océaniques

que sont les gyres tropicaux. Ainsi, ces gyres sont peu riches en nutriments et ont une variabilité saisonnière faible car fortement stratifiés, c'est-à-dire que la couche de mélange de surface est peu profonde avec un fort gradient de température lorsqu'on s'enfonce vers le fond limitant ainsi l'apport en nutriments par les eaux profondes. Toutefois au sein de ces larges structures océaniques, les tourbillons de plus petites échelles ont été caractérisés comme ayant un apport net positif en nutriments aux communautés planctoniques¹⁷³. Une structure beaucoup plus intéressante reste le gyre tempéré-subpolaire de l'atlantique nord qui connaît successivement une floraison (bloom en anglais) commençant au début du printemps aux latitudes tempérées puis se finissant en été aux plus hautes latitudes¹⁷³. On peut mentionner l'expérience historique North Atlantic Bloom Experiment (NABE) mené par le Joint Global Ocean Flux Study (JGOFS) en 1989-90, qui regroupa plus de 200 scientifiques et étudiants pour étudier le phénomène, pour comprendre l'importance et l'intérêt qui est porté à ce phénomène. La première explication du bloom de l'Atlantique Nord est amenée par Sverdrup en 1953 avec l'hypothèse de la profondeur critique²⁰⁸ (Figure 18). Cette hypothèse est directement basée sur la physique de l'océan et de la couche de mélange. En hiver dans l'Atlantique Nord les vents sont forts et les températures sont froides ce qui induit une couche de mélange relativement profonde dépassant la couche photique (~100 m, profondeur à partir de laquelle la lumière ne passe plus). Ainsi si on suppose que le phytoplancton est distribué de manière homogène dans cette couche, la croissance du phytoplancton est limitée par le nombre de cellules ayant accès à la lumière pendant cette période. Ensuite, quand vient le printemps, la couche de mélange rétrécit (par stratification liée à la diminution des vents et l'augmentation de la température de surface) et le phytoplancton est ainsi plus concentré dans la couche de mélange. Selon Sverdrup c'est lorsque la couche de mélange atteint une profondeur critique (la « critical depth ») telle que la croissance du phytoplancton devient suffisante pour compenser les pertes (broutage par le zooplancton, virus, respiration, cellules mortes coulant vers les fonds). Sverdrup eut déjà sans la développer l'idée de l'hypothèse suivante en notant qu'un bloom peut

tout de même se déclencher si la turbulence dans la couche de mélange est suffisamment faible. Quarante-cinq ans plus tard Jef Huisman de l'université d'Amsterdam reprend et formalise l'hypothèse de la turbulence critique^{209,210} (Figure 18) sans invalider l'hypothèse de Sverdrup. Cette hypothèse distingue ainsi la couche de mélange défini par la densité (ou la température) et la couche de mélange active (où il y a réellement de la turbulence). Ainsi lorsque le temps est suffisamment clé-ment (peu de mélange, apport de lumière et de chaleur), la croissance du phyto-plancton peut compenser le transport turbulent du phytoplancton vers les profon-deurs ce qui déclenche un bloom non loin de la surface de l'océan. Enfin on peut citer l'hypothèse de perturbation-récupération avancé en 2013 par Behrenfeld²¹¹ (Fi-gure 18). Cette hypothèse explique plus une augmentation de la biomasse totale de phytoplancton en fin d'hiver que l'enclenchement d'un bloom à proprement parler. En automne la perturbation de la couche de mélange (diminution de température et augmentation de la turbulence) a un impact plus important sur les brouteurs (zoo-plancton) que sur le phytoplancton entraînant une augmentation de la biomasse phytoplanctonique intégré sur l'ensemble de la couche de mélange mais pas une augmentation de sa concentration (Figure 18). En effet, au cours de l'hiver la couche de mélange ne cesse de s'approfondir et donc de diluer le phytoplancton. J'ai pris ici pour exemple le système du bloom Nord Atlantique qui est probablement le système océanique le plus étudié en océanographie biologique. Néanmoins, les mécanismes décrits ici restent valables pour de nombreux systèmes de blooms planctoniques bien qu'ils restent encore largement discutés et controversés car difficiles à prou-ver²¹². D'un point de vu compositionnel, l'hiver puis le bloom printanier de l'Atlantique Nord est en général dominé par des espèces de petits flagellés (<3 μm)²¹³ et le pic de biomasse peut éventuellement être dominé par de grosses cellules de diatomées (>10 μm)²¹⁴. La fin des blooms phytoplanctoniques peut avoir plusieurs causes comme un broutage surpassant le taux de division du phytoplancton, les at-taques virales^{215,216}, l'export de biomasse²¹⁷, ou encore la limitation par un macronu-triment ou par le fer^{218,219} par leur épuisement. De manière globale on peut considé-

rer que les océans sont « anémiés » et le manque de fer représente ainsi un important frein aux blooms planctoniques comme par exemple dans l'océan austral²²⁰. Ainsi de nombreuses expériences de fertilisation par le fer ont été menées à différents endroits des océans du globe²²¹⁻²²³. De manière intéressante, une étude récente basée sur la métagénomique a permis de caractériser les nutriments limitant de communautés de cyanobactéries type *Prochlorococcus* à l'échelle globale par l'étude de l'expression des gènes impliqués dans l'assimilation de l'azote, du phosphate et du fer²²⁴. Dans les autres phénomènes importants dans l'apport en nutriments aux communautés phytoplanctoniques, on peut citer la diazotrophie, la capacité de certains micro-organismes (des bactéries), à capter l'azote de l'atmosphère²²⁵ particulièrement active dans les gyres oligotrophiques²²⁶. Enfin les phénomènes de reminéralisation notamment par les communautés bactériennes, sont particulièrement importants dans les zones d'oxygène minimale par exemple^{227,228}. Dans le contexte complexe du seascape mêlant processus physiques de différentes échelles et processus biogéochimiques complexes et intriqués, quelle biogéographie du plancton émerge ?

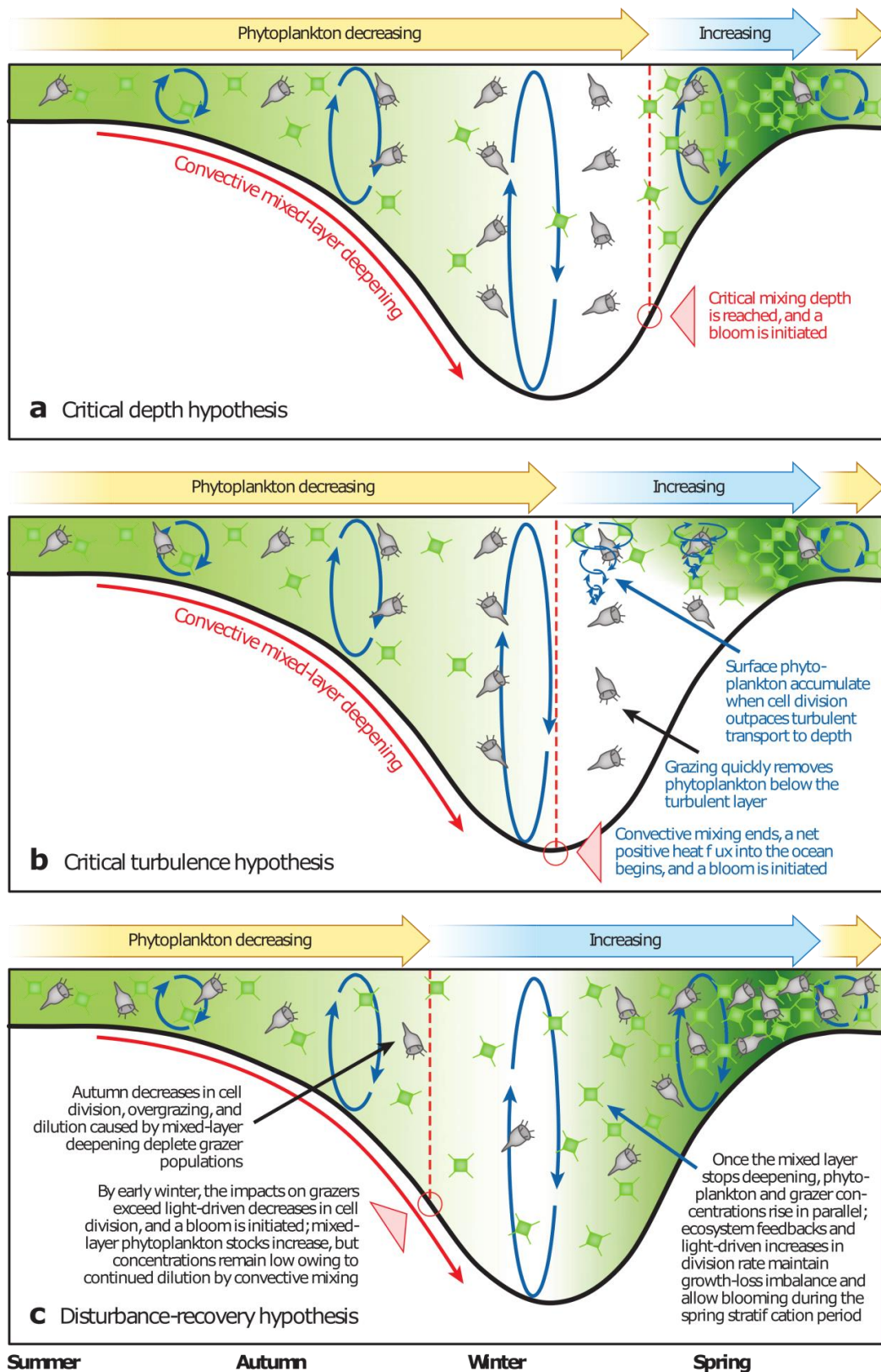


Figure 18 | Différentes hypothèses pour l'explication des « blooms » phytoplanktoniques²¹² : cas typique du gyre subpolaire de l'Atlantique Nord.

Figure issue de Behrenfeld et al.²¹². L'intensité du vert reflète ici la concentration en phytoplankton.

III.4. Etudes biogéographiques actuelles du plancton

Le partitionnement biogéographique de l'océan global est une question majeure de l'océanographie biologique moderne. Cette question est critique dans la compréhension biogéochimique de l'océan en lien avec les grands cycles des éléments tel que le carbone et l'azote. La réponse au changement climatique et la rétroaction sur le climat de la biogéographie marine apparaissent aussi au centre de cette question. Un partitionnement biogéographique très connu et de référence est aujourd'hui celui proposé par Alan Longhurst dans son livre « Ecological Geography of the Sea » paru en 2007⁵⁹. Il propose ainsi un partitionnement de l'océan en 56 BioGeoChemical Provinces (BGCPs)^{59,229} (Figure 19). Ces provinces sont caractérisées par 4 principaux paramètres biogéochimiques que sont la bathymétrie, la concentration en chlorophylle *a*, la température de surface et la salinité. En 2013 puis 2020, elles seront actualisées et leur variabilité temporelle au cours de l'année évaluée²²⁹ ainsi que leur réponse au réchauffement climatique avec l'émergence de nouvelles provinces²³⁰.

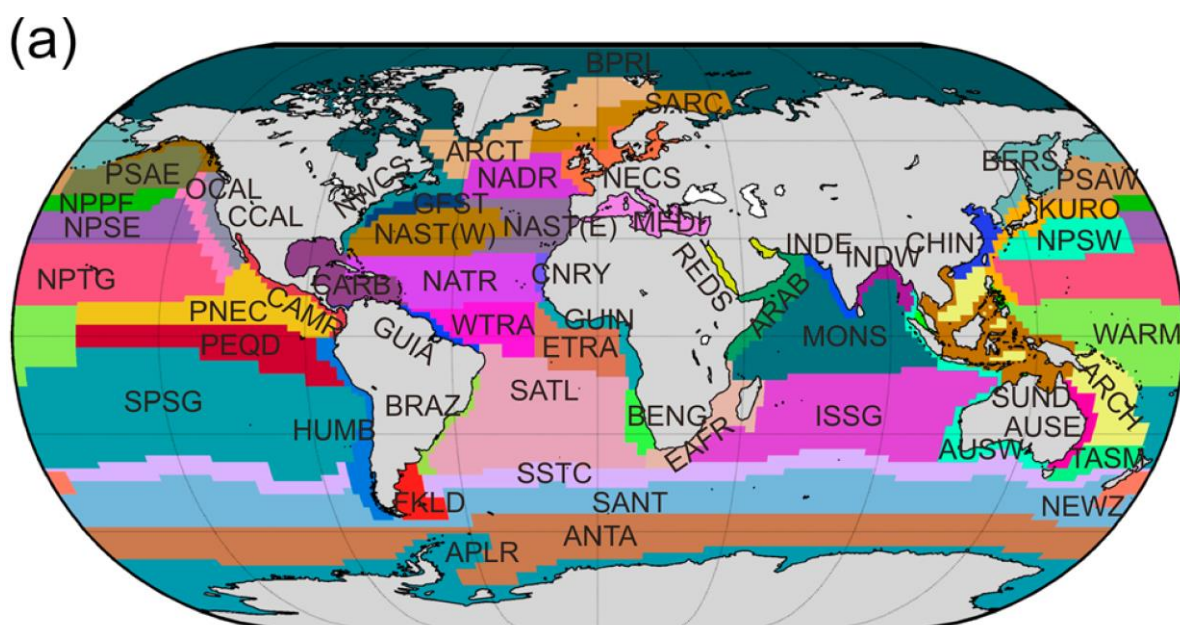


Figure 19 | Biogéographie des océans proposé par Longhurst en 2007^{59,229} : un partitionnement en 56 provinces biogéochimiques. Figure issue de Reygondeau *et al.*²²⁹, © 2020 Reygondeau.

Parmi les autres partitionnements biogéographiques qui ont été proposés on peut citer celui de Fay et McKingley²³¹ ou encore celui d'Oliver et Irwin²³², le premier étant basé sur la température de surface, la concentration en chlorophylle *a*, la fraction de glace et la profondeur maximale de la couche de mélange et le second sur la température de surface et la couleur de l'océan.

De nombreuses autres études utilisent une approche plus proche de la biologie pour caractériser la biogéographie des espèces marines : c'est l'approche de modélisation de la distribution des espèces. Cette approche consiste à modéliser la niche écologique réalisée (concept évoqué précédemment en section III.1. Le contexte du seascape et utilisée directement dans les articles 2 et 3 du chapitre I) d'une espèce par des approches probabilistes de machine learning. Ainsi, une étude récente propose un partitionnement des océans en biomes basé sur les modèles de distribution de 536

espèces de phytoplancton océanique provenant de sources variées⁶⁷. On peut aussi citer des études se concentrant sur des zones géographiques plus locales utilisant par exemple les données du Continuous Plankton Recorder dans l'Atlantique Nord et modélisant la répartition biogéographique de 87 espèces de diatomées et dinoflagellés et leur réponse au réchauffement climatique (taux de migration, paramètres influents...)²⁴ ou encore une étude sur les remplacements d'espèces de copépodes en mer méditerranée en réponse au réchauffement climatique⁶⁴. De nombreuses études portent aussi sur le réchauffement climatique récent et montrent des réponses variables au réchauffement climatique de différentes espèces de planctons qui suivent ou non leur niche thermique initiale en suivant ou non l'augmentation des températures²³³⁻²³⁵.

Ces nombreuses études projettent la réponse des communautés de plancton au réchauffement climatique, mais comment le plancton s'inscrit-il dans les cycles des éléments et comment le plancton est-il modélisé dans les modèles biogéochimiques et de réchauffement climatique actuels ? Je répondrai maintenant à ces questions.

IV. Les grands cycles biogéochimiques et la modélisation du climat

IV.1 La place du plancton dans les grands cycles biogéochimiques et dans le système Terre

La biomasse en carbone représentée par le plancton est tout à fait négligeable au sein du cycle du carbone et ne représente qu'un stock de 2.5 à 7.1 gigatonne de carbone (GtC) partagée de manière équivalente entre hétérotrophes et autotrophes^{236,237}. En comparaison, la biomasse des plantes terrestres représente un total de 450-650 GtC²³⁷ (Figure 20). La majeure partie du carbone océanique est ainsi dissoute (38000 GtC, Figure 20) ou sous forme de carbone organique dissous ou particulaire (DOC et POC constitués de débris, des fèces du zooplancton et des cellules mortes) (700 GtC, Figure 20)²³⁷. La DOC et de manière plus importante la POC sont à l'origine d'un flux qui ne figure pas dans la figure 20 car considéré par les auteurs comme négligeable et à l'équilibre, c'est-à-dire non impacté par les émissions de CO₂. Ce flux est la pompe à carbone biologique directement liée à l'activité planctonique (déjà évoqué en partie I.2. Position écologique du plancton). En effet, suite à la production primaire, la matière organique planctonique peut s'agréger et former des particules de matière qui sombrent vers les profondeurs avec plus ou moins d'efficacité selon leur taille et la température notamment²³⁸. Ainsi la composition des communautés²³⁹ de plancton et notamment la taille des organismes de la communauté (celle du phytoplancton²⁴⁰⁻²⁴² mais aussi des copépodes^{243,244}) sont des éléments cruciaux et complexes déterminant cette pompe biologique (voir article 2 du chapitre I). A 100 m de profondeur les estimations de ce puits de carbone est de l'ordre de 5 à 15 GtC par an^{238,245-248}. Je reprendrai certaines de ces estimations dans le chapitre I, article 2 en montrant notamment une potentielle chute de l'ordre de 4% de cette pompe biologique d'ici la fin du siècle en réponse au changement climatique¹³⁷. Suite à la reminéralisation tout au long de la colonne d'eau, les flux jusqu'au plancher océanique sont bien moindres et sont estimés être de l'ordre de 0.5 à 1.6 GtC par an²⁴⁸⁻²⁵¹. Néanmoins ces derniers sont probablement

sous-estimés comme le suggèrent certaines études montrant l'importance d'évènements ponctuels d'export de carbone organique vers les fonds²⁵². Pour résumer, dû à un turnover rapide des communautés et à la pompe biologique, le plancton ne représente qu'un réservoir de carbone faible dans le cycle du carbone. Néanmoins, via la production primaire et la pompe biologique il reste non négligeable dans son impact sur le cycle du carbone et les potentiels changements en production primaire, liés au réchauffement climatique, et affectant potentiellement la pompe à carbone biologique restent très incertains aujourd'hui²⁵³⁻²⁵⁵.

Le plancton est aussi partie intégrante du cycle de l'azote, élément chimique clé dans la biosynthèse des acides ribonucléiques (ADN et ARN) mais aussi des acides aminés et donc des protéines²⁵⁶. La chimie de l'azote et son cycle biogéochimique est complexe car l'azote possède 5 électrons de valences permettant un degré d'oxydation compris entre -3 et +5²⁵⁶. Le réservoir principal d'azote sur terre est l'atmosphère qui contient 78% d'azote (en volume) sous forme de diazote (N₂) gazeux et représente 3.7.10⁹ Tg d'azote²⁵⁷. Pour simplifier je décrirai ici principalement le cycle de l'azote marin. L'ensemble du cycle de l'azote incluant les perturbations anthropogéniques est décrit par la figure 21. On distingue deux principales sources d'azote pour la production primaire par le phytoplancton : les apports dits nouveaux et les apports issus du recyclage, concepts distingués par Dugdale et Goering en 1967²⁵⁸. Les apports nouveaux²⁵⁹, sous forme de nitrate (NO₃⁻), de nitrite (NO₂⁻) ou d'ammonium (NH₄⁺), sont issus des apports issus de la remontée des courants froids (upwelling principalement)²⁷, de la fixation de l'azote de l'atmosphère par les bactéries dites diazotrophes et principalement photosynthétiques (cyanobactéries²²⁵) mais aussi des apports provenant des terres (fleuves, ruissellement). L'apport par recyclage ou régénération est lui issu du recyclage de l'azote dans la zone photique principalement sous forme d'ammonium (NH₄⁺)²⁵⁸ issu notamment de la reminéralisation, de la lyse cellulaire, du broutage²⁶⁰ ou encore de fuites cellulaires provenant des cyanobactéries diazotrophes²⁶¹⁻²⁶³ fixant l'azote en excès par rapport à leur rapport cellulaire C:N.

The global carbon cycle

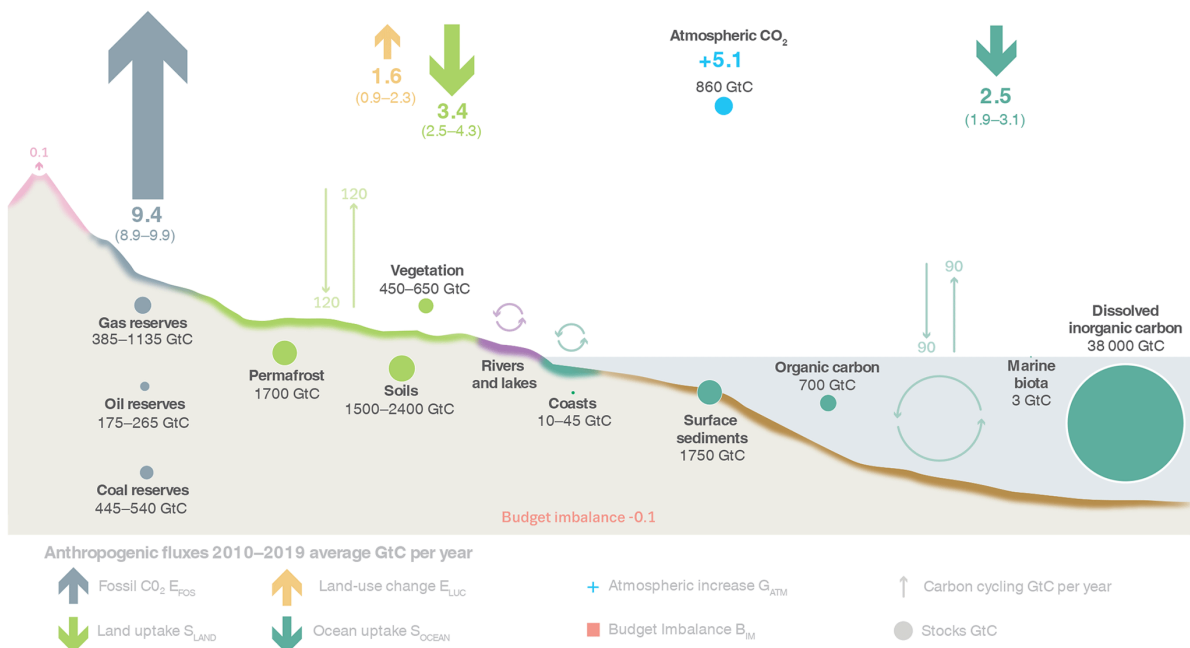


Figure 20 | Le cycle du carbone et l'excédent atmosphérique de CO₂ lié aux activités humaines (flèche bleue ciel sur la figure) : moyennes annuelles des flux et stocks sur la décennie 2010-19. Figure issue de Friedlingstein *et al.*²³⁷ © Friedlingstein 2020. Les flux liés aux activités humaines sont mis en évidence par les flèches en gras dans le haut de la figure.

L'apport par recyclage ou régénération est lui issu du recyclage de l'azote dans la zone photique principalement sous forme d'ammonium (NH₄⁺)²⁵⁸ issu notamment de la reminéralisation, de la lyse cellulaire, du broutage²⁶⁰ ou encore de fuites cellulaires provenant des cyanobactéries diazotrophes^{261–263} fixant l'azote en excès par rapport à leur rapport cellulaire C:N. De manière simplifiée, la fraction de la production primaire provenant d'apports par le nitrate (f-ratio)²⁵⁹ est utilisée pour mesurer le rapport entre apports nouveaux et apports par recyclage. En effet, principalement dans les zones aphotiques profondes, l'ammonium est en général oxydé en nitrite puis nitrate par le processus de nitrification²⁶⁴ réalisé en 2 étapes (nitritation puis nitratisation) par les bactéries et archées ce qui fait du nitrate la principale source d'apport nouveau à la zone photique. Il a néanmoins été mis en évidence que le processus de nitrification peut avoir lieu dans toute la colonne d'eau²⁶⁵ et ce jusqu'à 40% d'ammonium reminéralisé dans les gyres oligotrophiques

(spécifiquement le site BATS dans l'Atlantique Nord²⁶⁶).

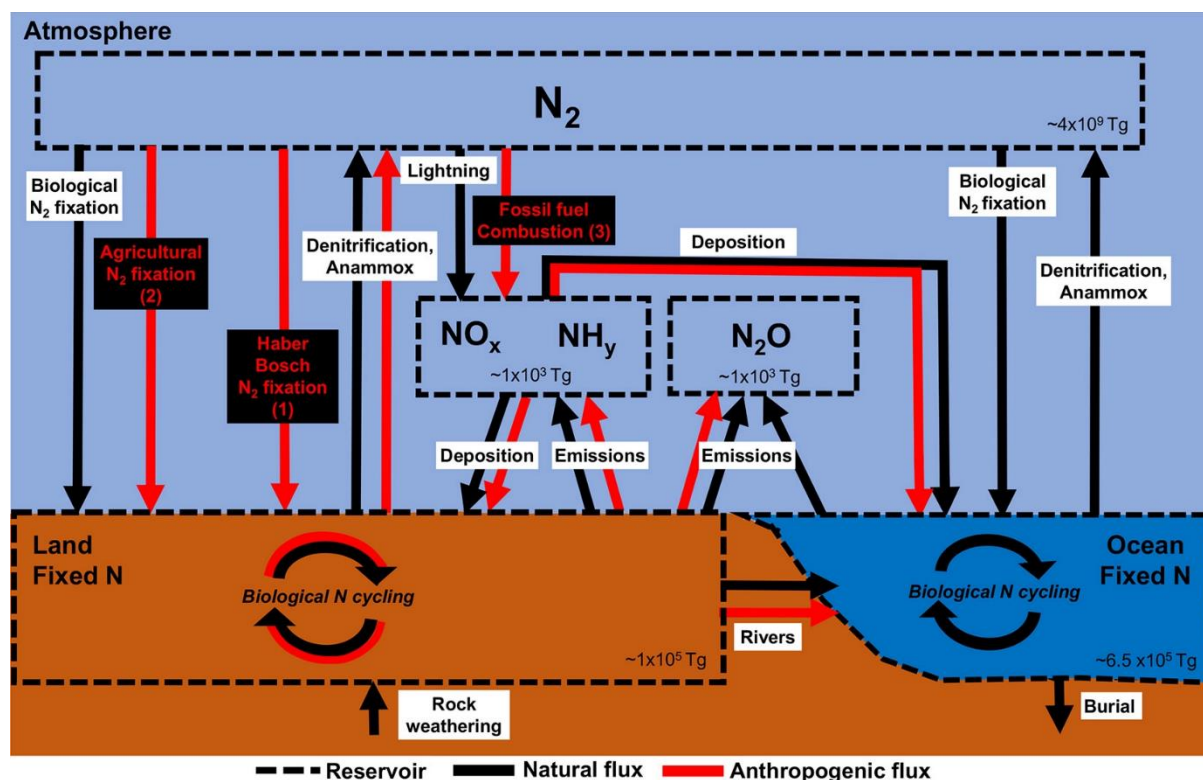


Figure 21 | Le cycle de l'azote²⁶⁷. Reproduit avec permission de Zhang *et al.*²⁶⁷.

Copyright 2020 American Chemical Society.

Avec le réchauffement climatique la nitrification mais aussi la dénitrification apparaissent comme des potentiels limitateurs de la production primaire en lien avec l'acidification des océans²⁶⁸ mais aussi la circulation des courants²⁶⁹. Un autre ratio important est l'e-ratio qui est le rapport entre la production primaire exportée (qui coule sous la zone photique) et la production primaire totale. En assumant sur le long terme que la masse en nitrate est à l'équilibre dans la zone photique (entrées=sorties), on peut confondre le e-ratio et le f-ratio d'où la notion de ef-ratio²⁷⁰. Ainsi en général on distingue dans les océans les zones de fortes productions nouvelles (upwelling, zones polaires (ef-ratio jusqu'à 0.7 en été)) des zones de recyclage important (gyres oligotrophiques)²⁷⁰. Néanmoins la production nouvelle est en partie soutenue dans les zones oligotrophiques par la diazotrophie qui en réponse au réchauffement climatique est prédite comme prenant une importance plus importante dans le pacifique notamment mais diminuerait à l'échelle globale bien que les modèles présentent des tendances contrastées²⁷¹. Nos

résultats du chapitre I, article 2 vont également dans le sens d'une augmentation des diazotrophes par expansions des gyres oligotrophiques tropicaux notamment dans le Pacifique¹³⁷. De manière intéressante les données génomiques de *Tara* Océans ont mis à jour un grand nombre de bactéries diazotrophes encore méconnues dites bactéries hétérotrophes diazotrophes et plus abondantes que les cyanobactéries diazotrophes mais dont le rôle dans la fixation d'azote reste encore à élucider^{132,135}. Un modèle récent semble montrer que celles-ci pourraient contribuer de manière significative²⁷². Le plancton participe à de nombreux autres cycles biogéochimiques tels que celui de l'oxygène, du phosphore, du soufre ou encore du silicate, mais je m'arrêterai ici. De manière importante les ratios entre les différents éléments peuvent être considérés comme constants au sein du plancton avec un rapport $-O_2:C:N:P$ de 138:105:16:1, appelé le Redfield ratio²⁷³ ($-O_2$ représentant l'oxygène nécessaire) bien qu'il existe tout de même des variations à ce ratio²⁷⁴.

De nombreux résultats que j'ai cités ici sont issus de modèles dit mécanistiques de la biogéochimie marine provenant de notre compréhension causale des phénomènes. Je m'attacherai maintenant à décrire ces modèles et en particulier au sein des modèles systèmes Terre très utilisés dans les approches de modélisation statistique (Chapitre I, article 2) de réponse du vivant, et en particulier du plancton, au réchauffement climatique.

IV.2. La modélisation biogéochimique au sein des modèles système terre

Bien qu'une grande partie de nos connaissances soient issues des observations des systèmes (quel qu'ils soient), notre compréhension mécanistique (entendu ici comme « causale »²⁷⁵) des phénomènes, permettant l'énoncé de lois mathématiques décrivant la dynamique du monde qui nous entoure, est une approche très utilisée et générant de nombreuses connaissances. L'approche mécanistique a cet atout de pouvoir ériger des lois générales mathématiques extrapolables à des situations non observables telles que le futur et en particulier le réchauffement climatique ou encore des crises passées²⁷⁶. Ainsi en 2021, le prix Nobel de physique a été attribué aux physiciens Giorgio Parisi, Syukuro Manabe et Klaus Hasselmann tous trois

précurseurs dans la modélisation du climat terrestre, de sa variabilité et dans la prédiction du réchauffement climatique²⁷⁷. Je décrirai ici le compartiment biogéochimique marin des modèles système Terre actuels qui nous intéressent ici (Figure 22). Ces modèles décrivent la biogéochimie marine dans une grille résolue spatialement représentant la topographie des océans, en intégrant dans l'espace et le temps, des équations décrivant la dynamique du plancton et des nutriments dans l'océan, en couplage avec l'atmosphère²⁷⁸, les modèles de courants océanique et de température¹⁹¹, et les modèles de glace océanique²⁷⁹. Je m'intéresserai en particulier aux modèles français (PISCESv2⁶⁰) et au modèle étasunien (MITecco^{149,280}). Le modèle français comporte 24 variables modélisées incluant notamment 2 compartiments de phytoplancton (diatomées et nanophytoplancton), deux classes de taille de zooplancton (micro- et mesozooplancton), la matière organique dissoute (DOM) et particulaire POM), une description de la chimie des carbonates et des principaux nutriments⁶⁰. PISCESv2 a pour particularité d'attacher une grande importance à la modélisation de la chimie du fer aussi bien dans le compartiment du phytoplancton avec un modèle de quota (dérivé du modèle de Droop²⁰⁰) que dans le modèle du fer dissous⁶⁰ (deux ligands et 5 formes de fer²⁸¹).

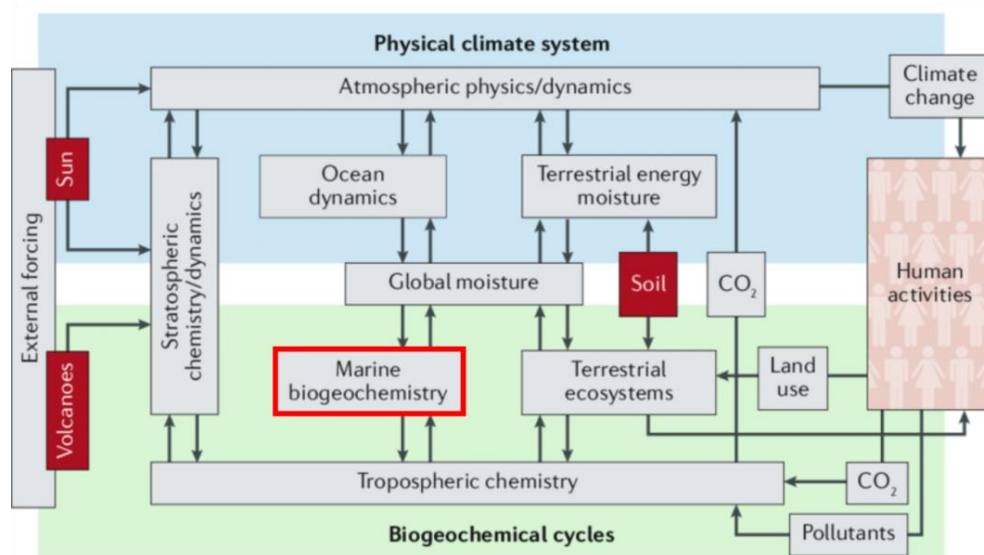


Figure 22 | Le diagramme de Betherton du système terre (NASA). Figure adaptée de Steffen *et al.*²⁸².

Les limitations par le phosphate, le nitrate et la silice (pour les diatomées) sont eux modélisés par des modèles de Monod^{60,202}. L'attachement particulier à la modélisation de la chimie complexe du fer et de ses sources dans PISCESv2 s'est démontré crucial dans la capture des patterns de limitation en nutriments et dans la distribution spatiale du fer²⁸³. Ainsi, les différents modèles biogéochimiques utilisés aujourd'hui présentent une grande variabilité dans leur représentation des compartiments biogéochimiques (modèles) et en conséquence de leur modélisation résultante (champs de sorties) comme je viens d'en prendre exemple avec la modélisation du fer²⁸³. En contraste du modèle français s'attachant particulièrement à la modélisation fine de la chimie et biochimie du fer, le modèle étasunien s'est lui attaché à une représentation plus résolue de la diversité phytoplanctonique marine et de son rôle dans la biogéochimie^{149,280}. Ainsi le modèle américain a d'abord inclu 14²⁸⁰ puis 350 types de phytoplancton dont la niche écologique est décrite par trois « traits » que sont la taille, la fonction biogéochimique et la tolérance à la température¹⁴⁹.

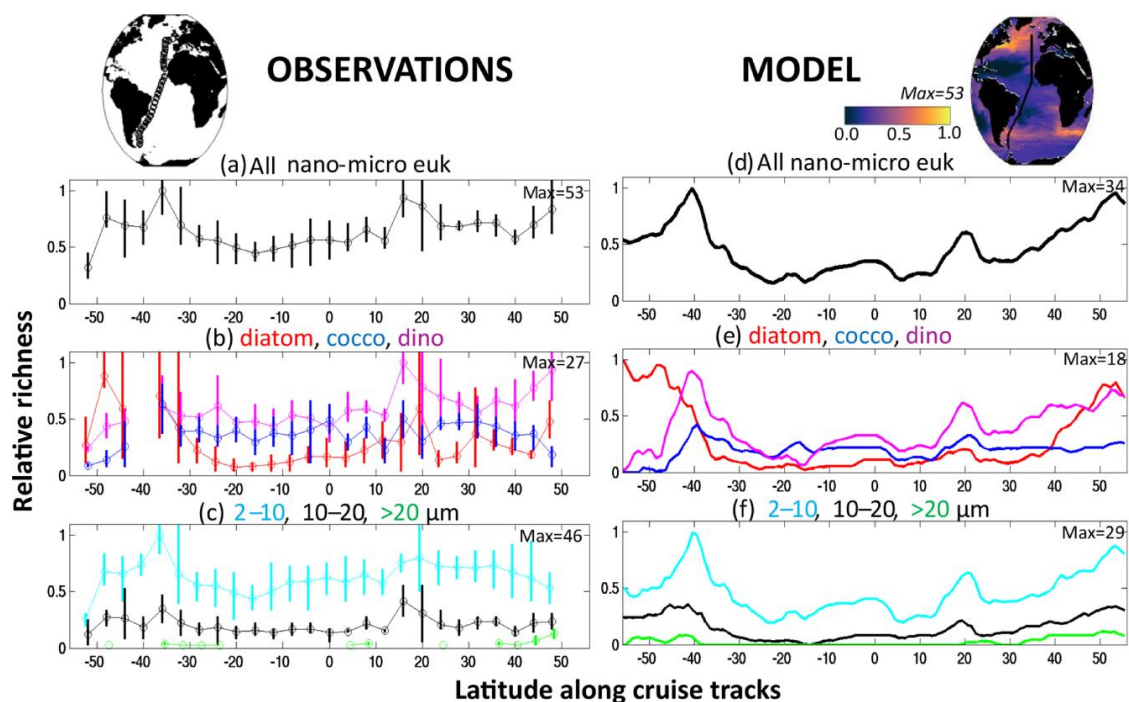


Figure 23 | Structures latitudinales globales de biodiversité tel qu'observé (à droite) et tel que modélisé par le modèle biogéochimique américain (à gauche).

Figure issue de Dutkiewicz *et al.*¹⁴⁹, © Dutkiewicz 2020.

Le modèle inclut ainsi 16 classes de tailles, 5 types de phytoplancton (les diatomées (utilisant la silice), les coccolithophores (carbonates), les mixotrophes (à la fois photosynthétique et brouteurs), les cyanobactéries fixant l'azote de l'atmosphère (diazotrophes) et le picophytoplancton) et 10 normes de niche thermique.

Cette approche a permis de démontrer que ces trois traits sont cruciaux dans la détermination de la distribution de la biodiversité phytoplanctonique dans les océans et a permis de reconstruire très correctement les structures observées de biodiversité dans l'espace comme notamment le gradient latitudinal de diversité¹⁴⁹ (Figure 23).

Pour résumer, les modèles systèmes terre sont des modèles complexes résolvant divers phénomènes physiques, chimiques et biologiques de manière variables selon les modèles et permettant de répondre à des questions variées^{149,271,283}. Une de leurs forces est leur capacité à pouvoir être extrapolés dans le futur notamment en simulant différents scénarios de changement climatique ce que je décrirai maintenant.

IV.3. Le changement climatique et sa modélisation

Le changement climatique ou réchauffement climatique est lié aux émissions de gaz à effet de serre par les activités humaines depuis la révolution industrielle de la deuxième moitié du XIX^{ème} siècle en Europe. Le principal gaz à effet de serre est le CO₂ qui est émis par la combustion des ressources naturelles de combustibles dits fossiles tels que le pétrole, le charbon et le gaz permettant des apports en énergie importants utilisés dans de nombreuses activités humaines. Comme je l'ai déjà mentionné, le pétrole est issu principalement de la décomposition de la matière organique marine dans certaines conditions de température, de pression et d'oxygénation issu de la pompe à carbone biologique décrite dans le paragraphe précédent²⁸⁴⁻²⁸⁶. La formation du pétrole peut prendre ainsi des milliers voire des millions d'années²⁸⁶, échelle de temps à contraster avec le rythme actuel auquel l'homme brûle le pétrole auquel il a accès. En effet, les réserves actuelles permettant de maintenir le rythme actuel de consommation permettrait encore 120 ans d'utilisation²⁸⁷ voir seulement 50 ans selon certaines études récentes^{288,289}. Le Groupe d'Experts Intergouvernemental sur l'évolution du Climat (GIEC) rédige ainsi des rap-

ports depuis la fin des années 1980 pour alarmer sur les conséquences de ces émissions sur le climat terrestre. Ainsi selon le dernier rapport du GIEC de 2021 sur les origines et les preuves physiques du réchauffement climatique²⁹⁰, depuis 1850, l'humanité a émis 2390 ± 240 Gt de CO₂ soit $\sim 900 \pm 90$ GtC dans l'atmosphère dont 55 ± 10 %²³⁷ a été absorbé par la pompe physico-chimique de l'océan (23 ± 5 %)²³⁷ entraînant notamment son acidification, et la pompe terrestre (29 ± 5 %)²³⁷. De manière remarquable, malgré l'augmentation des émissions atteignant 10 GtC en 2018, le CO₂ émis restant dans l'atmosphère depuis 1858 reste stable (45%)²⁹¹. Les émissions cumulatives ont entraîné un réchauffement de l'atmosphère en moyenne de 1.09 [0.95 à 1.20]°C (2011-2020 comparé à 1850-1900) dû à l'effet de serre du CO₂ et des autres gaz émis²⁹⁰ (Figure 24). Il est important de noter que l'effet de serre a été co-découvert par Eunice Foot²⁹² puis John Tyndall²⁹³ aussitôt que dans les années 1850 puis décrit comme pouvant avoir un effet sur le climat global par le chimiste Svante Arrhénius en 1896²⁹⁴. Le réchauffement moyen de l'atmosphère est plus important au niveau terrestre (+1.59°C) qu'au niveau des océans (+0.88°C)²⁹⁰. La température de surface des océans a quant à elle augmenté d'environ ~ 0.74 °C sur la période 1900-2019²⁹⁵ et les océans ont emmagasiné plus de 90% de la chaleur excédentaire liée au réchauffement climatique²⁹⁰. Parmi d'autres impacts sur les océans déjà cités ou non précédemment, on peut également se rappeler (parmi bien d'autres) la perturbation du cycle de l'azote²⁶⁸, le déplacement de nombreuses espèces planctoniques²³⁴, l'acidification des océans^{290,296}, l'augmentation significative des zones anoxiques^{290,297} ou encore le doublement des vagues de chaleur marine depuis 1980^{290,298}. Tous ces effets ont déjà bien sûr des conséquences directes sur les activités humaines marines notamment la pêche^{299,300} dont 3 milliards d'humains dépendent³⁰¹, la recherche de solutions d'adaptation et de mitigation des effets^{302,303} ou encore des changements politiques récents³⁰⁴. C'est principalement ici que les modèles système Terre entrent en jeu et permettent une vision à long terme des potentiels impacts du changement climatique sur le climat terrestre selon différents

sécenarios d'émissions de gaz à effet de serre. Je me concentrerai sur les effets projetés dans les océans.

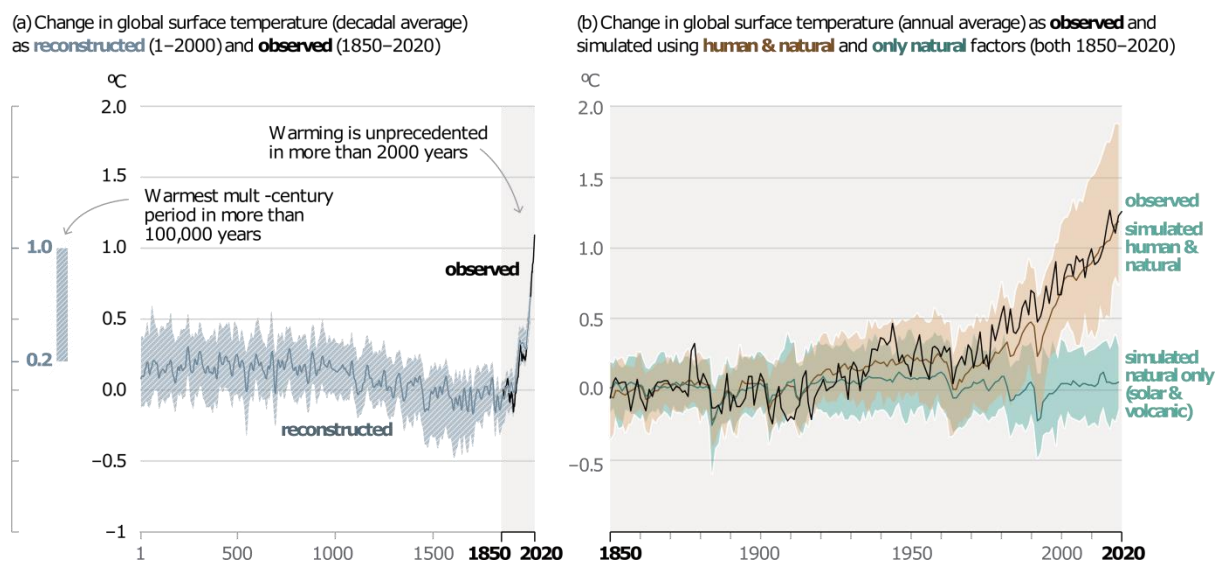


Figure 24 | Changements de la température de surface terrestre relatifs à la période 1850-1900²⁹⁰. Figure issue de Masson-Delmotte *et al.*²⁹⁰.

Comme je l'ai montré avec la Figure 22, les modèles système Terre incluent les émissions en gaz carbonique lié aux activités humaines ainsi que les changements d'usage des terres (déforestation/reforestation, urbanisation, augmentation de la surface agricole...). Ainsi, selon les politiques entreprises par les États, les scénarios géopolitiques et socio-économiques ou encore les avancées technologiques, l'humanité a la possibilité de suivre différents scénarios d'émission de gaz à effets de serre appelé « Representative Concentration Pathways » (RCPs) dans les modèles utilisés pour les prédictions du rapport 2013³⁰⁵ du GIEC et « Shared Socio-economic Pathways » (SSPs) dans le rapport plus récent de 2021²⁹⁵. Je décrirai ici les RCPs, utilisés dans les articles 2 et 3 du chapitre 1, ayant un recul insuffisant sur les nouveaux modèles suivant les SSPs. Il a été distingué ainsi 4 RCPs, respectivement le RCP2.6, RCP4.5, RCP6 et RCP8.5 pour lesquels l'atmosphère atteint d'ici 2100 une concentration en équivalent CO₂ (c'est-à-dire en incluant aussi les effets de serre des autres gaz) allant de 490 ppm (RCP2.6) à 1370 ppm (RCP8.5)⁶¹. La concentration 2006 en équivalent CO₂ de l'atmosphère était ainsi de 393 ppm³⁰⁶ en incluant les effets refroidissant des aérosols et a atteint 460 ppm en 2019³⁰⁷. Le scénario RCP2.6 prévoit

ainsi un pic d'émission vers 2020 et une neutralité carbone vers le début des années 2070⁶¹. Ainsi ce scénario peut être considéré comme à oublier selon la situation actuelle. En effet, bien que certains États s'engagent à la neutralité carbone d'ici 2050, les émissions continuent d'augmenter. Du point de vue des océans, les impacts moyens des différents scénarios tels que modélisés par un modèle moyen de plusieurs modèles système Terre sur la température de surface, le pH, la production primaire nette et la teneur en oxygène des océans sont résumés par la figure 25. Ainsi pour le modèle de réchauffement climatique important (RCP8.5), la température de surface des océans est projetée d'augmenter de $2.73 (\pm 0.72)^\circ\text{C}$, le pH de diminuer de $-0.33 (\pm 0.003)$ unités de pH, l'oxygène global de diminuer de $-3.45 (\pm 0.44) \%$ et la production primaire de diminuer de $-8.6 (\pm 7.9) \%$ ³⁰⁸.

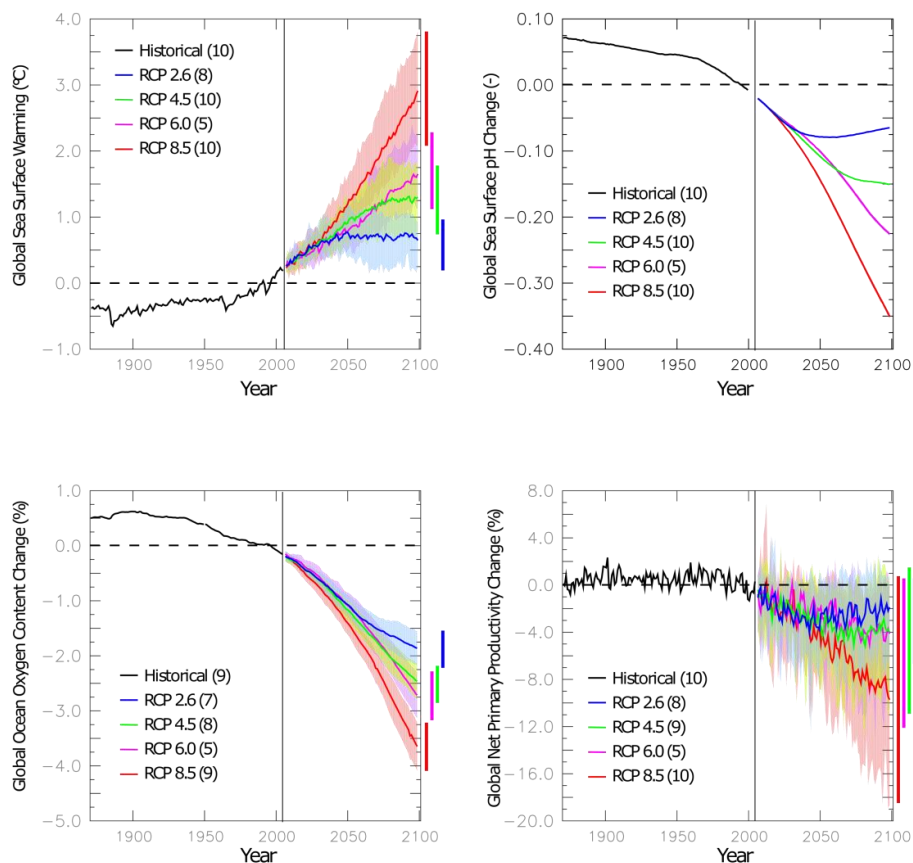


Figure 25 | Séries temporelles d'un modèle climatique moyen de quatre paramètres physico-chimiques des océans impacté par le réchauffement climatique³⁰⁸. Figure issue de Bopp *et al.*³⁰⁸, © Bopp 2013. Toutes les variables sont visualisées par rapport à la période 1990-99 et les zones assombries indiquent

l'écart-type lié aux écarts entre modèles.

J'utilise ce scénario avec un modèle moyen dans les articles 2 et 3 du chapitre I en incluant les changements projetés en température, salinité, macronutriments (nitrate, phosphate et silice), fer et saisonnalité des nutriments pour projeter la biogéographie des communautés planctoniques d'ici la fin du siècle. Pour résumer ce dernier paragraphe de l'introduction, le réchauffement climatique apparaît comme étant le problème le plus important et le plus urgent à résoudre dans ce siècle pour l'humanité et se trouve au cœur de nombreuses recherches scientifiques aussi bien sur l'écosystème océanique que terrestre. Il représente un défi d'envergure pour les gouvernements et populations actuels et futurs de tous les pays du globe. En effet, les conséquences sur l'humanité du réchauffement climatique d'ici la fin de siècle, si l'on suit les pires scénarios, seront dramatiques. Selon le scénario RCP8.5 et les scénarios démographiques actuels, une étude récente projette ainsi que 3.5 milliards d'humains pourraient habiter d'ici 2070 (ou non puisqu'ils mourront ou migreront) des zones atteignant des températures atmosphériques de surface considérées comme mortelles pour l'homme (températures moyennes annuelles supérieures à 29°C) et ce principalement dans les régions de forte pauvreté³⁰⁹.

DEMARCHE SCIENTIFIQUE

La démarche conceptuelle et scientifique de cette thèse a consisté dès son origine à explorer les différentes réponses possibles des organismes planctoniques à un changement environnemental en l'occurrence le réchauffement climatique. On peut résumer ces différentes réponses possibles par le diagramme présenté en figure 26. Comme je l'ai brièvement évoqué dans la partie III.3. Les facteurs limitant la croissance du phytoplancton : température, taille et nutriments, courants et lumière, j'ai identifié 4 réponses possibles du vivant (ceux-ci n'étant pas exclusifs au plancton) à un changement environnemental que nous décrivons maintenant. L'idée de la thèse a été d'explorer chez les communautés de plancton indépendamment les 3 premières possibilités de réponse que sont la migration, l'acclimatation et l'adaptation. Comme décrit dans la partie IV de l'introduction, les rétroactions sur le

climat de la réponse du plancton au réchauffement climatique ont aussi été explorées (chapitre I, article 2).

La première possibilité qu'a un organisme pour répondre à un changement environnemental est la migration qui consiste à se déplacer (passivement pour le plancton) vers des zones géographiques possédant les mêmes caractéristiques que son environnement initial. Il s'agit donc pour l'organisme de suivre sa niche écologique originale en suivant les changements de température comme cela a déjà été mis en évidence pour certaines espèces planctoniques^{234,310} bien que ce ne soit pas tout le temps le cas^{310,311}. Dans le cas où l'organisme ou la communauté d'organismes ne fait que migrer, on peut considérer que son génome est constant (à l'exception des mutations neutres) (Figure 26 : on part d'un génome G_0 à T_0 et on arrive à un génome G_0 à T_1). Le premier chapitre de la thèse est en grande partie consacré à l'exploration de ce mécanisme de migration (passive) des communautés et organismes planctoniques à l'horizon 2100 en réponse au réchauffement climatique. Le second mécanisme de réponse à un changement environnemental est l'acclimatation³¹² qui consiste en la régulation de l'expression des gènes d'un organisme ou d'une communauté d'organismes permettant une adaptation physiologique. Dans ce cas le génome reste aussi constant (Figure 26). Ainsi les données de *Tara* Océans par approche métatranscriptomique ont pu mettre en évidence des changements d'expression liés aux conditions environnementales tels que la disponibilité en fer^{126,127} ou encore en nitrate¹²⁸. Dans le chapitre 2, je m'intéresserai particulièrement à l'acclimatation des communautés de plancton le long de la transition entre l'Atlantique Nord et l'océan Arctique qui constitue un changement environnemental important avec une forte diminution des températures notamment. La troisième réponse possible des organismes aux changements environnementaux est l'adaptation génétique. Dans ce cas, la séquence du génome de l'organisme en question est modifiée. Ainsi, suite au changement environnemental le génome G_0 devient le génome G_1 (Figure 26). Il s'agit ici du bien connu moteur de l'évolution dont nous devons l'intuition à Charles Darwin dans son

œuvre *L'origine des espèces*³¹³. Le mécanisme de mutation génétique à l'œuvre dans la sélection naturelle ne sera que décrit au début du XX^{ème} siècle. Les lois de Mendel³¹⁴, la théorie chromosomique de l'hérédité par Morgan^{315,316}, et enfin la théorie synthétique de l'évolution³¹⁷ sont ainsi élaborés et constituent la base de nos connaissances modernes bien que celles-ci ne cessent d'évoluer³¹⁸. Au sein du plancton, certaines recherches récentes mettent ainsi en évidence des mutations locales permettant l'adaptation au sein de mêmes espèces à différents environnements¹⁹⁶⁻¹⁹⁸ ou encore l'adaptation de la niche écologique de certaines espèces phytoplanctoniques³¹¹. Dans le chapitre 3, je développerai les prémices d'un modèle de croissance du phytoplancton de type Modèle Basé Individu (IBM) dont l'une des finalités sera d'explorer l'influence des mécanismes d'adaptation sur la biogéographie du plancton en réponse à un changement environnemental tel que le réchauffement climatique. Enfin, la dernière réponse possible d'une espèce à un changement environnemental est l'extinction.

Ce phénomène touche de nombreuses espèces aujourd'hui. Il est en lien direct avec le réchauffement climatique et les pressions anthropogéniques³¹⁹. A plus long terme, cette crise de la biodiversité pourrait conduire à la sixième extinction de masse sur Terre et serait déjà en cours³²⁰. Je n'explorerai pas ce dernier mécanisme dans cette thèse.

Par soucis de simplicité, j'ai fait le choix pour cette thèse d'explorer indépendamment les uns des autres les différents mécanismes de réponse des organismes planctoniques aux changements environnementaux. La méthode scientifique s'attache ainsi de manière générale à décomposer les problèmes en sous-problèmes. Néanmoins, de manière plus vraisemblable les communautés de plancton répondent au réchauffement climatique en combinant ces différents mécanismes. Par exemple le mécanisme d'expression spécifique d'un allèle combine acclimatation puis adaptation¹⁹⁸. Une grande question sous-jacente à cette thèse est ainsi le futur de nos écosystèmes, marins dans notre cas, leur biodiversité et leur capacité d'adaptation face au réchauffement climatique.

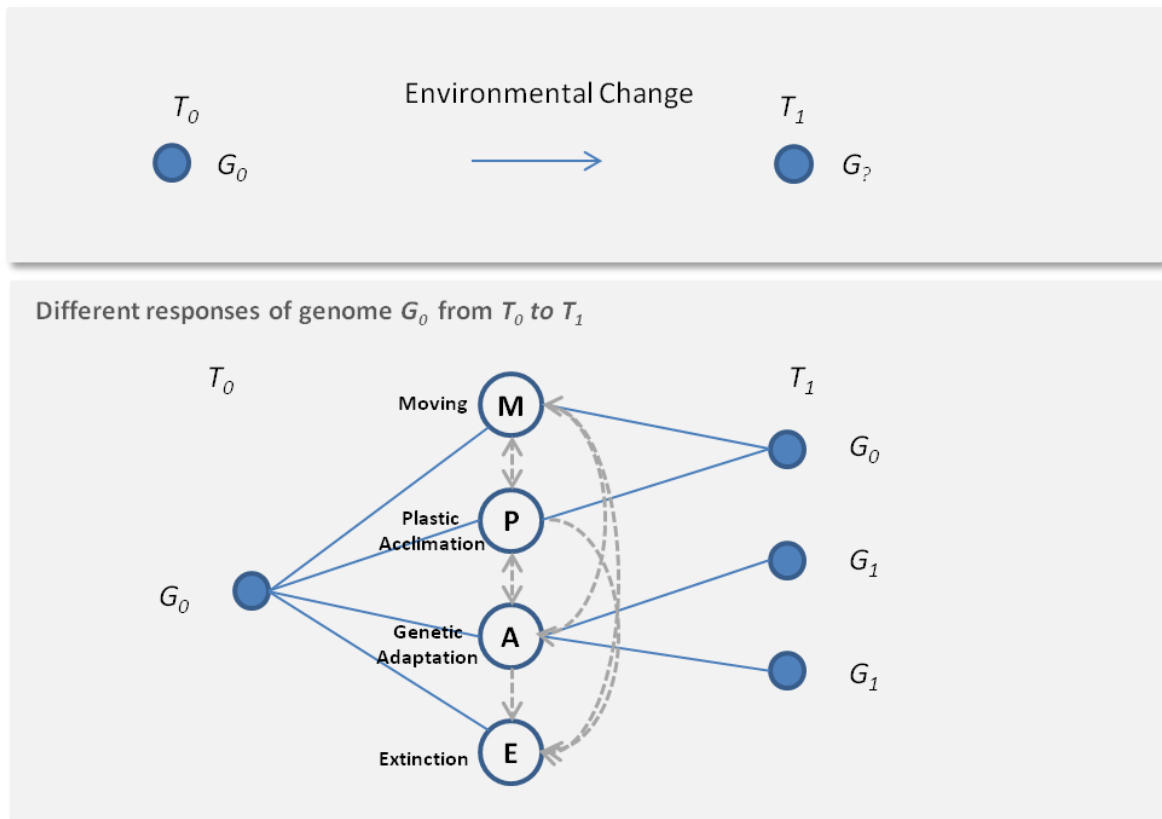


Figure 26 | Schéma conceptuel de la réponse d'un organisme et de son génome G_0 à un changement environnemental entre les temps T_0 et T_1 .

CHAPITRE 1 : Etude de la biogéographie génomique du plancton : changement climatique et cycles biogéochimiques

I. Introduction et résumé: une biogéographie génomique

Le premier chapitre de cette thèse est consacré à mes résultats portant sur la biogéographie du plancton basé sur la génomique et sa réponse au réchauffement climatique. Il comporte 3 articles complémentaires et répondant à différents aspects de la question que j'introduirai maintenant. Dans le contexte des résultats issus des expéditions *Tara Océans*, ces résultats s'inscrivent dans l'analyse des données métagénomiques de l'expédition. Dans un contexte plus large ils s'inscrivent dans la révolution permise par les données omiques environnementales, les progrès de séquençage de l'ADN, notamment par les technologies NGS (Next Generation Sequencing) mais aussi dans les recherches sur le réchauffement climatique. Les trois articles forment un tout cohérent et sont tout trois basés sur le même jeu de données : les données métagénomiques des expéditions *Tara Océans*. En quelques mots, le premier article explore les structures biogéographiques issues des données métagénomiques, le second extrapole ces structures à des biogéographies sur l'ensemble des océans et évalue les effets du changement climatique sur celles-ci ; enfin le troisième génère une base de données de génomes issus de l'assemblage des données métagénomiques et explore leurs biogéographies ainsi que leurs réponses au réchauffement climatique. Les résultats du chapitre sont résumés par l'article des pages suivantes qui sera publié dans *Les Comptes rendus Biologies de l'Académie des Sciences*. Etant uniquement co-auteurs des articles 1 et 3, j'ai fait le choix de ne présenter ici que leurs abstracts. Ces articles sont accessibles gratuitement en ligne.

Biogéographie du plancton au XXI^e siècle et impacts du changement climatique: avancées par la génomique

Paul Frémont^{1,2}, Marion Gehlen³, Olivier Jaillon^{1,2}

¹Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, Evry, France. ²Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans, Paris, France. ³LSCE-IPSL, CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France.

Correspondance : pfremond@genoscope.cns.fr, mgehlen@lsce.ipsl.fr, ojaillon@genoscope.cns.fr

Résumé : Cet article résume des avancées récentes dans nos connaissances sur la biogéographie du plancton obtenues par des approches de génomiques ainsi que les impacts du réchauffement climatique sur celle-ci. La comparaison à large échelle du contenu génomique d'échantillons de différentes tailles de plancton a révélé un partitionnement des océans en provinces génomiques ainsi que l'impact des principaux courants océaniques sur celles-ci. En définissant des niches écologiques, ces provinces sont extrapolées à l'ensemble des océans à l'exception de l'océan Arctique. D'ici la fin du XXI^e siècle, une restructuration majeure de ces provinces est projetée en réponse à un scénario de fortes émissions de gaz à effet de serre sur 50% de la surface des océans étudiés. Une telle restructuration pourrait engendrer une diminution de 4% de la production exportée. L'obtention de séquences assemblées d'un grand nombre de génomes de plancton définissant cette biogéographie a permis de mieux caractériser le contenu génomique des provinces et d'identifier les espèces les structurant. Ces génomes ont aussi permis de mieux caractériser les futurs changements potentiels de communautés de plancton sous changement climatique.

Mots clés : biogéographie du plancton, génomique, changement climatique

Paul Frémont est lauréat 2022 du Prix « Les Grandes Avancées Françaises en Biologie » de l'Académie des Sciences

La notion de biogéographie a été introduite par le naturaliste Allemand Alexander Von Humboldt au XIX^{ème} siècle. Il décrit, suite à son voyage en Amérique du sud, dans son ouvrage *Kosmos : Entwurf einer physischen Weltbeschreibung* [1] de nombreuses espèces végétales et fait pour la première fois le lien entre leur répartition géographique et le climat associé [1]. La biogéographie consiste ainsi en l'étude de la répartition spatiale et temporelle des espèces aussi bien actuelle, à travers les temps géologiques et plus récemment en réponse au changement climatique. Au cœur de cette discipline, on retrouve la notion de niche écologique dont on doit l'intuition à Humboldt. Elle est mathématisée par le zoologiste George Evelyn Hutchinson dans les années 1950 [2] et se retrouve au centre des recherches actuelles de l'écologie contemporaine [3]. La niche écologique d'une espèce ou communauté d'espèces est ainsi conceptualisée comme l'hypervolume des paramètres environnementaux permettant la vie de cette ou ces espèces [2].

Bien qu'elle ait d'abord été définie pour les organismes terrestres, on peut définir une niche écologique pour les organismes marins et en particulier pour le plancton qui nous intéresse

ici. Mais pourquoi s'intéresser au plancton ? Plusieurs facteurs d'importance nous poussent aujourd'hui à étudier ces microorganismes aquatiques. Tout d'abord, le phytoplancton, réalise la photosynthèse et contribue ainsi fortement à la production primaire réalisée dans les océans, quasi équivalente en masse à la production terrestre [4]. Ainsi, l'ensemble des réseaux trophiques dépend de cette production primaire, incluant les bactéries reminéralisatrices, les grands mammifères marins (baleines) mais aussi les hommes (pêcheries) [5]. Ensuite le plancton est en partie à l'origine de la pompe à carbone biologique c'est-à-dire l'export de carbone organique vers les fonds océaniques et son stockage dans les sédiments [6], [7]. Ce phénomène se produit *via* de multiples processus physiques (agrégations particulaires) et biologiques (blooms, boulettes fécales du zooplancton) [6], [7]. Enfin, la structuration spatiale du plancton dans les océans, sa biogéographie, est une question scientifique cruciale et fondamentale [8] en raison, entre autres, du transport par les courants à travers les océans. Cette question liée à celle des niches écologiques du plancton est encore très débattue parmi la communauté scientifique. Aujourd'hui dans le contexte du changement climatique, ces processus pivots pour le climat terrestre sont mis en danger et il est plus que jamais nécessaire de mieux les comprendre pour mieux anticiper et gérer les changements globaux à venir. Certaines expéditions du XIX^{ème} siècle ont inclus des objectifs scientifiques [9]. Cependant les écosystèmes marins sont restés les moins compris et étudiés, ce qui a conduit relativement récemment au lancement d'expéditions contemporaines [10]–[12]. Ces récentes expéditions sont concomitantes de la révolution omique qui permet un séquençage massif d'ADN de plancton, soit à partir de culture soit à partir de prélèvements d'eau filtrée, ou non, par métagénomique. Cette approche a largement enrichi nos connaissances sur le plancton marin [10]–[13].

Dans cet article, nous faisons la synthèse d'avancées récentes sur nos connaissances de la biogéographie du plancton, qui ont été permises par la production et l'analyse des données métagénomique des expéditions *Tara Océans* (2009-2013) à travers les résultats de trois articles publiés cette année 2022 [14]–[16].

Le premier article Richter *et al.* [14], fondateur, porte sur les structures biogéographiques à l'échelle de l'ensemble des océans du globe (à l'exception de l'Arctique¹) observés à partir des données de métagénomiques, de metabarcoding et d'imagerie des expéditions *Tara Océans* de 2009 à 2012. L'article est centré sur le lien entre les structures biogéographiques et les courants de large échelle et, de manière plus secondaire, les paramètres environnementaux.

Les structures biogéographiques observées sont basées sur des estimations de bêta-diversité c'est-à-dire le niveau de similarité génomique entre deux communautés. Des estimations ont été réalisées à 3 échelles différentes entre chaque paire d'échantillons d'une même fraction de taille de plancton de l'expédition: au niveau métagénomique, au niveau taxonomique par l'approche de metabarcoding, et taxonomique par classification morphométrique d'images. Pour la métagénomique, un algorithme spécifique appelé Simka [17] a été développé par une équipe de l'INRIA de Rennes pour estimer la bêta-diversité par comparaison du contenu en k-mer (sous partie de k nucléotides de la séquence d'ADN) des paires d'échantillons. Ces

¹ L'Arctique n'a pas été inclut à cette étude car les échantillons de l'expédition *Tara Polar Circle* n'était pas encore séquencés

données de bêta-diversité ont été comparées avec les temps de parcours minimaux suivis par les courants entre stations calculés à partir du modèle de circulation MITgcm [18]. En analysant la corrélation de façon cumulative, c'est-à-dire en comparant les échantillons uniquement connectés par un temps $t \leq t_0$ (en augmentant petit à petit le t_0), un maximum de corrélation à 1,5 ans de transport ($t_0=1.5$ an) entre la bêta-diversité et le temps de transport est apparu pour chacune des six fractions de taille de plancton allant des virus (0-0.2 μm) aux petits métazoaires (180-2000 μm). Ce temps de parcours de 1,5 an correspond au temps caractéristique de parcours du courant de surface, et donc du plancton, au sein des bassins océaniques étudiés. La bêta-diversité des petites fractions de taille (virus 0-0.2 μm , bactéries 0.22-3 μm et protistes 0.8-5 et 5-20 μm) s'est avérée être corrélée plus fortement à ces temps de transports appuyant l'hypothèse d'une plus grande dépendance de ces organismes aux courants (sans pour autant signifier que ces organismes soient transportés à des vitesses différentes).

Les données de metabarcoding se sont avérées être relativement moins bien corrélées aux courants que les données de métagénomique et encore moins celle d'imagerie. Ce résultat montre la capacité de la métagénomique (séquençage aléatoire de fragments d'ADN) à détecter à haute résolution des structures biogéographiques. La métagénomique a une plus grande résolution que les autres méthodes car elle se base sur la comparaison de millions de k-mers, le metabarcoding sur quelques centaines de milliers d'Unités Taxonomiques Opérationnelles (OTUs) et dont la résolution est très inégale selon les taxons avec le risque de confondre certaines espèces, et l'imagerie (ici les larges fractions de taille) seulement sur quelques centaines de taxons résolus avec un risque important de confusion d'espèces. Malgré les progrès de séquençage, la métagénomique fournit très certainement une très petite proportion de l'ADN des communautés, et donc fournit de l'information qui provient majoritairement des organismes les plus abondants. Il en résulte qu'une information à très haute résolution d'un petit nombre d'organismes abondants apporte un signal biogéographique plus fin qu'une description taxonomique floue d'une grande partie de la diversité existante.

L'article fait aussi le lien entre bêta-diversité et facteurs environnementaux, montrant une forte corrélation entre les concentrations en nutriments et la biogéographie des petits organismes jusqu'à 1,5 année de transport tandis que l'influence de la température se mesure à plus large échelle. Enfin, en groupant les sites d'échantillonnages en fonction de leur bêta-diversité à l'aide de l'algorithme UPGMA [19] (Unweighted Pair group Method with Arithmetic mean), l'étude a permis de définir des provinces génomiques pour chaque fraction de taille de plancton évoquée précédemment. Ces provinces sont donc des zones géographiques, non nécessairement continues, contenant des organismes parmi les plus abondants et qui ont une similarité génomique très importante, très probablement des espèces identiques. Ce résultat est fort et nouveau car il démontre un partitionnement des océans basés pour la première fois sur la génomique. Ce partitionnement est probablement stable en raison de la façon dont l'échantillonnage a été réalisé et malgré le fait que les échantillons étudiés ne soient issus que d'un unique prélèvement temporel à chaque site. En effet, alors qu'il existe une bonne corrélation entre beta diversité et transport par les courants, il faut noter que les temps de transport sont des résultats de modèles, alors que le

parcours du bateau n'a pas suivi les courants ni dans le temps ni l'espace. Ainsi, les provinces ont une couverture moyenne en saisons et en mois relativement importante (3 saisons et 5 mois en moyenne pour une province). A noter aussi que ce constat n'exclut pas une certaine variabilité saisonnière des provinces dans leur contenu biologique et dans leur topologie. Les provinces sont à la base de l'article suivant qui extrapole ces provinces génomiques à l'ensemble des océans (à l'exception de l'Arctique). De manière intéressante, cette première étude montre aussi que ce partitionnement est variable selon les fractions de taille de plancton avec de plus petites provinces pour les petits organismes que pour les grands organismes.

Le second article Frémont *et al.* [15] extrapole à l'ensemble des océans les provinces génomiques du premier article en définissant les niches environnementales de celles-ci, étudie l'impact du changement climatique sur celles-ci (déplacement, expansion/rétrécissement) et enfin analyse les potentiels impacts écologiques et biogéochimiques de cette réorganisation.

A partir d'un sous ensemble de 38 des 48 provinces génomiques originales du premier article, cette seconde étude a dans un premier temps mis en évidence, à l'aide de techniques de machine learning (randomforest [20], boosted regression trees [21], single hidden layer neural network [22] et generalized additive models[23]), que 27 d'entre elles correspondent à des niches environnementales définies statistiquement de manière robuste. Ce résultat a été permis par l'association statistique entre les provinces et les paramètres environnementaux les caractérisant: la température, la salinité, trois macronutriments (nitrate, phosphate et silice dissous), le fer et un indice de saisonnalité du nitrate. Le choix de ces paramètres est basé sur les analyses du premier article [14], bien que d'autres paramètres tels que la profondeur de la couche de mélange ou la durée du jour auraient aussi pu être utilisés. Dans cet article les provinces ont ainsi été appelées provinces climato-génomiques du fait qu'elles sont caractérisées par leur contenu génomique et leur climat physico-chimique. Un second résultat majeur est la mise en évidence de génomes signatures (issus du troisième article décrit plus bas pour les eucaryotes [16] et d'un autre article pour les procaryotes [24]) pour ces provinces dont les biogéographies coïncident fortement avec les provinces elles-mêmes.

Cette structuration constitue un résultat fort et inconnu auparavant. En effet il montre que la structuration en provinces génomiques relativement abstraites initialement (car basées uniquement sur la comparaison des k-mers issus des métagénomés) s'explique par la présence de génomes permettant de documenter la nature des similarités métagénomiques. A l'aide des climatologies globales du World Ocean Atlas 2013 [25] des différents paramètres environnementaux choisis, les différentes niches environnementales ont ensuite été extrapolées à l'ensemble des océans (excepté l'arctique) définissant leur partitionnement pour les six fractions de tailles de plancton utilisées (Figure 2 de Frémont *et al.* [15]). L'extrapolation à l'ensemble des océans a confirmé les différences des biogéographies des petits et grands organismes, impliquant un découplage trophique, et a aussi permis de montrer une différence importante entre les partitionnements génomiques et les partitionnements biogéographiques connus [26]–[28] (suppléments de l'article). Néanmoins ce résultat est sûrement à nuancer au sens où la structuration biogéographique par les

provinces climato-génomiques décrites n'est issue que d'un échantillonnage partiel des océans avec une distance moyenne de 300 km entre deux échantillons proximaux. Ce résultat appelle donc à de nouvelles études sur la biogéographie du plancton avec des prélèvements répétés dans le temps ainsi qu'à des échelles plus résolutive spatialement comme pourrait le permettre par exemple l'échantillonnage issu de futures expéditions [29]. Le second volet de l'article a consisté en l'analyse de l'impact du changement climatique sur l'organisation des provinces génomique à l'aide d'un modèle système Terre moyen à l'horizon de la fin du siècle selon le scénario RCP8.5 (Representative Concentration Pathway) d'émission importante de gaz à effet de serre. Les modèles systèmes Terre permettent de projeter, selon des lois physiques et chimiques, la dynamique du système climatique et des grands cycles biogéochimiques (par exemple du carbone, de l'azote et du phosphore) en réponse à différents scénarios d'émission de gaz à effet de serre. Selon le scénario RCP8.5, l'atmosphère atteindrait des concentrations en CO₂ supérieures à 1000 ppm (parties par millions) d'ici 2100. Ces projections permettent d'appréhender l'évolution des conditions physico-chimiques de l'océan sous changement climatique. A l'aide de ces projections, il a été possible d'extrapoler les provinces génomiques actuelles à l'horizon 2100 et comparer leurs distributions présentes et futures. Les taux de déplacement des provinces trouvés sont comparables à ceux existant dans la littérature pour le phytoplancton notamment [30] et les provinces tropicales s'étendent au détriment des provinces tempérées. Environ 50% de l'aire des océans de surface considérés est projetée comme pouvant subir un changement important de communauté planctonique. Ce chiffre est à nuancer, il pourrait être sûrement bien plus élevé. En effet dans de nombreuses zones, les paramètres environnementaux projetés d'ici la fin du siècle dépassent les valeurs extrêmes présentes (tropiques et antarctique notamment, voir suppléments de l'article) donc en dehors du domaine de validité des techniques de machine learning (nos méthodes ne peuvent pas projeter de changements dans ces zones). Ainsi, ce chiffre de 50% est relatif aux provinces actuelles observées, c'est-à-dire que 50% de l'océan considéré est projeté comme pouvant à l'avenir être colonisé par des communautés de plancton existants aujourd'hui. La réponse biologique dans les zones dépassant les extrêmes actuels semble très incertaine et de nouvelles pressions de sélection y sont attendues. De manière importante, les zones de pêches et zones économiques exclusives sont largement comprises dans la réorganisation des provinces génomiques. La restructuration des provinces s'accompagne aussi de changements compositionnels des communautés de plancton en taille et taxonomie (analyse basée sur l'abondance relative des génomes du troisième article [16]). L'analyse a été restreinte à trois grands groupes de plancton : les copépodes, les phototrophes et les diazotrophes (organismes capables de métaboliser le diazote atmosphérique (N₂)). Les changements projetés dans ces groupes pourraient notamment créer de nouvelles interactions trophiques proie/prédateur d'ici la fin du siècle.

L'étude s'est aussi attachée aux impacts potentiels de la réorganisation de la biogéographie du plancton sur les cycles biogéochimiques du carbone et de l'azote. Tout d'abord, une augmentation des diazotrophes dans l'océan pacifique tropico-équatorial est projetée en accord avec d'autres études projetant une augmentation de la diazotrophie dans ces zones [31]. Enfin, la structuration biogéographique en provinces et particulièrement l'assemblage

de ces provinces (à travers les six fractions de taille) s'est aussi avérée être un bon indicateur statistique de l'export de carbone (à 100m de profondeur) vers les fonds océaniques. Ainsi, une diminution moyenne de 4% de la pompe biologique est projetée à l'horizon 2100 et est mise en lien avec les changements de communautés. Ce résultat fort est en accord avec les modèles actuels [32] et aurait un effet de rétroaction positive (c'est-à-dire aggravant) sur le réchauffement climatique.

Pour résumer, cette étude a permis, pour la première fois, de générer une biogéographie génomique du plancton à l'échelle globale (à l'exception de l'océan Arctique). De plus, elle a permis d'évaluer les conséquences du changement climatique sur les communautés de plancton à l'horizon 2100 selon un scénario d'importantes émissions de gaz à effet de serre. De manière intéressante, la génomique va dans le même sens que de nombreuses projections des modèles mécanistiques actuels, démontrant son potentiel et confortant la force de ces modèles. Une partie des analyses de cette étude a été permise par l'étude du troisième article que nous présenterons maintenant.

Le troisième article Delmont *et al.* [16] présente une collection de Génomes Assemblés par Métagénomes (MAGs) eucaryotes issue des données métagénomiques des expéditions *Tara Océans* (2009-2013). Le principal résultat de cet article est la création d'une collection de 683 MAGs et 30 SAGs (génomes cellule unique) eucaryotes à partir des données métagénomiques de 5 fractions de taille enrichies en eucaryotes des expéditions *Tara Océans* (de 0.8 μm à 2 mm). L'assemblage par l'algorithme MEGAHIT [33] a utilisé au total 280 milliards de lectures de séquençage Illumina issus de 939 métagénomes échantillonnés à travers 143 sites et couvrant une grande partie des océans mondiaux (incluant les échantillons de l'océan l'Arctique issu de l'expédition *Tara Polar Circle*). Cependant la séquence de ces génomes est souvent incomplète, les génomes générés sont estimés complets en moyenne à $40 \pm 24 \%$ (allant jusqu'à 93.7%), avec des tailles allant de 10 Mbp (Méga paire de bases) à 1.3 Gbp (Giga paire de bases). Ils correspondent au total à 39.1 milliards des 280 milliards de lectures soumises à l'assembleur (14.0%). L'assembleur ne pouvant absorber d'un seul coup l'ensemble des données métagénomique (limite mémoire), les MAGs ont été obtenus par 11 sous-ensembles de données métagénomiques qui ont été chacun co-assemblés (contenant chacun entre 6 et 38 milliards de lectures). Ces ensembles correspondent à des sites de prélèvement répartis de manière cohérente géographiquement.

Dans cette collection de MAGs, un ensemble de 10 millions de gènes a été identifié. De manière novatrice, elle couvre un ensemble de lignées eucaryotes mal caractérisées et étendant largement les efforts de culture de micro-organismes marins principalement côtiers. Une phylogénie de la collection a été générée à partir de l'ADN codant pour l'ARN polymérase et regroupe la plupart des grands groupes d'eucaryotes microbiens marins connus, tels que les Opisthocontes (principalement des copépodes), les Archaeplastida, les Straménopiles, les Alvéolés, les Rhizaria (les trois derniers formant le supergroupe SAR), les Cryptista, les Excavata, les Haptista et un nouveau groupe putatif. On note toutefois l'absence des Dinoflagellés, groupe très important et abondant mais aux génomes très grands [34]. Un effort de séquençage bien plus important serait nécessaire pour l'assemblage du génome de ces organismes. Un classement des génomes a aussi été réalisé

sur la base de l'annotation fonctionnelle des gènes les constituant. Ce classement a notamment révélé des convergences fonctionnelles entre groupes distants phylogénétiquement comme les diatomées et d'autres algues telles que les Mamiellales. Enfin, la modélisation des niches écologiques des 713 génomes ont été testées et, au total, 374 niches robustes statistiquement (52%) ont pu être définies. La même méthodologie et les mêmes paramètres environnementaux que pour le second article [15] a été utilisée pour cette analyse.

L'ensemble des résultats de cette analyse et les projections climatiques à la fin du siècle ont été compilés sur un site internet ([http://end.mio.osupytheas.fr/Ecological Niche database/](http://end.mio.osupytheas.fr/Ecological_Niche_database/)). Il est hébergé à l'institut océanographique de Marseille (MIO). Le MAG dont la niche s'étend le plus d'ici la fin du siècle correspond à un copépode tropico-équatorial, de même que le MAG dont la niche se rétrécit le plus. Ces résultats vont dans le sens d'une réorganisation potentielle forte de la chaîne trophique marine des océans globaux. Le site comporte aussi un onglet présentant les climatologies des différents paramètres environnementaux ainsi qu'un onglet permettant à l'utilisateur de définir des niches environnementales virtuelles et de les visualiser au temps présent et à la fin du siècle.

Pour résumer, ces trois articles basés sur le même jeu de données ont permis d'incrémenter significativement les connaissances sur la biogéographie du plancton grâce à la métagénomique. Ainsi, les génomes et provinces structurant spatialement le plancton dans les océans étaient jusqu'alors inconnus. Ce résultat est nouveau par sa résolution fine liée à la génomique et le découplage partiel entre les biogéographies des différentes fractions de taille de plancton. Ainsi il existe un découplage au niveau trophique entre proies (phytoplancton) et prédateurs (zooplancton) à l'échelle des océans globaux. L'impact potentiel du réchauffement climatique sur la biogéographie des différents organismes planctoniques a été évalué d'ici la fin du siècle dans un scénario d'émissions importantes de gaz à effet de serre. Il est ainsi projeté que les provinces climato-génomiques actuelles pourraient se réorganiser sur environ 50 % des océans (hors océan Arctique) d'ici 2100. De plus, les provinces tropicales s'étendraient aux détriments des provinces tempérées. De nombreuses zones océaniques futures pourraient atteindre des climats extrêmes dépassant les valeurs maximales (ou minimales) actuelles de température et concentrations en nutriments. Dans ces zones, les provinces actuelles sont maintenues, cependant on peut s'attendre à la disparition de certaines niches et à l'apparition de nouvelles communautés sous pression de sélection inconnue à ce jour. Enfin, la restructuration de la biogéographie du plancton pourrait engendrer une diminution de 4% de la pompe à carbone biologique entraînant un aggravement du réchauffement climatique.

References

- [1] A. von Humboldt, Kosmos : Entwurf einer physischen Weltbeschreibung. Ed. Stuttgart/Tübingen, Cotta, 1847.
- [2] G. E. Hutchinson, Concluding remarks, *Cold Spring Harbor Symposia on Quantitative Biology*, 1957.

- [3] A. Pocheville, The Ecological Niche: History and Recent Controversies , in *Handbook of Evolutionary Thinking in the Sciences*, P. and L. G. and S. M. Heams Thomas and Huneman, Éd. Dordrecht: Springer Netherlands, 2015, p. 547-586. doi: 10.1007/978-94-017-9014-7_26.
- [4] C. B. Field, M. J. Behrenfeld, J. T. Randerson, et P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components , *Science*, 1998, doi: 10.1126/science.281.5374.237.
- [5] F. Azam, T. Fenchel, J. Field, J. Gray, L. Meyer-Reil, et F. Thingstad, The Ecological Role of Water-Column Microbes in the Sea , *Marine Ecology Progress Series*, 1983, doi: 10.3354/meps010257.
- [6] R. W. Eppley et B. J. Peterson, Particulate organic matter flux and planktonic new production in the deep ocean , *Nature*, vol. 282, n° 5740, 1979, doi: 10.1038/282677a0.
- [7] L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp- D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, et G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean , *Nature*, 2016, doi: 10.1038/nature16942.
- [8] M. T. Kavanaugh, M. J. Oliver, F. P. Chavez, R. M. Letelier, F. E. Muller-Karger, et S. C. Doney, Seascapes as a new vernacular for pelagic ocean monitoring, management and conservation , *ICES Journal of Marine Science*, 2016, doi: 10.1093/icesjms/fsw086.
- [9] R. H. Fleming, Oceanography: The past , *Deep Sea Research Part A. Oceanographic Research Papers*, vol. 28, n° 11, 1981, doi: 10.1016/0198-0149(81)90044-3.
- [10] D. B. Rusch, A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcón, V. Souza, G. Bonilla-Rosso, L. E. Eguarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Birmingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, et J. C. Venter, The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific , *PLoS Biology*, vol. 5, n° 3, 2007, doi: 10.1371/journal.pbio.0050077.
- [11] C. M. Duarte, Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition , *Limnology and Oceanography Bulletin*, 2015, doi: 10.1002/lob.10008.
- [12] E. Karsenti, S. G. Acinas, P. Bork, C. Bowler, C. de Vargas, J. Raes, M. Sullivan, D. Arendt, F. Benzoni, J. M. Claverie, M. Follows, G. Gorsky, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, U. Krzic, F. Not, H. Ogata, S. Pesant, E. G. Reynaud, C. Sardet, M. E. Sieracki, S. Speich, D. Velayoudon, J. Weissenbach, P. Wincker, C. Abergel, D. Arslan, S. Audic, J. M. Aury, N. Babic, L. Beaufort, L. Bittner, E. Boss, C. Boutte, J. Brum, M. Carmichael, R. Casotti, A. Chambouvet, P. Chang, C. Chica, C. Clerissi, S. Colin, F. M. Cornejo-Castillo, C. Da Silva, S. De Monte, J. Decelle, Y. Desdevises, C. Dimier, J. Dolan, M. Duhaime, X. Durrieu de Madron, F. d'Ortenzio, F. d'Ovidio, I. Ferrera, L. Garczarek, M. J. Garet-Delmas, S. Gasmi, J. M. Gasol, N. Grimsley, R. Heilig, J. Ignacio-Espinoza, J. L.

- Jamet, L. Karp-Boss, M. Katinka, H. Khalili, Z. Kolber, N. Le Bescot, H. Le Goff, G. Lima-Mendez, F. Mahé, M. G. Mazzocchi, M. Montresor, P. Morin, B. Noel, C. Pedrós-Alió, E. Pelletier, Y. Perez, M. Picheral, G. Piganeau, O. Poirot, J. Poulain, N. Poulton, F. Prejger, J. Prihoda, I. Probert, J. Rampal, G. Reverdin, S. Romac, J. B. Romagnan, F. Roullier, C. Rouviere, G. Samson, S. Santini, H. Sarmento, A. Sciandra, S. Solonenko, L. Stemmann, L. Subirana, S. Sunagawa, A. Tanaka, P. Testor, A. Thompson, V. Tichanné-Seltzer, L. Tirichine, E. Toulza, S. Tozzi, A. Veluchamy, et A. Zingone, A holistic approach to marine Eco-systemsbiology , *PLoS Biology*, 2011, doi: 10.1371/journal.pbio.1001177.
- [13] S. Sunagawa, S. G. Acinas, P. Bork, C. Bowler, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels, L. Karp-Boss, E. Karsenti, M. Lescot, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, M. Sieracki, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, D. Eveillard, G. Gorsky, L. Guidi, D. Iudicone, E. Karsenti, F. Lombard, H. Ogata, S. Pesant, M. B. Sullivan, P. Wincker, et C. de Vargas, Tara Oceans: towards global ocean ecosystems biology , *Nature Reviews Microbiology*, 2020, doi: 10.1038/s41579-020-0364-5.
- [14] D. J. Richter, R. Watteaux, T. Vannier, J. Leconte, P. Frémont, G. Reygondeau, N. Maillet, N. Henry, G. Benoit, O. Da Silva, T. O. Delmont, A. Fernández-Guerra, S. Suweis, R. Narci, C. Berney, D. Eveillard, F. Gavory, L. Guidi, K. Labadie, E. Mahieu, J. Poulain, S. Romac, S. Roux, C. Dimier, S. Kandels, M. Picheral, S. Searson, T. O. Coordinators, S. Pesant, J.-M. Aury, J. R. Brum, C. Lemaitre, E. Pelletier, P. Bork, S. Sunagawa, F. Lombard, L. Karp-Boss, C. Bowler, M. B. Sullivan, E. Karsenti, M. Mariadassou, I. Probert, P. Peterlongo, P. Wincker, C. de Vargas, M. Ribera d'Alcalà, D. Iudicone, et O. Jaillon, Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems , *eLife*, vol. 11, p. e78129, août 2022, doi: 10.7554/eLife.78129.
- [15] P. Frémont, M. Gehlen, M. Vrac, J. Leconte, T. O. Delmont, P. Wincker, D. Iudicone, et O. Jaillon, Restructuring of plankton genomic biogeography in the surface ocean under climate change , *Nature Climate Change*, 2022, doi: 10.1038/s41558-022-01314-8.
- [16] T. O. Delmont, M. Gaia, D. D. Hinsinger, P. Frémont, C. Vanni, A. Fernandez-Guerra, A. M. Eren, A. Kourlaiev, L. d'Agata, Q. Clayssen, E. Villar, K. Labadie, C. Cruaud, J. Poulain, C. Da Silva, M. Wessner, B. Noel, J.-M. Aury, S. Sunagawa, S. G. Acinas, P. Bork, E. Karsenti, C. Bowler, C. Sardet, L. Stemmann, C. de Vargas, P. Wincker, M. Lescot, M. Babin, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, O. Jaillon, S. Kandels, D. Iudicone, H. Ogata, S. Pesant, M. B. Sullivan, F. Not, K.-B. Lee, E. Boss, G. Cochrane, M. Follows, N. Poulton, J. Raes, M. Sieracki, S. Speich, C. de Vargas, C. Bowler, E. Karsenti, E. Pelletier, P. Wincker, et O. Jaillon, Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean , *Cell Genomics*, vol. 2, n° 5, p. 100123, 2022, doi: <https://doi.org/10.1016/j.xgen.2022.100123>.
- [17] G. Benoit, P. Peterlongo, M. Mariadassou, E. Drezén, S. Schbath, D. Lavenier, et C. Lemaitre, Multiple comparative metagenomics using multiset k-mer counting , *PeerJ Computer Science*, vol. 2016, n° 11, 2016, doi: 10.7717/peerj-cs.94.

- [18] S. Clayton, S. Dutkiewicz, O. Jahn, C. Hill, P. Heimbach, et M. J. Follows, Biogeochemical versus ecological consequences of modeled ocean physics , *Biogeosciences*, vol. 14, n° 11, 2017, doi: 10.5194/bg-14-2877-2017.
- [19] R. R. Sokal, A statistical method for evaluating systematic relationships , *Univ Kans Sci Bull*, vol. 38, 1958.
- [20] L. Breiman et A. Cutler, Breiman and Cutler's random forests for classification and regression , *Package randomForest* , 2012, doi: 10.5244/C.22.54.
- [21] G. Ridgeway, gbm: Generalized Boosted Regression Models , *R package version 1.6-3.1*, 2010.
- [22] W. N. Venables et B. D. Ripley, *Modern Applied Statistics with S Fourth edition* by. 2002.
- [23] S. N. Wood, Stable and efficient multiple smoothing parameter estimation for generalized additive models , *Journal of the American Statistical Association*, 2004, doi: 10.1198/016214504000000980.
- [24] T. O. Delmont, J. J. P. Karlusich, I. Veseli, J. Fuessel, A. M. Eren, R. A. Foster, C. Bowler, P. Wincker, et E. Pelletier, Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean , *ISME J*, oct. 2021, doi: <https://doi.org/10.1038/s41396-021-01135-1>.
- [25] T. P. Boyer, J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson, R. a Locarnini, A. V. Mishonov, T. D. O'Brien, C. R. Paver, J. R. Reagan, D. Seidov, I. V. Smolyar, M. M. Zweng, T. D. O. Brien, C. R. Paver, J. R. Reagan, D. Seidov, I. V. Smolyar, M. M. Zweng, et K. D. Sullivan, WORLD OCEAN DATABASE 2013, NOAA Atlas NESDIS 72 , *Sydney Levitus, Ed.; Alexey Mishonoc, Technical Ed.*, 2013, doi: 10.7289/V5NZ85MT.
- [26] A. R. Longhurst, *Ecological Geography of the Sea*. 2007. doi: 10.1016/B978-0-12-455521-1.X5000-1.
- [27] G. Reygondeau, A. Longhurst, E. Martinez, G. Beaugrand, D. Antoine, et O. Maury, Dynamic biogeochemical provinces in the global ocean , *Global Biogeochemical Cycles*, 2013, doi: 10.1002/gbc.20089.
- [28] A. R. Fay et G. A. McKinley, Global open-ocean biomes: Mean and temporal variability , *Earth System Science Data*, 2014, doi: 10.5194/essd-6-273-2014.
- [29] G. Gorsky, G. Bourdin, F. Lombard, M. L. Pedrotti, S. Audrain, N. Bin, E. Boss, C. Bowler, N. Cassar, L. Caudan, G. Chabot, N. R. Cohen, D. Cron, C. De Vargas, J. R. Dolan, E. Douville, A. Elineau, J. M. Flores, J. F. Ghiglione, N. Haëntjens, M. Hertau, S. G. John, R. L. Kelly, I. Koren, Y. Lin, D. Marie, C. Moulin, Y. Moucherie, S. Pesant, M. Picheral, J. Poulain, M. Pujo-Pay, G. Reverdin, S. Romac, M. B. Sullivan, M. Trainic, M. Tressol, R. Troublé, A. Vardi, C. R. Voolstra, P. Wincker, S. Agostini, B. Banaigs, E. Boissin, D. Forcioli, P. Furla, P. E. Galand, E. Gilson, S. Reynaud, S. Sunagawa, O. P. Thomas, R. L. V. Thurber, D. Zoccola, S. Planes, D. Allemand, et E. Karsenti, Expanding Tara Oceans Protocols for Underway, Ecosystemic Sampling of the Ocean-Atmosphere Interface During Tara Pacific Expedition (2016–2018), *Frontiers in Marine Science*, 2019, doi: 10.3389/fmars.2019.00750.

- [30] J. J. Polovina, J. P. Dunne, P. A. Woodworth, et E. A. Howell, Projected expansion of the subtropical biome and contraction of the temperate and equatorial upwelling biomes in the North Pacific under global warming , *ICES Journal of Marine Science*, 2011, doi: 10.1093/icesjms/fsq198.
- [31] L. Wrightson et A. Tagliabue, *Quantifying the Impact of Climate Change on Marine Diazotrophy: Insights From Earth System Models*, vol. 7. 2020. Disponible sur: <https://www.frontiersin.org/article/10.3389/fmars.2020.00635>
- [32] C. Laufkötter, M. Vogt, N. Gruber, O. Aumont, L. Bopp, S. C. Doney, J. P. Dunne, J. Hauck, J. G. John, I. D. Lima, R. Seferian, et C. Völker, Projected decreases in future marine export production: The role of the carbon flux through the upper ocean ecosystem , *Biogeosciences*, vol. 13, n° 13, 2016, doi: 10.5194/bg-13-4023-2016.
- [33] D. Li, C.-M. Liu, R. Luo, K. Sadakane, et T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph , *Bioinformatics*, vol. 31, n° 10, p. 1674-1676, mai 2015, doi: 10.1093/bioinformatics/btv033.
- [34] B. Song, S. Chen, et W. Chen, Dinoflagellates, a unique lineage for retrogene research , *Frontiers in Microbiology*, vol. 9, n° JUL, 2018, doi: 10.3389/fmicb.2018.01556.

II. Article 1: Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems

Dans cet article⁸⁵, j'ai réalisé un important travail de vérification et reproduction des résultats, j'ai aussi apporté certains résultats notamment les analyses sur les distances environnementales et les analyses complémentaires sur l'imagerie et les MAGs. J'ai ainsi réalisé la refonte et la création de nombreuses figures et figures supplémentaires de l'article, accessibles en annexe. Cet article est fondateur par rapport aux deux articles suivants: il pose les bases des structures biogéographiques observées notamment par la définition des provinces génomiques dont les niches écologiques ont été calculées et extrapolées à l'ensemble des océans dans le second article³²¹. Étant co-auteur, je ne présente ici que l'abstract de l'article. Sa version intégrale est accessible en ligne à l'adresse suivante : <https://elifesciences.org/articles/78129>.

Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems

Daniel J Richter^{1,2†}, Romain Watteaux^{3,4†}, Thomas Vannier^{5,6,7†}, Jade Leconte^{5,6}, Paul Frémont^{5,6}, Gabriel Reygondeau^{8,9}, Nicolas Maillet¹⁰, Nicolas Henry^{1,6}, Gaëtan Benoit¹¹, Ophélie Da Silva^{6,12}, Tom O Delmont^{5,6}, Antonio Fernández-Guerra^{13,14,15}, Samir Suweis¹⁶, Romain Narci¹⁷, Cédric Berney^{1,6}, Damien Eveillard^{6,18}, Frederick Gavory⁵, Lionel Guidi^{6,12}, Karine Labadie¹⁹, Eric Mahieu¹⁹, Julie Poulain^{5,6}, Sarah Romac^{1,6}, Simon Roux²⁰, Céline Dimier^{1,21}, Stefanie Kandels^{22,23}, Marc Picheral^{6,12}, Sarah Searson^{6,12}, Tara Oceans Coordinators, Stéphane Pesant^{24,25}, Jean-Marc Aury⁵, Jennifer R Brum^{20,26}, Claire Lemaitre¹¹, Eric Pelletier^{5,6}, Peer Bork^{22,27,28}, Shinichi Sunagawa^{22,29}, Fabien Lombard^{6,12,30}, Lee Karp-Boss³¹, Chris Bowler^{6,21,21}, Matthew B Sullivan^{20,32,33,34}, Eric Karsenti^{6,21,23}, Mahendra Mariadassou¹⁷, Jan Probert^{1,6}, Pierre Peterlongo¹¹, Patrick Wincker^{5,6}, Colombar de Vargas^{1,6*}, Maurizio Ribera d'Alcalá^{3*}, Daniele Iudicone^{3*}, Olivier Jaillon^{5,6*}

*For correspondence: vargas@sb-roscoff.fr (CV); maurizio@szn.it (MRd'A); iudicone@szn.it (DI); ojailon@genoscope.crs.fr (OJ)

†These authors contributed equally to this work

Group author details: [Tara Oceans Coordinators See page 16](#)

Competing interest: [See page 17](#)

Funding: [See page 17](#)

Preprinted: 06 December 2019
Received: 23 February 2022
Accepted: 06 June 2022
Published: 03 August 2022

Reviewing Editor: Stilianos Louca, University of Oregon, United States

© Copyright Richter, Watteaux, Vannier *et al.* This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

¹Sorbonne Université, CNRS, Station Biologique de Roscoff, UMR7144, ECOMAP, Roscoff, France; ²Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta, Barcelona, Spain; ³Stazione Zoologica Anton Dohrn, Villa Comunale, Naples, Italy; ⁴CEA, DAM, DIF, F-91297, Arpajon Cedex, France; ⁵Génomique Métabolique, Genoscope, Institut de Biologie François Jacob, CEA, CNRS, Université Evry, Université Paris-Saclay, Evry, France; ⁶Research Federation for the study of Global Ocean systems ecology and evolution, FR2022/Tara GOSEE, Paris, France; ⁷Aix Marseille Univ., Université de Toulon, CNRS, IRD, MIO UM, Marseille, France; ⁸Changing Ocean Research Unit, Institute for the Oceans and Fisheries, University of British Columbia. Aquatic Ecosystems Research Lab, Vancouver, Canada; ⁹Ecology and Evolutionary Biology, Yale University, New Haven, CT, United States; ¹⁰Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, France; ¹¹Univ Rennes, CNRS, Irria, IRISA-UMR 6074, Rennes, France; ¹²Sorbonne Universités, CNRS, Laboratoire d'Océanographie de Villefranche, LOV, Villefranche-sur-Mer, France; ¹³Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark; ¹⁴MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany; ¹⁵Max Planck Institute for Marine Microbiology, Bremen, Germany; ¹⁶Dipartimento di Fisica e Astronomia 'G. Galilei' & CNISM, INFN, Università di Padova, Padova, Italy; ¹⁷MaIAGE, INRAE, Université Paris-Saclay, Jouy-en-Josas, France; ¹⁸Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, Nantes, France; ¹⁹Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France; ²⁰Department of Microbiology, The Ohio State University, Columbus, United States; ²¹Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France; ²²Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany; ²³Directors' Research European Molecular Biology

Laboratory, Heidelberg, Germany; ²⁴MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany; ²⁵PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany; ²⁶Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, United States; ²⁷Yonsei Frontier Lab, Yonsei University, Seoul, Republic of Korea; ²⁸Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany; ²⁹Institute of Microbiology, Department of Biology, ETH Zurich, Vladimir-Prelog-Weg, Zurich, Switzerland; ³⁰Institut Universitaire de France (IUF), Paris, France; ³¹School of Marine Sciences, University of Maine, Orono, United States; ³²EMERGE Biology Integration Institute, The Ohio State University, Columbus, United States; ³³Center of Microbiome Science, The Ohio State University, Columbus, United States; ³⁴Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, United States

Abstract Biogeographical studies have traditionally focused on readily visible organisms, but recent technological advances are enabling analyses of the large-scale distribution of microscopic organisms, whose biogeographical patterns have long been debated. Here we assessed the global structure of plankton geography and its relation to the biological, chemical, and physical context of the ocean (the 'seascape') by analyzing metagenomes of plankton communities sampled across oceans during the *Tara* Oceans expedition, in light of environmental data and ocean current transport. Using a consistent approach across organismal sizes that provides unprecedented resolution to measure changes in genomic composition between communities, we report a pan-ocean, size-dependent plankton biogeography overlying regional heterogeneity. We found robust evidence for a basin-scale impact of transport by ocean currents on plankton biogeography, and on a characteristic timescale of community dynamics going beyond simple seasonality or life history transitions of plankton.

III. Article 2: Restructuring of plankton genomic biogeography in the surface ocean under climate change

Cette article est publié dans la revue *Nature Climate Change*. La version publiée est présentée ici et reproduite avec la permission de Springer Nature. J'ai réalisé l'ensemble des analyses de l'article.



Restructuring of plankton genomic biogeography in the surface ocean under climate change

Paul Frémont^{1,2} , Marion Gehlen³ , Mathieu Vrac³ , Jade Leconte^{1,2} , Tom O. Delmont^{1,2},
Patrick Wincker^{1,2}, Daniele Iudicone⁴  and Olivier Jaillon^{1,2} 

The impact of climate change on diversity, functioning and biogeography of marine plankton remains a major unresolved issue. Here environmental niches are evidenced for plankton communities at the genomic scale for six size fractions from viruses to meso-zooplankton. The spatial extrapolation of these niches portrays ocean partitionings south of 60° N into *climato-genomic* provinces characterized by signature genomes. By 2090, under the RCP8.5 future climate scenario, provinces are reorganized over half of the ocean area considered, and almost all provinces are displaced poleward. Particularly, tropical provinces expand at the expense of temperate ones. Sea surface temperature is identified as the main driver of changes (50%), followed by phosphate (11%) and salinity (10%). Compositional shifts among key planktonic groups suggest impacts on the nitrogen and carbon cycles. Provinces are linked to estimates of carbon export fluxes which are projected to decrease, on average, by 4% in response to biogeographical restructuring.

Planktonic communities are composed of complex and heterogeneous assemblages of small animals, single-celled eukaryotes (protists), bacteria, archaea and viruses. They contribute to the regulation of the Earth system through primary production via photosynthesis¹ and carbon export to the deep ocean^{2,3} and are at the base of the food webs that sustain the whole trophic chain in the oceans⁴.

The composition of planktonic communities varies over time at a given site with daily⁵ to seasonal fluctuations⁶ following environmental variability⁷ (for example, temperature, nutrients). Overlying these relatively short-scale spatiotemporal variations, a more macroscale partitioning of the ocean has been revealed by different combinations of biological and physicochemical data^{8–10} and recently documented at the resolution of community genomics¹¹. The basin-scale biogeographical structure has been proposed to result from a combination of multiple biophysicochemical processes named the seascape⁷. These processes include both abiotic and biotic interactions¹², neutral genetic drift¹³, natural selection^{14,15}, temperature variations, nutrient supply but also advection and mixing along currents^{11,13}.

Today, knowledge of global-scale plankton biogeography at the DNA level is in its infancy. We lack understanding and theoretical explanations for the emergence and maintenance of biogeographical patterns at genomic resolution. Omics data (that is, the DNA/RNA sequences representative of the variety of coding and non-coding sequences of organisms) provide the appropriate resolution to track and record global biogeographical features¹¹, modulation of the repertoire of expressed genes in a community in response to environmental conditions^{2,16,17} and eco-evolutionary processes^{13–15}. Moreover, as recently demonstrated by global expeditions^{2,18–21}, metagenomic sequencing can be consistently analysed across plankton organisms. The strong links between plankton and environmental conditions suggest potentially major consequences of climate change on community composition and biogeography^{22,23}.

Time series observations have highlighted recent changes in the planktonic ecosystem such as changes in community composition²⁴ or poleward shifts of some species²⁵, mainly attributed to temperature increase caused by anthropogenic greenhouse gas emissions. These changes are expected to intensify with ongoing global warming²⁶ and could lead to major reorganization of plankton community composition²² with a potential decline in diversity^{27–29}. Another major consequence of global reorganization of the seascape on biological systems would be a decrease of primary production at mid-latitudes and an increase at higher latitudes²⁶.

Here we report the global structure of plankton biogeography south of 60° N based on metagenomic data using niche models and its putative modifications under climate change. First, we define environmental niches³⁰, that is, the envelope of environmental parameters suitable for an organism or a population at the scale of genomic provinces across size organism size fractions representing major plankton groups from nano- (viruses) to meso-zooplankton (small metazoans). Then, we spatially extrapolate their niches into *climato-genomic* provinces to derive the structure of plankton biogeography for each size fraction individually and for all combined. Next, considering the same niches, we assess the spatial reorganization of these provinces under climate change by the end of the century. We study compositional changes among planktonic groups known to be critical for two major biogeochemical cycles: nitrogen with nitrogen-fixing bacteria and carbon with phototrophs and copepods. By correlating our provinces with carbon export fluxes, these compositional changes would lead to a reduction of approximately 4% of carbon export. Finally, we quantify the relative importance of the environmental drivers explaining projected changes.

Niche models and signature genomes from genomic provinces

We define and validate environmental niches using four machine learning techniques for 27 previously defined genomic provinces¹¹.

¹Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, Evry, France. ²Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans, Paris, France. ³LSCE-IPSL, CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France. ⁴Stazione Zoologica Anton Dorn, Naples, Italy. ✉e-mail: pfremont@genoscope.cns.fr; marion.gehlen@lsce.ipsl.fr; ojaillon@genoscope.cns.fr

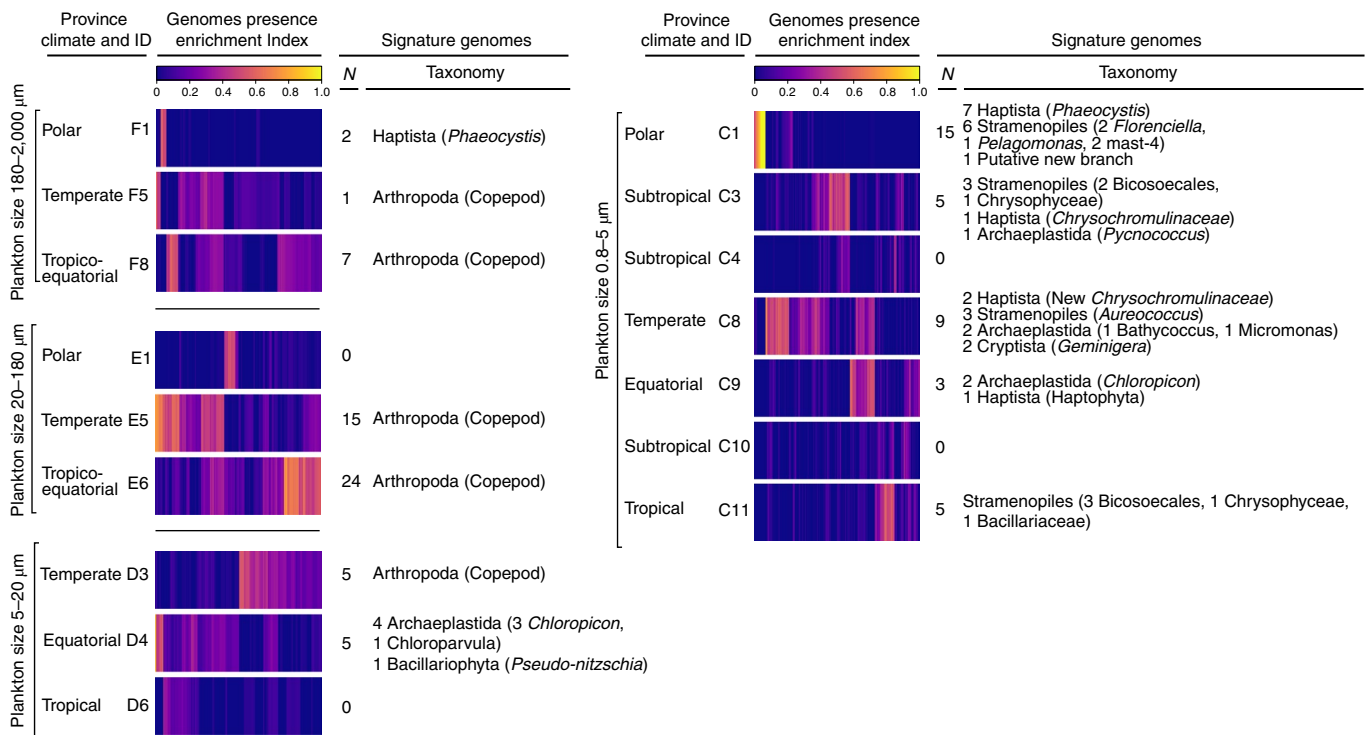


Fig. 1 | Eukaryotic signature genomes of provinces of eukaryote enriched size classes. For each plankton size class, indexes of presence enrichment (equation (1)) for 713 genomes of eukaryotic plankton³¹ in corresponding provinces are clustered and represented in a colour scale. Signature genomes (Methods) are found for almost all provinces; their number (N) and taxonomies are summarized (detailed list in Supplementary Table 6). For each size class, a genome is considered to be signature of a province if its presence enrichment index is superior to 0.5 for that province and inferior to 0.1 for other provinces.

These provinces cover major ocean regions except the Arctic for which no data were available at the time of this study. They correspond to 529 metagenomes for six size fractions (ranging from 0 μm to 2,000 μm) sampled at 95 sites (Supplementary information 1, Extended Data Fig. 1 and Supplementary Fig. 1). Niche definition relies on predictor variables that significantly differentiate genomic provinces¹¹: sea surface temperature (SST), salinity, dissolved silica, nitrate, phosphate and iron, plus a seasonality index of nitrate (equation (2)).

The signal of ocean partitioning is probably due to abundant and small genomes whose geographical distributions closely match provinces. Within a collection of 1,778 bacterial, 110 archaeal and 713 eukaryotic environmental genomes^{31,32} characterized from *Tara* Oceans samples without cultivation, we find a total of 324 signature genomes covering all but four provinces and displaying a taxonomic signal coherent with the size fractions (Fig. 1 for eukaryotes and Extended Data Fig. 2 for Bacteria and Archaea). Some of the signature genomes correspond to unexplored lineages with no cultured representatives, highlighting the knowledge gap for organisms that structure plankton biogeography and the strength of a rationale devoid of any a priori on reference genomes or species.

Structure of present-day biogeography of plankton

Niches are extrapolated to a global ocean biogeography for each size fraction at $1^\circ \times 1^\circ$ spatial resolution using environmental predictors from 2006–2013 World Ocean Atlas 2013 (WOA13) climatology³³. At each grid point, provinces with the highest probability of presence computed from the mean projection of four machine learning models are retained (Fig. 2 and Supplementary Fig. 2). They are referred to as *dominant* provinces and manually annotated based on their climate (Supplementary Table 1).

In agreement with previous observations¹¹, provinces of large size fractions ($>20\mu\text{m}$) are wider and partially decoupled from those of smaller size fractions, probably due to differential responses to oceanic circulation and environmental variations, different life cycle constraints, lifestyles^{7,11} and trophic network positions³⁴. Biogeographies of small metazoans that enrich the largest size fractions (180–2,000 μm and 20–180 μm) are broadly aligned with latitudinal bands (tropico-equatorial, temperate and subpolar) dominated by a single province (Fig. 2a,b). A more complex oceanic structuring emerges for the smaller size fractions ($<20\mu\text{m}$) (Fig. 2c–f) with several provinces per large geographical region. For size fraction 0.8–5 μm enriched in small protists (Fig. 2d), distinct provinces are identified for tropical oligotrophic gyres and for the nutrient-rich equatorial upwelling region. A complex pattern of provinces, mostly latitudinal, is found for the bacteria- (Fig. 2e, 0.22–3 μm) and the virus enriched size classes (Fig. 2f, 0–0.2 μm), although less clearly linked to large-scale oceanographic regions. A single province extending from temperate to polar regions emerges for size fraction 5–20 μm enriched in protists (Fig. 2c) for which fewer samples were available (Supplementary Fig. 1b,c), a likely source of bias. A consensus map combining all size fractions, built using the PHATE algorithm³⁵, summarizes the main characteristics of the biogeographies described above (Supplementary Discussion 2 and Supplementary Fig. 3).

Finally, we compare genomic biogeographies and existing ocean partitionings^{8–10}. Though each of them is unique (Supplementary Figs. 6–8), common borders highlight a global latitudinal partitioning independent of the type of data (Supplementary Discussion 3).

Future changes in plankton biogeography structure

We assess the impacts of climate change on plankton biogeography at the end of the century following the Representative Concentration

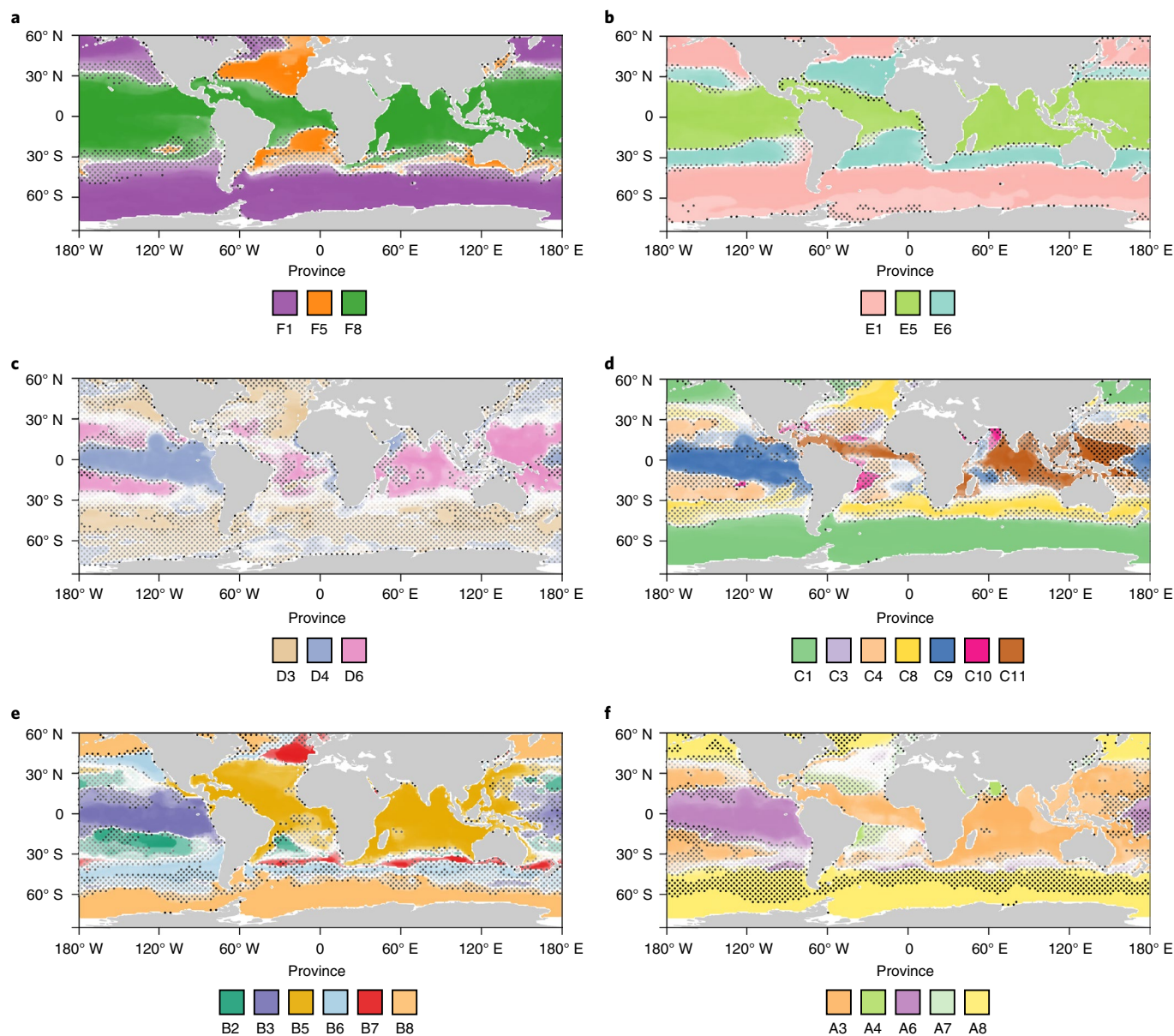


Fig. 2 | Global biogeographies of size-structured plankton provinces projected on WOA13 dataset. a–f, Maps showing biogeographies of metazoans enriched (180–2,000 μm) (**a**), small metazoans enriched (20–180 μm) (**b**), protist enriched (5–20 μm) (**c**), protist enriched (0.8–5 μm) (**d**), bacteria enriched (0.22–3 μm) (**e**) and viruses enriched (0–0.2 μm) (**f**) size fraction. At each grid point of the maps, the *dominant* province is represented using a darkness of colour proportional to its presence probability. Dots represent areas of uncertainty (where the delta of probability between the *dominant* and another province is inferior to 0.5). Simple biogeographies are observed in large size fractions (>20 μm) with a partitioning in three major oceanic areas: tropico-equatorial, temperate and polar. More complex geographic patterns and patchiness are observed in smaller size fractions. Maps generated using R package ‘maps’⁵¹.

Pathway 8.5 (RCP8.5)³⁶ greenhouse gas concentration trajectory. To consistently compare projections of present and future biogeographies, we use the bias-adjusted mean of six Earth System Model (ESM) climatologies (Supplementary Table 2 and Supplementary Fig. 9). The highest SST warming (7.2 °C) is located off the east coast of Canada in the North Atlantic while complex patterns of salinity and nutrient variations are projected in all oceans (Supplementary Fig. 10). Following this trajectory, future temperature at most sampling sites will be higher than the mean and maximum contemporary temperature within their current province (Extended Data Fig. 3).

Our projections indicate multiple large-scale changes in biogeographical structure, including expansions, shrinkages and shifts in plankton organism size-dependent provinces (Fig. 3a–d, Extended

Data Fig. 4 and Supplementary Figs. 11 and 12). A change in the *dominant* province in at least one size fraction would occur over 60.1% of the ocean surface studied, ranging from 12% (20–180 μm) to 31% (0.8–5 μm) (Fig. 3 and Table 1). Consistent with previous studies^{22,37}, the majority of shifts are poleward (Supplementary Discussions 2, 4 and 5).

We calculate a dissimilarity index (equation (4)) at each grid point between probabilities of future and present *dominant* provinces for all size fractions combined (Fig. 4a). Large dissimilarities are obtained over northern (25° to 60°) and southern (–25° to –60°) temperate regions (mean of 0.29 and 0.24, respectively), mostly reflecting the poleward shift of temperate provinces (red arrows in Fig. 4a). In austral and equatorial regions, despite important environmental changes (Supplementary Fig. 10) and previously

projected changes in diversity^{27–29} and biomass³⁸, the contemporary provinces remain the most probable at the end of the century (mean dissimilarities of 0.18 and 0.02, respectively).

To further study the future decoupling between provinces of different size fractions, we analyse the assemblages of *dominant* provinces of each size fraction. Whether a threshold is applied to the dissimilarity index ($>1/6$; Methods), from 45.3% to 57.1% of the ocean surface, mainly located in temperate regions, would be inhabited in 2090–2099 by assemblages that exist elsewhere in 2006–2015 (Fig. 4b versus Fig. 4c). Contemporary assemblages would disappear on 3.5% to 3.8% of the surface, and, conversely, novel assemblages not encountered today would cover 2.9% to 3.0% of the surface. Changes of assemblages impact key economic zones with over 41.8% to 51.8% of the surface of the main fisheries and 42.2% to 55.2% of exclusive economic zones for which the future assemblage would differ from the contemporary one (Extended Data Fig. 5).

Impact on plankton groups driving carbon and nitrogen cycles

To assess the potential biogeochemical impact of biogeographical restructuring, we focus on compositional changes among nitrogen-fixing bacteria (diazotrophs), copepods and phototrophs, three groups considered important for the carbon and nitrogen cycles^{39,40} that are well represented among the *Tara* Oceans environmental genomes. Focusing on marine areas where *dominant* provinces are projected to be replaced, we compare the present and future distribution of environmental genomes corresponding to 27 diazotrophs, 198 copepods and 231 phototrophs^{31,32}.

We project statistically significant changes in composition for all three groups (Fig. 3e,f and Extended Data Figs. 6–9, Wilcoxon test, $p < 0.05$). Here we describe only changes in diazotrophs for which the genome collection most likely encompasses the whole diversity of the group (Supplementary Discussion 6 addresses changes in copepods and phototrophs).

Diazotrophy, the biotic fixation of atmospheric nitrogen, is an important process for both nitrogen and carbon cycles. It supports biological productivity in the nitrogen-limited tropical oceans⁴¹. Marine diazotrophs include cyanobacteria (for example, *Trichodesmium*) described as the principal nitrogen fixers⁴² and various heterotrophic bacterial diazotrophs (HBDs) that lack cultured representatives or in situ imaging^{32,41}. Forty-eight of the bacterial environmental genomes are diazotrophs (eight cyanobacteria and 40 HBDs). These genomes encapsulate 92% of the metagenomic signal of the known marker gene responsible for microbial nitrogen fixation³² (*nifH*, the nitrogenase subunit H). Twenty-seven are found in at least five samples of a given size class and have been used for the analysis.

In size fraction 0.8–5 μm , we project significantly higher relative abundances in cyanobacteria in the tropico-equatorial regions of the Pacific Ocean (C9 to C11 and C4; Fig. 3f), which might point towards an increase in nitrogen fixation in this region as previously suggested by other models⁴⁰. Supporting this result, we find similar significant compositional changes towards an increase in cyanobacteria in size fraction 5–20 μm (Extended Data Fig. 7c). We also project significant compositional changes for some clades of HBDs

(for example, increase in gammaproteobacteria: C8 to C3; Fig. 3f) though no global trend can be identified here. In the other size classes (20–180 μm , 180–2,000 μm and 0.22–3 μm), compositional changes are not significant (Extended Data Fig. 8). To summarize, although we cannot estimate nitrogen fixation rates using genomic data alone, genomic measurements of nitrogen-fixing cyanobacteria are in agreement with a potential increase in nitrogen fixation in the tropics, echoing results from other models⁴⁰.

Linking assemblages of provinces and carbon export fluxes

Carbon export refers to the processes by which organic carbon is transferred from the surface ocean to depth mainly by sinking particles that are derived from surface ocean plankton ecosystem processes⁴³. Surface plankton communities are reported to be important in determining carbon export fluxes through the water column². We test this hypothesis by calculating mean fluxes of carbon export that can be associated to the assemblages of our *climate-genomic* provinces from three datasets of extrapolated carbon fluxes^{3,43,44}. We find significant differences of carbon fluxes for 46% of pairs of province assemblages (pairwise Wilcoxon test, Holm correction $p < 0.05$). Total carbon export values are comparable to those computed from the reference datasets over the same regions (from 2.6 using Henson et al.² to 6.7 Pg C per year using Eppley et al.⁴³; Supplementary Table 9). Our projections of the reorganization of *climate-genomic* provinces correspond with a decrease in carbon export of 4.0% on average by the end of the century (from 3.3% for Laws et al.⁴⁴ to 4.4% for Eppley et al.⁴³; Fig. 4d).

We use the Apriori algorithm⁴⁵ to identify associations between changes in the genome composition of communities and changes in carbon export among four main latitudinal zones of the ocean: equatorial, subtropical north/south and temperate/subpolar (Extended Data Fig. 10). In temperate/subpolar regions, no significant association rules are found (Extended Data Fig. 10a). In northern and southern subtropical regions, decreases in carbon export are associated with global decreases in nano-diatoms, nano-algae and pico-algae (0.8–5 μm and 5–20 μm ; Extended Data Fig. 10b,c). In equatorial regions, the decrease in carbon export is mainly associated with decreases in nano-diatoms and nano-algae (5–20 μm and 0.8–5 μm) and increases in diazotrophs (0.8–5 μm and 5–20 μm ; Extended Data Fig. 10d). No significant rules are found for carbon export increase or changes in copepods and micro-algae.

Drivers of plankton biogeography reorganization

We quantify the relative importance of environmental predictors (SST, salinity, dissolved silica, phosphate, nitrate, iron and seasonality of nitrate) into niche definitions and in driving future changes of the structure of plankton biogeography (equation (6)). Among these environmental properties, temperature is the first influential parameter for niche definition (for 19 out of 27 niches) but with a relative contribution of only 22.6% on average (Supplementary Fig. 16a).

For each site highly dissimilar between 2006–2015 and 2090–2099 (Bray-Curtis $> 1/6$), the relative impact of each environmental parameter in driving this change is calculated (equation (6) (ref. 22); Fig. 5a). Overall, comparing the relative contribution of each parameter to the reorganization of the provinces, SST would be

Fig. 3 | Modelled present and future global biogeographies of the three and seven provinces for the small metazoan enriched size fraction (20–180 μm) and the protist enriched size fraction (0.8–5 μm) and associated compositional shifts in marine hexanauplia and diazotrophs. a–d, Presence probability and uncertainties are represented as in Fig. 2. **e–f**, At each grid point of the maps, the dominant province is represented using a darkness of colour proportional to its presence probability. Dots represent areas of uncertainty (where the delta of probability between the dominant and another province is inferior to 0.5). Top: locations of *dominant* province shifts in size fraction 20–180 μm (**e**) and 0.8–5 μm (**f**). Bottom: Summary of significant compositional shifts in marine Hexanauplia (copepods) (**e**) classified by size ('not classified' when no preferential size class was found) or bacterial diazotrophs (**f**). Each type of transition is coloured according to the one on the map or in grey (when less frequent). Bar plots represent mean relative abundances based on genome abundance in the given province. Statistically significant compositional changes in a type of genome are represented by triangles of the associated transition colour (Wilcoxon test, $p < 0.05$). Maps generated using R package 'maps'⁵¹.

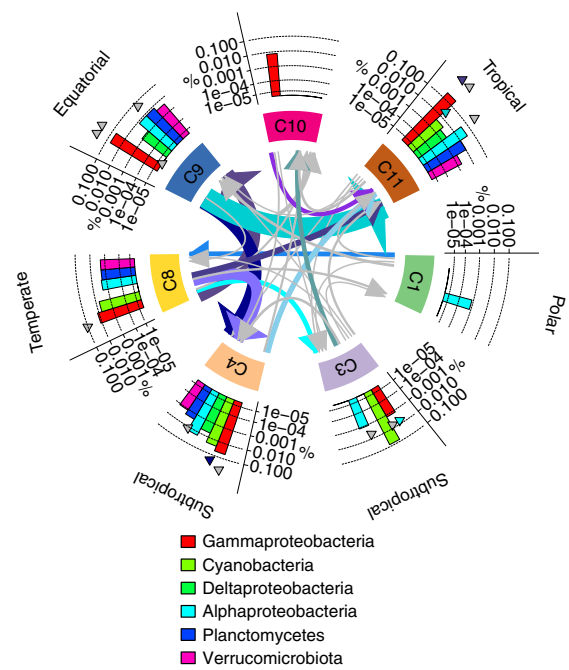
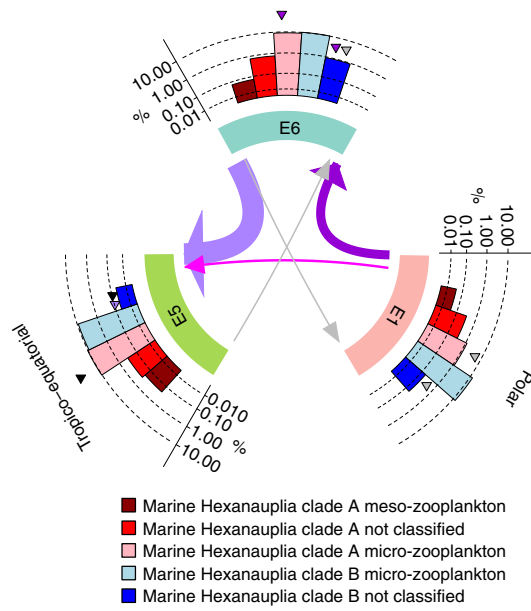
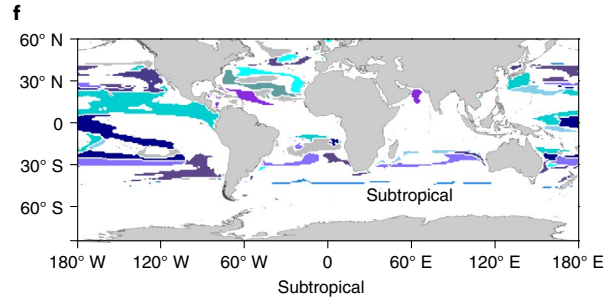
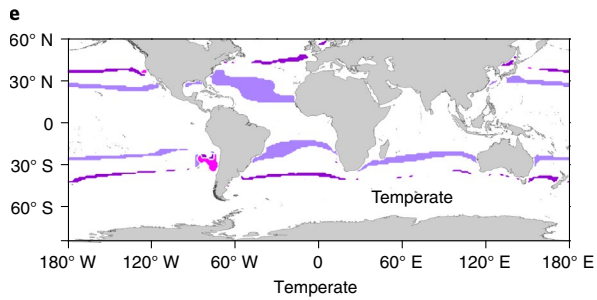
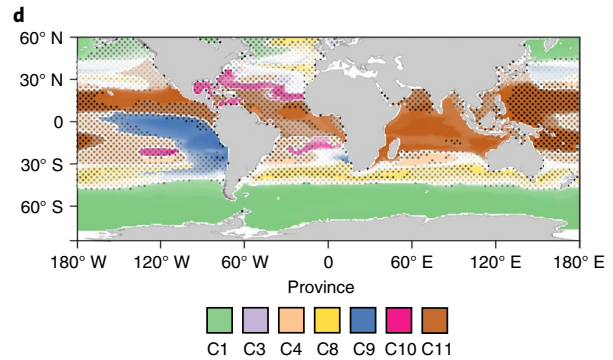
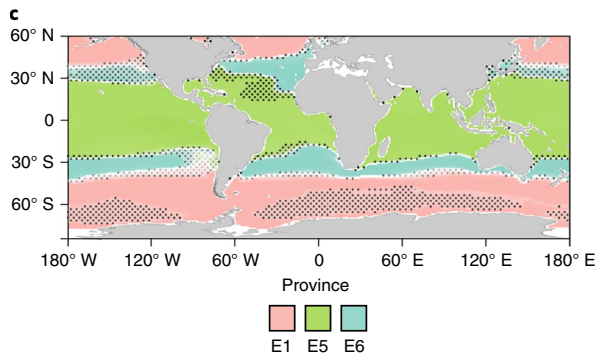
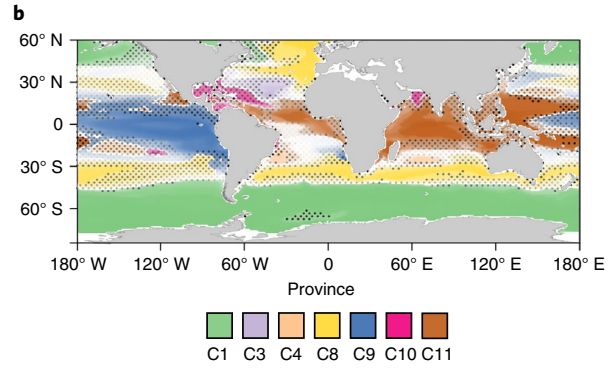
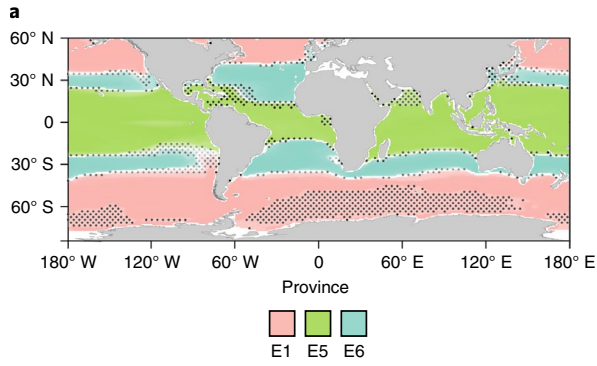


Table 1 | Global statistics of covered areas and province changes and transitions

Size fraction (μm)	Present-day covered area (%)	End-of-century covered area (%)	% area with a change of <i>dominant</i> province	Most frequent transition	%	Second-most frequent transition	%
180–2,000	74	74	13	temperate → tropico-equatorial (F5 → F8)	67	polar → temperate (F1 → F5)	14
20–180	78	77	12	temperate → tropico-equatorial (E6 → E5)	67	polar → temperate (E1 → E6)	29
5–20	45	49	22	temperate → equatorial (D3 → D4)	47	equatorial → tropical (D4 → D6)	25
0.8–5	56	59	31	equatorial → tropical (C9 → C11)	22	temperate → equatorial (C8 → C9)	15
0.22–3	60	61	15	temperate → tropical (B7 → B5)	22	polar → temperate (B8 → B6)	16
0–0.2	64	66	16	equatorial → tropical (A6 → A3)	32	temperate → equatorial (A8 → A6)	31
Total			60				

From 12% to 31% of the total covered area is estimated to be replaced by a different province at the end of the century compared with present day depending on the size fraction. In total, considering all size fractions, this represents 60% of the total covered area with at least one predicted change of *dominant* province across the six size fractions.

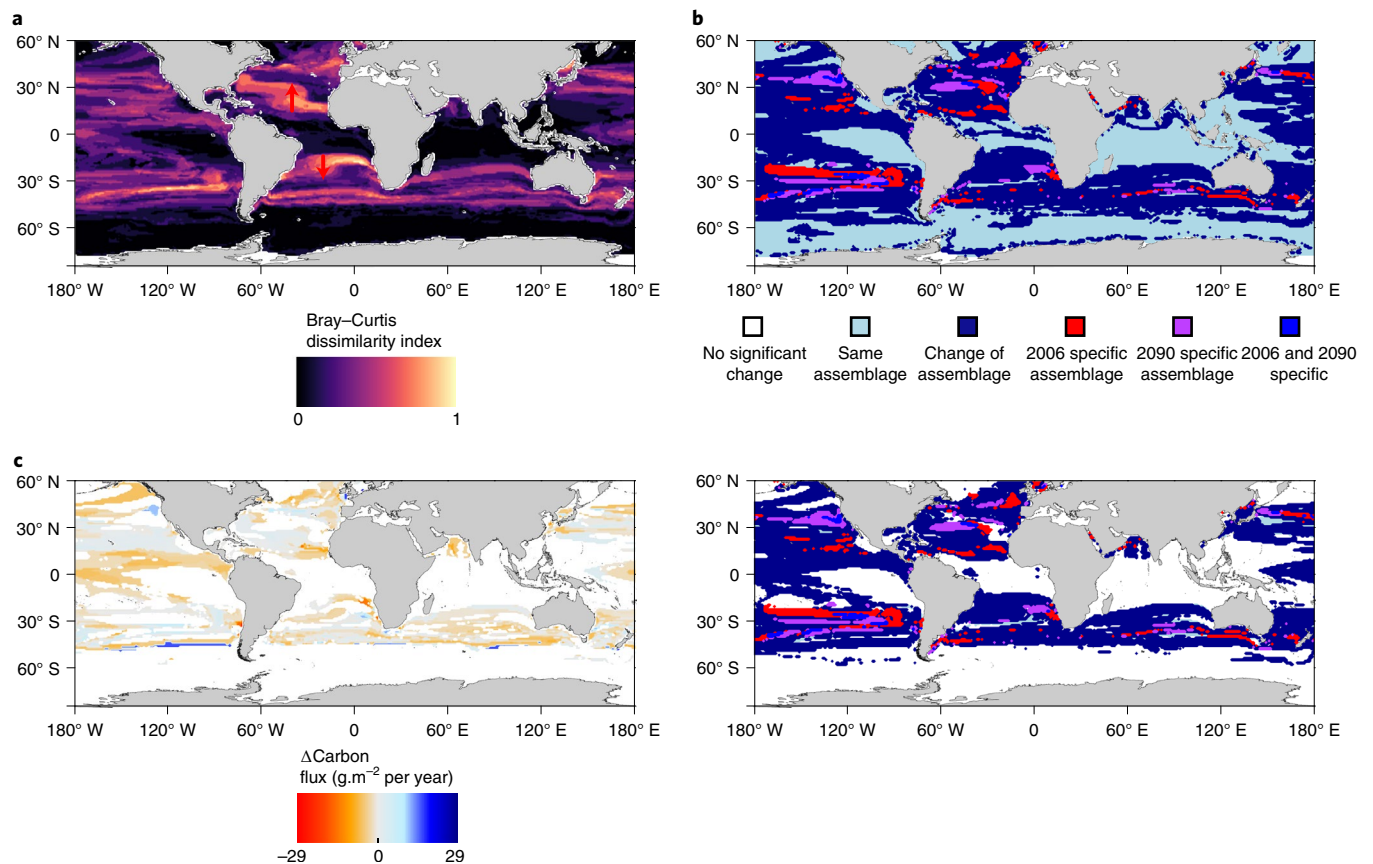


Fig. 4 | Global impact of climate change on plankton community assemblages and carbon export. **a**, Bray–Curtis dissimilarity index (equation (4)) is calculated by integrating all the *dominant* provinces presence probabilities over the six size fractions. **b**, An assemblage is the combined projected presence of the dominant province of each size class. Assemblage reorganization is either mapped on all considered oceans (top) or with a criterion on the Bray–Curtis index (Bray–Curtis > 1/6) (bottom). New assemblages are expected to appear in 2090 (purple + blue), whereas some 2006 specific assemblages are projected to disappear (red + blue). **c**, Difference in carbon export by the end of the century based on present day mean export within each assemblage (Henson et al.³). Maps generated using R package ‘maps’⁴⁵.

responsible for 50% on average followed by phosphate (11%) and salinity (10.3%) (Supplementary Fig. 17). Over the majority of the ocean, SST is the primary driver of the reorganization (Fig. 5a).

In some regions, salinity (for example, eastern North Atlantic) and phosphate (for example, equatorial region) dominate (Fig. 5a). When excluding the effect of SST, salinity and phosphate become

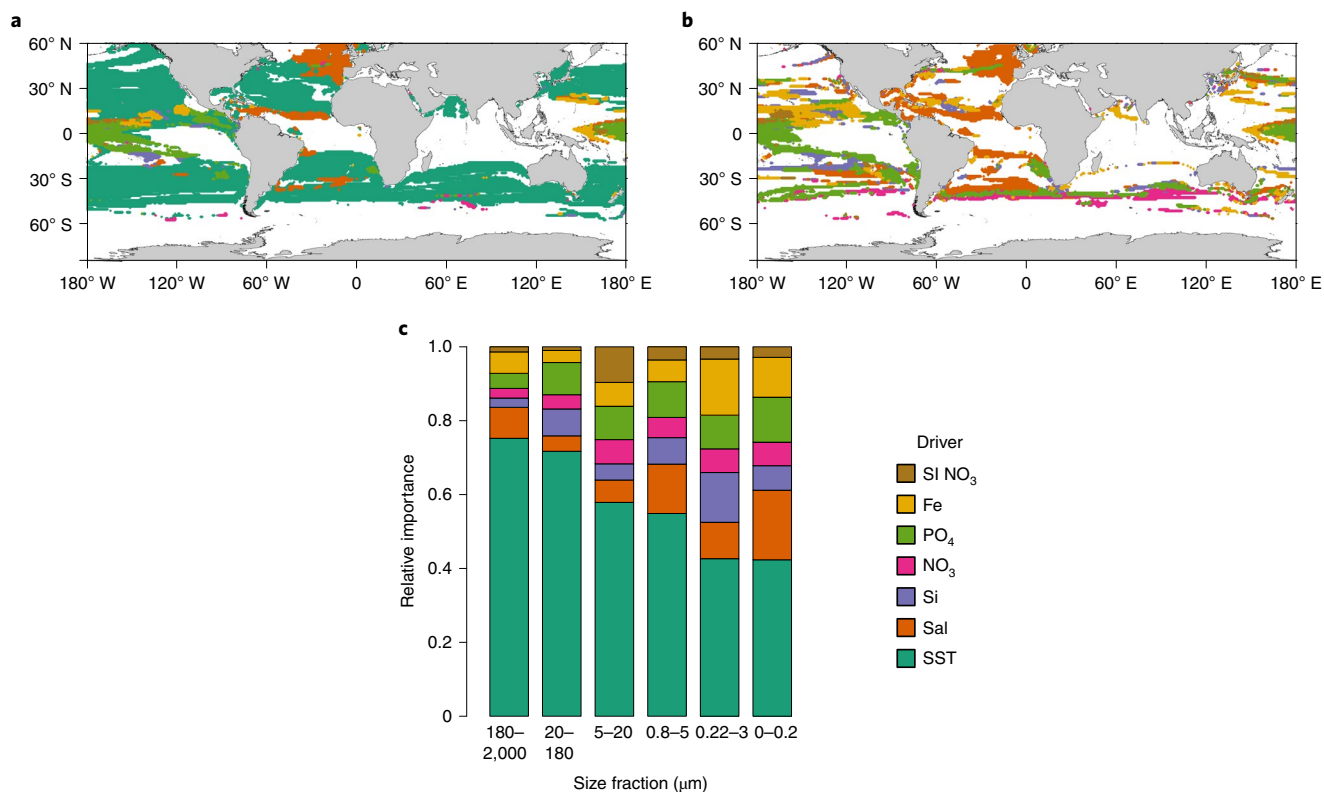


Fig. 5 | Drivers of plankton biogeographical restructuring in response to climate change. **a**, Map of most impacting drivers on *dominant* province changes. Temperature appears as the top impacting driver on the majority of the ocean with a significant change of province. **b**, Most impacting driver without considering temperature change. Salinity (Sal) and phosphate are found to be the second and third drivers of province reorganization; notably, they dominate at tropical and subpolar latitudes. **c**, Relative importance of the drivers in the different size fractions. Sea Surface Temperature (SST) is found to be the most important driver for all size classes but has a more important impact in large size classes ($>20\ \mu\text{m}$). SI NO₃ is a seasonality index of nitrate (Methods). Maps generated using R package ‘maps’⁵¹.

the primary drivers of the reorganization of the provinces (Fig. 5b). The impact of SST varies across size classes with a significantly higher contribution in large size classes ($>20\ \mu\text{m}$) compared with the small ones (mean of $\sim 73\%$ versus $\sim 49\%$, t -test $p < 0.05$; Fig. 5c). Though the contribution of combined nutrients to niche definition is similar for small and large size classes (mean of $\sim 56\%$ versus $\sim 61\%$; Supplementary Fig. 16 and Supplementary Table 3), their future projected variations have a higher relative impact on the reorganization of biogeographies of small organisms (mean of $\sim 39\%$ versus $\sim 20\%$, t -test $p < 0.05$; Supplementary Fig. 16 and Supplementary Table 3). For instance, in the tropical zone, the shrinkage of the equatorial province C9 (size fraction $0.8\text{--}5\ \mu\text{m}$; Fig. 2b,d and Supplementary Fig. 18e) is 24% driven by a reduction of dissolved phosphate concentrations and 25% by a SST increase. In contrast, SST drives 56% of the shrinkage of the temperate province F5 (size fraction $180\text{--}2,000\ \mu\text{m}$; Supplementary Fig. 18d). Finally, non-poleward shifts are found only within small size fractions ($<20\ \mu\text{m}$) (Extended Data Fig. 4 and Supplementary Fig. 11), highlighting differential responses to nutrients and SST changes between large and small organisms, the latter being enriched in phytoplankton that directly rely on nutrient supplies.

Discussion

We propose a novel partitioning of the ocean into plankton size-dependent *climato-genomic* provinces, complementing previous efforts based on other biophysicochemical data^{9–11}. Though initially built at genomic scale, our biogeographies paradoxically reveal basin-scale provinces that are larger than Longhurst’s biogeochemical provinces¹⁰ and Fay and McKingley biomes⁹. We propose

that this apparent paradox emerges from the combination of the scale, nature and resolution of sampling. First, two proximal samples from the *Tara* Oceans expedition are separated by $\sim 300\ \text{km}$ on average, sampled over three years. This relatively large spatiotemporal scale overlies shorter-scale compositional variations previously observed^{5,6}. At the global scale, the limited effect of seasons on the frontiers of biogeochemical provinces¹⁰ does not affect their global topology which might be the case for the global structuring of the genomic biogeography. Second, our estimates of plankton community dissimilarities are highly resolutive as they are computed at genomic scale with billions of small DNA fragments^{11,18}, thus smoothing out the more discrete species-level signals. Together, from these combinations of processes and patterns occurring at multiple scales emerge basin-scale provinces associated with coherent environmental niches and signature genomes.

These *climato-genomic* provinces are structured in broad latitudinal bands with smaller organisms ($<20\ \mu\text{m}$) displaying more complex patterns and partially decoupled from larger organisms. This decoupling is the result of distinct statistical links between provinces based on organism size fractions and environmental parameters and could reflect their respective trophic modes^{11,34}.

Complex changes of the parameters defining the niches are projected under climate change, leading to the reorganization of size-dependent provinces. Assuming a constant relationship to environmental drivers that define the *climato-genomic* provinces, climate change is projected to restructure them over approximately 50% of surface oceans south of $60^\circ\ \text{N}$ by the end of the century (Fig. 4). The largest reorganization is detected in subtropical and temperate regions in agreement with other studies^{29,37} and is accompanied

by appearance and disappearance of size-fractionated provinces' assemblages. In the tropical-equatorial and southern regions, environmental conditions are projected to exceed current extremes by the end of the century. While some studies extrapolate important diversity and biomass changes in these zones^{27–29,38}, here we project shifts of their boundaries and maintenance of their climatic label. The present approach does not account for putative changes in community composition or the emergence of novel niches over these regions for which novel environmental selection pressure is expected.

Despite these limitations, genomic data allow us to project compositional shifts due to the reorganization of the provinces at a high phylogenetic and functional resolution. We find significant changes in composition of diazotrophs, copepods and phototrophs including changes in size, a key trait for carbon export. These changes include increase of large copepods in subpolar regions (180–2,000 µm), increase in nitrogen-fixing cyanobacteria in subtropical regions in agreement with modelling studies⁴⁰ and complex changes in phototrophs. These changes might lead to novel prey–predator interactions, for example, in regions where new assemblages of communities are projected. Critically, carbon export fluxes are significantly different among *climato-genomic* province assemblages in agreement with theory^{3,26} and previous genomic studies². By the end of the century, changes in these assemblages are linked to a global decrease of the carbon pump of ~4%, a consistent estimation compared with most models^{26,46}. This decrease is linked to decreases in nano-algae and diatoms and an increase in diazotrophs but not to changes in copepods or micro-algae. These results are to be taken with caution, especially for copepods and phototrophs, as our genome database does not account for the total diversity of these groups and is biased towards small and the most abundant genomes. Nevertheless, the presented association of metagenomic data with physicochemical data paves the way for the integration of multiple global-scale data for a better understanding of biogeochemical cycles.

Overall, our projections for the end of the century do not take into account possible future changes of major biophysicochemical factors such as the dynamics of community mixing, trophic interactions through transport⁴⁷, the dynamics of the genomes^{13–15} (adaptation or acclimation) and biomass variations³⁸. New sampling in current and future expeditions⁴⁸ and ongoing technological improvements in biophysicochemical characterization of seawater samples^{31,48,49} will soon refine functional^{16,17}, environmental (micronutrients⁵⁰) and phylogenetic³¹ characterization of plankton ecosystems for various biological entities (genotypes, species or communities) and spatiotemporal scales⁴⁸. Ultimately, integrating this varied information will allow a better understanding of the conditions of emergence of ecological niches in the seascape and their response to a changing ocean.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41558-022-01314-8>.

Received: 28 October 2020; Accepted: 15 February 2022;
Published online: 04 April 2022

References

- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* <https://doi.org/10.1126/science.281.5374.237> (1998).
- Guidi, L. et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* <https://doi.org/10.1038/nature16942> (2016).

- Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Glob. Biogeochem. Cycles* <https://doi.org/10.1029/2011GB004099> (2012).
- Azam, F. et al. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* <https://doi.org/10.3354/meps010257> (1983).
- Saab, M. A. Day-to-day variation in phytoplankton assemblages during spring blooming in a fixed station along the Lebanese coastline. *J. Plankton Res.* <https://doi.org/10.1093/plankt/14.8.1099> (1992).
- Djurhuus, A. et al. Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-14105-1> (2020).
- Kavanaugh, M. T. et al. Seascapes as a new vernacular for pelagic ocean monitoring, management and conservation. *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsw086> (2016).
- Longhurst, A. R. *Ecological Geography of the Sea* (Elsevier, 2007).
- Fay, A. R. & McKinley, G. A. Global open-ocean biomes: mean and temporal variability. *Earth Syst. Sci. Data* <https://doi.org/10.5194/essd-6-273-2014> (2014).
- Reygondeau, G. et al. Dynamic biogeochemical provinces in the global ocean. *Glob. Biogeochem. Cycles* <https://doi.org/10.1002/gbc.20089> (2013).
- Richter, D. J. et al. Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. Preprint at *bioRxiv* <https://doi.org/10.1101/867739> (2020).
- Dutkiewicz, S. et al. Dimensions of marine phytoplankton diversity. *Biogeosciences* <https://doi.org/10.5194/bg-17-609-2020> (2020).
- Hellweger, F. L., Van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* <https://doi.org/10.1126/science.1254421> (2014).
- Laso-Jadart, R. et al. Investigating population-scale allelic differential expression in wild populations of *Oithona similis* (Cyclopoida, Claus, 1866). *Ecol. Evol.* <https://doi.org/10.1002/ece3.6588> (2020).
- Delmont, T. O. et al. Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *eLife* <https://doi.org/10.7554/eLife.46497> (2019).
- Carradec, Q. et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* <https://doi.org/10.1038/s41467-017-02342-1> (2018).
- Salazar, G. et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* <https://doi.org/10.1016/j.cell.2019.10.014> (2019).
- Alberti, A. et al. Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci. Data* <https://doi.org/10.1038/sdata.2017.93> (2017).
- Pesant, S. et al. Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* <https://doi.org/10.1038/sdata.2015.23> (2015).
- Karsenti, E. et al. A holistic approach to marine eco-systems biology. *PLoS Biol.* <https://doi.org/10.1371/journal.pbio.1001177> (2011).
- Duarte, C. M. Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. *Limnol. Oceanogr. Bull.* <https://doi.org/10.1002/lob.10008> (2015).
- Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1519080113> (2016).
- Benedetti, F., Guilhaumon, F., Adloff, F. & Ayata, S. D. Investigating uncertainties in zooplankton composition shifts under climate change scenarios in the Mediterranean Sea. *Ecography* <https://doi.org/10.1111/ecog.02434> (2018).
- Beaugrand, G. et al. Prediction of unprecedented biological shifts in the global ocean. *Nat. Clim. Change* **9**, 237–243 (2019).
- Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. & Levin, S. A. Marine taxa track local climate velocities. *Science* <https://doi.org/10.1126/science.1239352> (2013).
- Bopp, L. et al. Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosciences* <https://doi.org/10.5194/bg-10-6225-2013> (2013).
- Thomas, M. K., Kremer, C. T., Klausmeier, C. A. & Litchman, E. A global pattern of thermal adaptation in marine phytoplankton. *Science* <https://doi.org/10.1126/science.1224836> (2012).
- Ibarbalz, F. M. et al. Global trends in marine plankton diversity across kingdoms of life. *Cell* <https://doi.org/10.1016/j.cell.2019.10.008> (2019).
- Bussen, G. et al. Large scale patterns of marine diatom richness: drivers and trends in a changing ocean. *Glob. Ecol. Biogeogr.* <https://doi.org/10.1111/geb.13161> (2020).
- Hutchinson, G. E. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415–427 (1957).
- Delmont, T. O. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.10.15.341214> (2020).

32. Delmont, T. O. et al. Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *ISME J.* <https://doi.org/10.1038/s41396-021-01135-1> (2021).
33. Boyer, et al. *World Ocean Database 2013, NOAA Atlas NESDIS 72* (National Oceanic and Atmospheric Administration, 2013); <https://doi.org/10.7289/V5NZ85MT>
34. Sunagawa, S. et al. Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/s41579-020-0364-5> (2020).
35. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-019-0336-3> (2019).
36. van Vuuren, D. P. et al. The representative concentration pathways: an overview. *Climatic Change* <https://doi.org/10.1007/s10584-011-0148-z> (2011).
37. Polovina, J. J., Dunne, J. P., Woodworth, P. A. & Howell, E. A. Projected expansion of the subtropical biome and contraction of the temperate and equatorial upwelling biomes in the North Pacific under global warming. *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsq198> (2011).
38. Flombaum, P., Wang, W. L., Primeau, F. W. & Martiny, A. C. Global picophytoplankton niche partitioning predicts overall positive response to ocean warming. *Nat. Geosci.* <https://doi.org/10.1038/s41561-019-0524-2> (2020).
39. Richardson, A. J. In hot water: zooplankton and climate change. *ICES J. Mar. Sci.* **65**, 279–295 (2008).
40. Wrightson, L. & Tagliabue, A. Quantifying the impact of climate change on marine diazotrophy: insights from Earth system models. *Front. Mar. Sci.* **7**, 635 (2020).
41. Zehr, J. P. & Capone, D. G. Changing perspectives in marine nitrogen fixation. *Science* **368**, eaay9514 (2020).
42. Luo, Y.-W. et al. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data* **4**, 47–73 (2012).
43. Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677–680 (1979).
44. Laws, E. A., Falkowski, P. G., Smith, W. O., Ducklow, H. & McCarthy, J. J. Temperature effects on export production in the open ocean. *Glob. Biogeochem. Cycles* **14**, 1231–1246 (2000).
45. Agrawal, R. & Srikant, R. in *Proceedings of the 20th International Conference on Very Large Data Bases* (eds Bocca, J. B. et al.) 487–499 (Morgan Kaufmann, 1994).
46. Laufkötter, C. et al. Projected decreases in future marine export production: the role of the carbon flux through the upper ocean ecosystem. *Biogeosciences* **13**, 4023–4047 (2016).
47. Iudicone, D. Some may like it hot. *Nat. Geosci.* <https://doi.org/10.1038/s41561-020-0535-z> (2020).
48. Gorsky, G. et al. Expanding Tara Oceans protocols for underway, ecosystemic sampling of the ocean–atmosphere interface during Tara Pacific expedition (2016–2018). *Front. Mar. Sci.* <https://doi.org/10.3389/fmars.2019.00750> (2019).
49. Istace, B. et al. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* <https://doi.org/10.1093/gigascience/giw018> (2017).
50. Grand, M. M. et al. Developing autonomous observing systems for micronutrient trace metals. *Front. Mar. Sci.* <https://doi.org/10.3389/fmars.2019.00035> (2019).
51. Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P. & Deckmyn, A. maps: Draw geographical maps. R version 3.5.0 <https://cran.r-project.org/web/packages/maps/index.html> (2021).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Genomic provinces of plankton. Environmental niches are computed for trans-kingdom plankton genomic provinces from Richter et al.¹¹. These provinces consist of the clustering of each of six metagenomic dissimilarity matrices (based on the amount of DNA *k*-mers shared between pairs of samples) from six available size fractions with sufficient metagenomic data from the *Tara* Oceans dataset. The six size fractions (0–0.2 μm, 0.22–3 μm, 0.8–5 μm, 5–20 μm, 20–180 μm and 180–2,000 μm) were initially sampled to represent major plankton groups^{18–20,34}. Two large size classes (180–2,000 μm and 20–180 μm) are enriched in zooplankton dominated by arthropods (mainly copepods) and cnidarians. Size classes 5–20 μm and 0.8–5 μm are enriched in smaller eukaryotic algae such as dinofytes (5–20 μm), pelagophytes and haptophytes (0.8–5 μm). Finally, size classes 0.22–3 μm and 0–0.2 μm are respectively enriched in bacteria and viruses. Within each size fraction (from large to small), there are respectively 8, 8, 11, 6, 6 and 8 (48 in total) provinces defined in Richter et al.¹¹ formed by *Tara* Oceans metagenomes (644 metagenomes sampled either at the surface or at the deep chlorophyll maximum across 102 sites).

Genome signature of the provinces. We analysed the distribution of 713 eukaryotic and 1,888 prokaryotic genomes^{31,32} within the genomic provinces. These genomes are metagenome-assembled genomes (MAGs) obtained from *Tara* Oceans metagenomes. For each size class, we select MAGs that are present (according to a criteria defined in Delmont et al.³¹) in at least five samples. We computed an index of presence enrichment of MAGs within provinces as the Jaccard index (J)⁵², defined as follows:

$$J = \frac{M_{11}}{M_{11} + M_{01} + M_{10}} \quad (1)$$

M_{11} is the number of samples where the MAG is present and matches a sample of the province. M_{01} and M_{10} are, respectively, the number of samples where the MAG is not present in a sample of the province and, reciprocally, the number of samples where the MAG is present outside the province. A MAG is considered to be signature of a province if the Jaccard index is superior to 0.5 for this province and inferior to 0.1 for all other provinces of the given size class (Fig. 1 and Extended Data Fig. 2).

WOA13 data. Physicochemical parameters proposed to have an impact on plankton genomic provinces¹¹ are used to define environmental niches: SST, salinity (Sal), dissolved silica (Si), dissolved nitrate (NO₃), dissolved phosphate (PO₄), dissolved iron (Fe) and a seasonality index of nitrate (SINO₃). With the exception of Fe and SINO₃, these parameters are extracted from the gridded WOA13³⁵. Climatological Fe fields are provided by the biogeochemical model PISCES-v2³⁵. SINO₃ is defined as the range of nitrate concentrations over the year in one grid cell divided by the maximum range encountered in WOA13 at the *Tara* Oceans sampling stations:

$$SI([\text{NO}_3](x, y, z)) = \frac{\max_t([\text{NO}_3](x, y, z, t)) - \min_t([\text{NO}_3](x, y, z, t))}{\max_s(\max_t([\text{NO}_3](x, y, z, t)) - \min_t([\text{NO}_3](x, y, z, t)))} \quad (2)$$

[NO₃] is the nitrate concentration. The indices t and s , respectively, refer to time (months) and spatial (*Tara* samples) coordinates. All parameters are co-located with the corresponding stations and extracted at the month corresponding to the *Tara* Oceans sampling. To compensate for missing physicochemical samples in the *Tara* Oceans in situ dataset, climatological data (WOA13) are used. The correlation between in situ samples and corresponding values extracted from WOA13 are high (r^2 : SST: 0.96, Sal: 0.83, Si: 0.97, NO₃: 0.83, PO₄: 0.89). In the absence of corresponding WOA13 data, a search is done within 2° around the sampling location and values found within this square are averaged.

Nutrients such as NO₃ and PO₄ display a strong collinearity when averaged over the global ocean (correlation of 0.95 in WOA13), which could complicate disentangling their respective contributions to niche definition. However, observations and experimental data allow identification of limiting nutrients at the regional scale, characterized by specific plankton communities⁵⁶. The projection of niches into the future climate would yield spurious results if the present-day collinearity is not maintained⁵⁷, but there is up to now no evidence for large-scale changes in global nutrient stoichiometry⁵⁸.

ESM and bias correction. Outputs from six ESMs (Supplementary Table 2) are used to project environmental niches under greenhouse gas concentration trajectory RCP8.5³⁶. Environmental drivers are extracted south of 60°N for present-day (2006–2015) and end-of-century (2090–2099) conditions for each model, and the multi-model mean is computed. A bias correction method, the cumulative distribution function transform (CDFt)⁵⁹, is applied to adjust the distributions of SST, Sal, Si, NO₃ and PO₄ of the multi-model mean to the WOA13 database. CDFt is based on a quantile mapping approach to reduce the bias between modelled and observed data while accounting for climate change. Therefore, CDFt does not rely on the stationarity hypothesis and present and future

distributions can be different. CDFt is applied on the global fields of the mean model simulations. By construction, CDFt preserves the ranks of the simulations to be corrected. Thus, the spatial structures of the model fields are preserved.

Environmental niche models: training, validation and projections. Provinces with similar metagenomic content are retrieved from Richter et al.¹¹. From a total of 48 initial provinces, 10 provinces are removed either because they are represented by too few samples (7 out of 10) or they are found in environments not resolved by ESMs (for example, lagoons of Pacific Ocean islands, 3 out of 10). This narrows down the number of samples from 644 to 595 metagenomes. Four machine learning methods are applied to compute environmental niches for each of the 38 provinces: gradient boosting machine (gbm)⁶⁰, random forest (rf)⁶¹, single hidden layer neural networks (nn)⁶² and generalized additive models (gam)⁶³. Hyper parameters of each technique (except gam) are optimized. These are (1) for gbm, the interaction depth (1, 3 and 5), learning rate (0.01, 0.001) and the minimum number of observations in a tree node (1 to 10); (2) for rf, the number of trees (100 to 900 with step 200 and 1,000 to 9,000 with step 2,000) and the number of parameters used for each tree (1 to 7); (3) for nn, the number of neurons of the network (1 to 10) and the decay (1.10⁻⁴ to 9.10⁻⁴ and 1.10⁻⁵ to 9.10⁻⁵). For gam the number of splines is set to 3, respectively 2 only when not enough points are available (for fraction 0–0.2, 65 points). R packages 'gbm' (2.1.3), 'randomForest' (4.6.14), 'mgcv' (1.8.16) and 'nnet' (7.3.12) are used for gbm, rf, nn and gam models.

To define the best combination of hyper parameters for each model, we perform random cross-validation by training the model on 85% of the dataset randomly sampled and by calculating the area under the curve⁶⁴ (AUC) on the 15% remaining points of the dataset. This process is repeated over 30 random subsets of the entire training set for each combination of hyperparameters. We work in a presence/absence framework, that is, for each province separately, the dataset consists of the variable to predict ('presence' or not of the province in the sample) and the predictors consisting of the environmental variables for each sample. A fixed probability threshold of 0.5 for presence/absence detection is used to calculate the AUC. Fixing the probability threshold allows optimization of all models according to this threshold so that the *dominant* province has a reasonably high probability of presence (at least in regions with similar environmental parameters to the training dataset) and for the four types of statistical model we use. The best combination of hyper parameters is the one for which the mean AUC over the cross-validation is the highest. A model is considered valid if at least three out of the four techniques have a mean AUC superior to 0.65, which is the case for 27 out of the 38 provinces (Supplementary Fig. 1a). A climatic annotation is given to the 27 validated niches (Supplementary Table 2). Final models are trained on the full dataset and only the techniques that have a mean AUC higher than 0.65 are considered to make the projections. The vast majority (23) of the 27 validated niches is validated by all 4 models and the remaining 4 niches by only 3 models. Relative influences of each parameter in defining environmental niches are calculated using the 'feature_importance' function from the DALEX R package⁶⁵ for all four statistical methods (Supplementary Fig. 16a). To evaluate the consistency and coherence of environmental niche models, we first make global projections on the 2006–2013 WOA2013 climatology. Projections are consistent with sampling regions for provinces encompassing vast oceanic areas. For example, the genomic province sampled in temperate Atlantic regions of size fraction 180–2,000 μm is projected to be present in the north and south temperate Atlantic but also other temperate regions (Supplementary Fig. 2). For model training and projections, physicochemical variables are scaled to have a mean of 0 and a variance of 1 (using means and standard deviations of the training data). This standardization procedure allows for better performance of neural networks models. Finally, as statistical models often disagree on projection sets whereas they give similar predictions on the training set (Supplementary Figs. 4, 5), we use the ensemble model approach for global-scale projections of provinces⁶⁶, that is, the mean projections of the validated machine learning techniques.

Combined size class provinces and ocean partitioning comparisons. To combine all size classes' provinces, we use the PHATE algorithm^{65,67} from the R package 'phateR'. This algorithm allows visualization of high dimensional data in the requested number of dimensions while best preserving the global data structure⁶⁷. We choose to train PHATE separately on WOA13 projections and present-day and end-of-century projections including presence probabilities of *non dominant* provinces. We use three dimensions and set hyper parameter *k*-nearest neighbours (knn) and decay, respectively, to 1,000 for WOA13 and 2,000 for model data as in this case there are twice as many points. The hyper parameter knn reflects the degree to which the mapping of PHATE from high to low dimensionality should respect the global features of the data. We argue that 1,000 and 2,000 are good choices as it will be sufficient to have a highly connected graph, conserve global structure, allow visualization of structures of the size of the provinces (mean number of points in a province: 4,867) and have a reasonable computational time. Decay is set to 20 in both cases. Then we cluster the resulting distance matrix using the *k*-medoids algorithm⁶⁸, and the silhouette average width criteria⁶⁹ is used as an indicator of good fit. The silhouette criterion is maximal for two, three and four clusters, and two peaks are found at seven and 14 clusters (not shown). We choose

to present the four and seven cluster geographical patterns as they seem more relevant with respect to the resolutions of our environmental datasets (WOA13 and climate models). We compare the three polar clusters of the seven cluster geographical patterns with Antarctic Circumpolar Currents fronts⁷⁰ by overlying them on the map (black lines in Supplementary Fig. 3b).

To visualize the global biogeography structure, the resulting three vectors of PHATE are plotted using an RGB colour code. Each coordinate of each vector is respectively assigned to a given degree of colour component between 0 and 255 (8 bits red, green or blue) using the following formula (Supplementary Figs. 3 and 13):

$$C_{\text{col}}(i) = \frac{C(i) - \min(C(1), C(2), C(3))}{\max(C(1), C(2), C(3)) - \min(C(1), C(2), C(3))} \times 255 \quad (3)$$

$C(i)$ is the i th component of the PHATE axes and $C_{\text{col}}(i)$ is the corresponding degree of color of $C(i)$. Respectively, components 1, 2 and 3 are assigned to red, green and blue.

To compare the six size fraction biogeographies, the combined size class biogeography and existing biogeochemical partitions of the oceans^{9,10}, we use the adjusted rand index⁷¹ (Supplementary Figs. 6–8) and overlay their boundaries above our partitions.

Centroids and migration shifts. The centroid of each province is defined as the average latitude and longitude for which the probability of presence is superior to 0.5 and weighted by both the probability of presence at each grid point and the grid cell area. The migration shift is calculated as the distance between the present-day and the end-of-the-century centroids considering the Earth as a perfect sphere of radius 6,371 km. For consistency (that is, to avoid long distance aberrant shifts), it is only calculated for provinces with an area of dominance larger than 10^6 km^2 in the given basin.

Bray–Curtis dissimilarity index. Climate change impact on global projections is calculated at each grid point as the Bray–Curtis dissimilarity index^{72,73} defined as follows:

$$BC = \frac{\sum_n |P_n^{\text{future}} - P_n^{\text{present}}|}{\sum_n |P_n^{\text{future}} + P_n^{\text{present}}|} \quad (4)$$

where (P_n^{present} and P_n^{future}) are, respectively, the probability of presence of the province n in present day and at the end of the century. Only the probabilities of *dominant* provinces are non-null, and all others are set to zero. The mask of main fisheries⁵³ (chosen as the first four deciles) and exclusive economic zones⁵⁴ are overlaid on the Bray–Curtis map.

Change in province assemblages. A province assemblage is defined as the assemblage of *dominant* provinces of each size fraction at a given grid point of the considered ocean. We consider two criteria of change in province assemblage between present-day and end-of-the-century conditions. The first one, more straightforward and less stringent, considers that a province assemblage occurs when a change of *dominant* province is found in at least one size fraction. In a more stringent way, a change of assemblage is considered important for $BC > \frac{1}{6}$ (previous section). This threshold corresponds to an idealized case where each *dominant* province has a probability of 1, and a change of *dominant* province is found in only one size fraction. This criterion allows us to discard assemblage changes for which the changes in probability of presence of *dominant* provinces are very low.

Differential genomic composition across provinces. We characterize the composition in marine Hexanauplia (copepods), marine diazotrophs and phototrophs of provinces by considering the mean relative abundances of groups of MAGs^{31,32} characterized taxonomically and by size (for copepods and phototrophs) in each province for all size fractions except the viral one and for size fractions $>5 \mu\text{m}$ (for marine Hexanauplia). Marine Hexanauplia are annotated taxonomically (either as belonging to clade A (109 MAGs) or clade B (105 MAGs)) from which 198 are found in at least five samples that we use (98 clade A and 100 clade B). To attribute a preferential size class to these MAGs, mean relative abundances over all sites of each of them are compared across size fraction 180–2,000 μm , 20–180 μm and 5–20 μm using Welch ANOVA⁷⁴ (p value <0.05). When the test is significant, the MAG is annotated to its preferential size class (Meso-zooplankton: 180–2,000 μm , Micro-zooplankton: 20–180 μm or 5–20 μm). When the test is not significant, the MAG is annotated as unclassified. Then for each group of MAG (defined by the preferential size class plus the clade), the mean of the sum over the MAGs from this group is calculated in each province to characterize the province. The same procedure is applied for 27 out of 48 bacterial diazotrophs (present in at least five samples) from the prokaryotic MAG collection³², distinguishing groups at the phylum level (Gammaproteobacteria $n=8$, Cyanobacteria $n=8$, Deltaproteobacteria $n=2$, Alphaproteobacteria $n=4$, Planctomycetes $n=3$, Verrucomicrobiota $n=2$) and for 231 phototrophs (present in at least five samples) distinguishing algae ($n=172$), diatoms (Bacillariophyta, $n=11$) and cyanobacteria ($n=48$) and their preferential size class (pico: 0.22–3 μm , nano: 0.8–5 μm and

5–20 μm and micro: 20–180 μm and 180–2,000 μm). Finally, significant differential composition between provinces are annotated using Holm corrected⁷⁵ pairwise Mann–Whitney U test⁷⁶ ($p <0.05$ for significance).

Estimation of carbon export fluxes in assemblages of provinces. We use three sets of carbon export data at 100 m depth (extrapolated data)^{34,33,44} to estimate an average particulate organic carbon (POC) flux for each assemblage of *climato-genomic* provinces. These extrapolations are based on estimates of carbon export efficiency from in situ measurements either using the POC/²³⁴Th radioisotope ratio³ or the f -ratio technique^{43,44}. They both derive a relation between carbon export efficiency and SST. Then to obtain extrapolations of carbon export, these ratios are multiplied by estimates of primary production from satellite estimates of sea surface chlorophyll concentrations. For each dataset, we associate each extrapolation of POC fluxes to the assemblage of present-day provinces from the nearest point of the flux estimate. An average flux is associated with an assemblage if at least three data points are associated with that assemblage. Each grid point is associated to a given assemblage and is assigned to its average carbon flux. Thus, we calculate carbon export estimates where an assemblage from the set of present-day assemblages can be projected in the future. Next we calculate total carbon export for current and projected end-of-century assemblages, summing fluxes over these areas (Supplementary Table 9). Similar conclusions are reached using median values of carbon export. We verified that using whole assemblages as predictors of carbon export is more robust than using single size fraction dominant provinces (ANOVA⁷⁷, Akaike Information Criterion⁷⁸).

Linking changes in carbon export and compositional changes. We use the Apriori algorithm⁴⁵ to identify associations between changes in community composition and changes in carbon export. From the relative abundance values of phototroph, diazotroph and copepod MAGs^{31,32}, we consider only significant changes in their composition among communities by a pairwise Wilcoxon test (Holm correction $p <0.05$). We also consider only phototrophic and copepod MAGs for which a preferred size class was found. The Apriori algorithm determines association rules between sets of ‘transactions’, here community changes with carbon export. The algorithm identifies sufficiently frequent associations between transactions and computes a lift as follows:

$$\text{lift}(A \rightarrow B) = \frac{\text{confidence}(A \rightarrow B)}{\text{support}(B)} \quad (5)$$

where $\text{confidence}(A \rightarrow B)$ is the number of transactions where both A and B occur divided by the number of transactions where A occurs, and $\text{support}(B)$ is the percentage of transactions where B occurs in the whole dataset. The lift corresponds to the increase factor in likelihood that B occurs when A occurs. The Apriori algorithm is launched for four latitudinal zones of the ocean: equatorial, subtropical (north and south) and temperate/subpolar zones. Only the directionality of changes is considered (increase/decrease in either carbon flux or a type of organism). If rules overlap perfectly, only the largest association rules are kept (maximum rule length is set to 7). Only association rules pointing towards changes in carbon fluxes are kept ($A = \text{change(s)}$ in one or several plankton group and $B = \text{change in carbon export in equation (5)}$). Minimum support is set to 0.05.

Driver analysis. To assess the relative importance of each driver in province changes, the methodology from Barton et al.²² is adopted. For a set N of n provinces (individual provinces or all provinces together), the probability of presence of each province is recalculated for present-day conditions except for driver d (from the set of drivers D) for which the end-of-the-century condition is used (P_n^{future} for d th driver only). The set of driver D can be either all drivers (Fig. 5a,c) or all drivers except SST (Fig. 5b). The relative importance of driver d at a given grid point for the set of N of provinces is computed as follows:

$$RI(d) = \frac{\sum_{n \in N} |P_n^{\text{future}} \text{ for } d\text{th driver only} - P_n^{\text{present}}|}{\sum_{d \in D} \sum_{n \in N} |P_n^{\text{future}} \text{ for } d\text{th driver only} - P_n^{\text{present}}|} \quad (6)$$

$RI(d)$ is computed at grid cells where $BC > \frac{1}{6}$ and is calculated with either the set of all drivers (Fig. 5a,c) or all drivers except SST (Fig. 5b). When $RI(d)$ is calculated for individual provinces (Fig. 5c and Supplementary Fig. 18d,e), it is computed only at grid cells where $BC > \frac{1}{6}$ and the concerned province is either *dominant* in present-day and/or end-of-the-century conditions.

Maps. We use the continents map background from R package ‘maps’⁵¹ for all map figures in this study.

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

Data Availability

All data used are available at <https://github.com/institut-de-genomique/NCLIM-20102618B>. All coordinates of ocean partitionings from this study

are available at https://figshare.com/articles/dataset/Biogeographies_genomic_provinces/19071620⁹.

Code Availability

All codes used are available at <https://github.com/institut-de-genomique/NCLIM-20102618B>.

References

52. Jaccard, P. Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales. *Bull. Murith.* **31**, 81–92 (1902).
53. Watson, R. A. A database of global marine commercial, small-scale, illegal and unreported fisheries catch 1950–2014. *Sci. Data* <https://doi.org/10.1038/sdata.2017.39> (2017).
54. *Maritime Boundaries Geodatabase: Maritime Boundaries and Exclusive Economic Zones (200NM), version 11* (Flanders Marine Institute, 2019); <https://doi.org/10.14284/386>
55. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* <https://doi.org/10.5194/gmd-8-2465-2015> (2015).
56. Bibby, T. S. & Moore, C. M. Silicate:nitrate ratios of upwelled waters control the phytoplankton community sustained by mesoscale eddies in sub-tropical North Atlantic and Pacific. *Biogeosciences* <https://doi.org/10.5194/bg-8-657-2011> (2011).
57. Brun, P., Kjørboe, T., Licandro, P. & Payne, M. R. The predictive skill of species distribution models for plankton in a changing climate. *Glob. Change Biol.* <https://doi.org/10.1111/gcb.13274> (2016).
58. Redfield, A. C. in *James Johnstone Memorial Volume* (ed. Daniel, R. J.) 176–192 (Liverpool Univ. Press, 1934).
59. Michelangeli, P. A., Vrac, M. & Loukos, H. Probabilistic downscaling approaches: application to wind cumulative distribution functions. *Geophys. Res. Lett.* <https://doi.org/10.1029/2009GL038401> (2009).
60. Ridgeway, G. gbm: Generalized boosted regression models. R version 1.6–3.1 <https://cran.r-project.org/web/packages/gbm/gbm.pdf> (2010).
61. Breiman, L. & Cutler, A. randomForest: Breiman and Cutler's random forests for classification and regression. R package 4.1.0 <https://www.stat.berkeley.edu/~breiman/RandomForests/> (2012).
62. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4th edn (Springer, 2002).
63. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* <https://doi.org/10.1198/01621450400000980> (2004).
64. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* <https://doi.org/10.1016/j.patrec.2005.10.010> (2006).
65. Biecek, P. DALEX: explainers for complex predictive models. *J. Mach. Learn. Res.* **19**, 1–5 (2018).
66. Jones, M. C. & Cheung, W. W. L. Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsu172> (2015).
67. Vallejos, C. A. Exploring a world of a thousand dimensions. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-019-0330-9> (2019).
68. Kaufman, L. and Rousseeuw, P.J. in *Statistical Data Analysis Based on the L1 Norm and Related Methods* (ed. Dodge, Y.) 405–416 (North-Holland, 1987).
69. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7) (1987).
70. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Res. Part I* [https://doi.org/10.1016/0967-0637\(95\)00021-W](https://doi.org/10.1016/0967-0637(95)00021-W) (1995).
71. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* <https://doi.org/10.1007/BF01908075> (1985).
72. Somerfield, P. J. Identification of the Bray–Curtis similarity index: comment on Yoshioka (2008). *Mar. Ecol. Prog. Ser.* <https://doi.org/10.3354/meps07841> (2008).
73. Bloom, S. Similarity indices in community studies: potential pitfalls. *Mar. Ecol. Prog. Ser.* <https://doi.org/10.3354/meps005125> (1981).
74. Welch, B. L. The generalisation of student's problems when several different population variances are involved. *Biometrika* **34**, 28–35 (1947).
75. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
76. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
77. Sthle, L. & Wold, S. Analysis of variance (ANOVA). *Chemom. Intell. Lab. Syst.* **6**, 259–272 (1989).
78. Bozdogan, H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika* **52**, 345–370 (1987).
79. Frémont, P. et al. Biogeographies of genomic provinces from 'Restructuring of plankton genomic biogeography in the surface ocean under climate change'. *figshare*. https://figshare.com/articles/dataset/Biogeographies_genomic_provinces/19071620 (2022).

Acknowledgements

P.F. was supported by a CFR doctoral fellowship and the NEOGEN impulsion grant from the Direction de la Recherche Fondamentale of the CEA. This study received funding from the European Union's Horizon 2020 Blue Growth research and innovation programme under grant agreement number 862923 (project AtlantECO), ATIGE Genopole postdoctoral fellowship (T.O.D.), HYDROGEN/ANR-14-CE23-0001 (T.O.D.) and ANR-11-IDEX-0004-17-EURE-0006. M.G. acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 820989 (project COMFORT). This study benefited from access to high-performance computing resources through GENCI- [TGCC/CINES/IDRIS] and the ESPRI computing and data centre (<https://mesocentre.ipsl.fr>), which is supported by CNRS, Sorbonne Université, Ecole Polytechnique and CNES and through national and international grants. We thank the commitment of the Research Federation for the Study of Global Ocean Systems Ecology and Evolution (FR2022/TaraGOSEE) and of Stazione Zoologica Anton Dohrn. We thank T. Roy for preparation of the climatic data, S. Henson for providing carbon export data, LAGE (Laboratoire d'Analyses Génomiques des Eucaryotes, CEA) members for stimulating discussions on this project, M. Mariadassou, S.D. Ayata and B.H. Mele for discussions on statistics and climate envelope models, C. Scarpelli and members of the scientific computation team from Genoscope for support on computations, L. Bopp for initial discussions on this project and on climate models and N. Le Bescot (TernogDesign) for help with the figures. We thank all members of the Tara Oceans consortium for maintaining a creative environment and for their constructive criticism. Tara Oceans would not exist without the Tara Ocean Foundation and the continuous support of 23 institutes (<https://oceans.taraexpeditions.org/>).

This article is contribution number 128 of Tara Oceans.

Author contributions

P.F., M.G. and O.J. conceived the study. P.F. computed the results and compiled and analysed the data. M.G., O.J. and J.L. conducted a preliminary study. M.V. wrote the bias correction algorithm. P.F. wrote the initial draft of the paper. T.O.D., P.F., M.G., O.J., D.I., J.L., M.V. and P.W. discussed the results and contributed to writing the paper.

Competing interests

The authors declare no competing interests.

Additional information

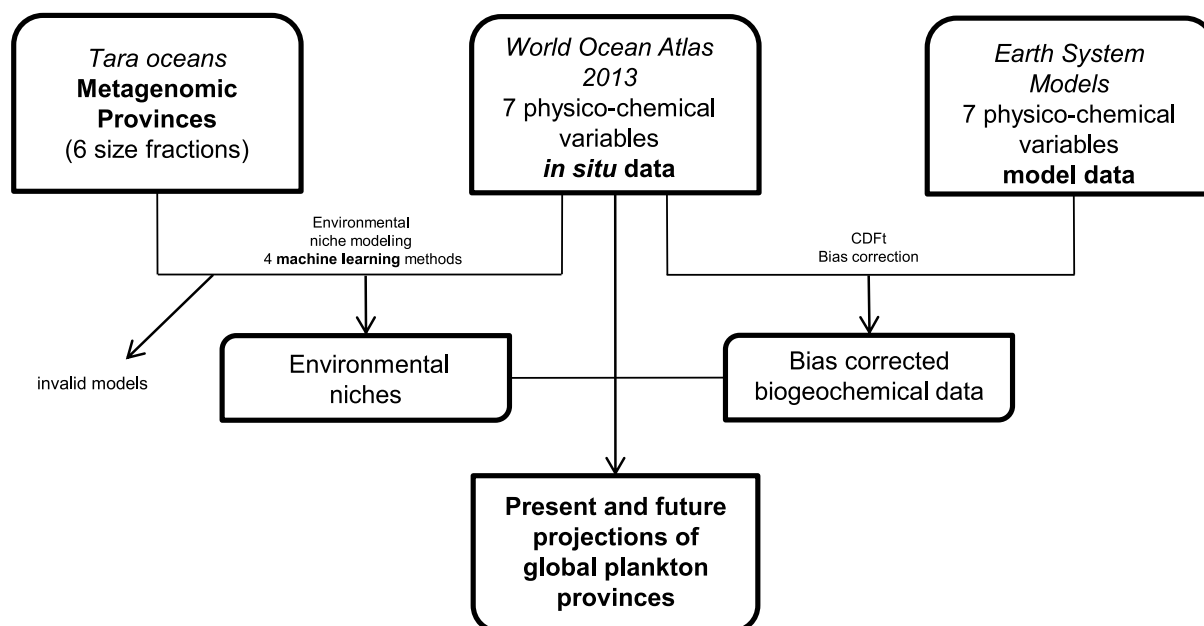
Extended data is available for this paper at <https://doi.org/10.1038/s41558-022-01314-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41558-022-01314-8>.

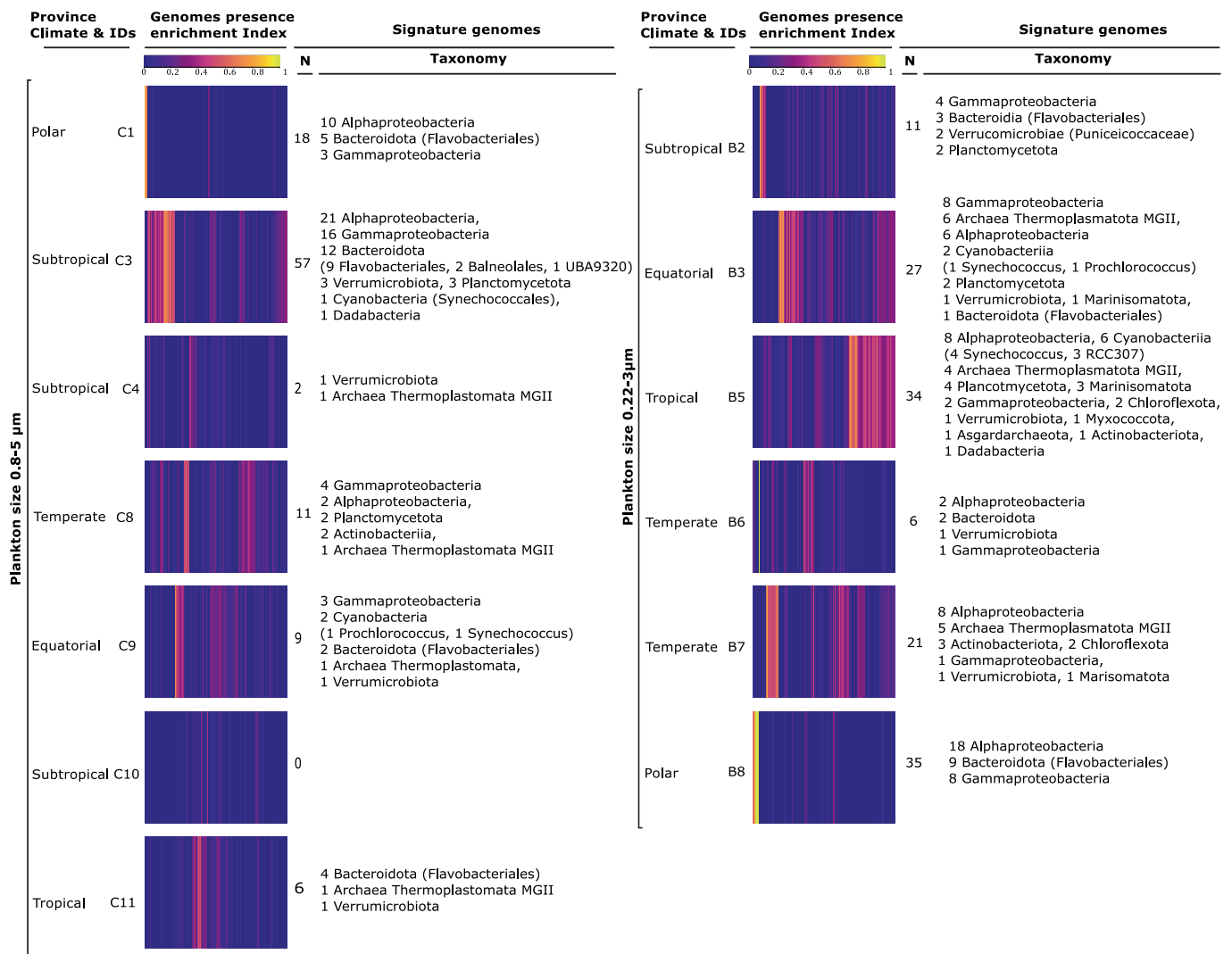
Correspondence and requests for materials should be addressed to Paul Frémont, Marion Gehlen or Olivier Jaillon.

Peer review information *Nature Climate Change* thanks Levente Bodrossy, Robert Ptacnik and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

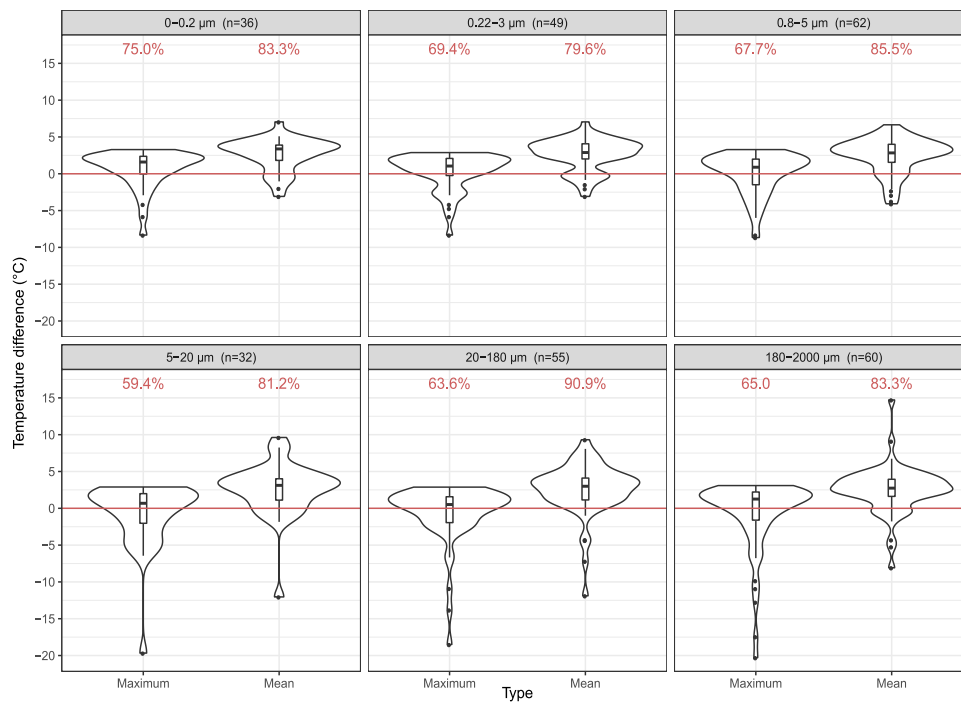
Reprints and permissions information is available at www.nature.com/reprints.



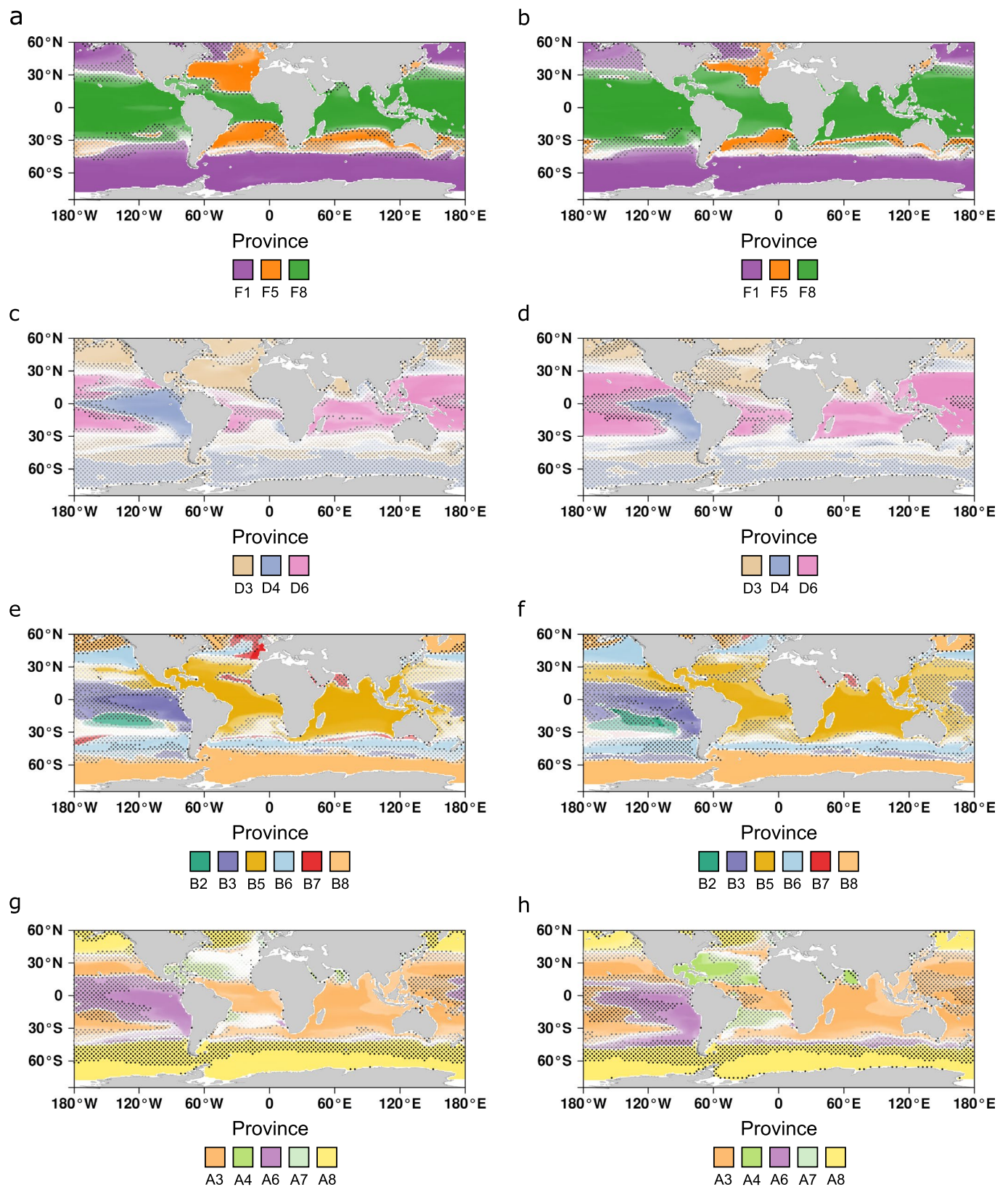
Extended Data Fig. 1 | Study pipeline. Metagenomic data from the 2009–2013 Tara Oceans expedition and in situ measurements of physicochemical variables (World Ocean Atlas 2013, WOA13)³³ are combined to define environmental niches at the plankton community level across 6 size fractions. Bias corrected outputs from a mean model of 6 Earth System Models (Supplementary Table 1) and WOA13 data are then used to project global plankton provinces for present day and end of the century conditions under a high warming scenario (RCP8.5)³⁶. Variables are Sea Surface Temperature (SST), Salinity (Sal), Dissolved silica (Si), Nitrate (NO₃), Phosphate (PO₄), Iron (Fe) and a seasonality index of nitrate (SI NO₃).



Extended Data Fig. 2 | Prokaryotic signature genomes of provinces of the prokaryote (0.22–3 µm) and protist (0.8–5 µm) enriched size classes. Indexes of presence enrichment⁵² for 1888 genomes of prokaryotic plankton³² in corresponding provinces are clustered and represented in a colour scale. Signature genomes (see Methods) are found for almost all provinces, their number and taxonomies are summarized (detailed list in Supplementary Table 6). A genome is considered to be signature of a province if the presence enrichment index is superior to 0.5 with this province and inferior to 0.1 for all other provinces of the given size class.

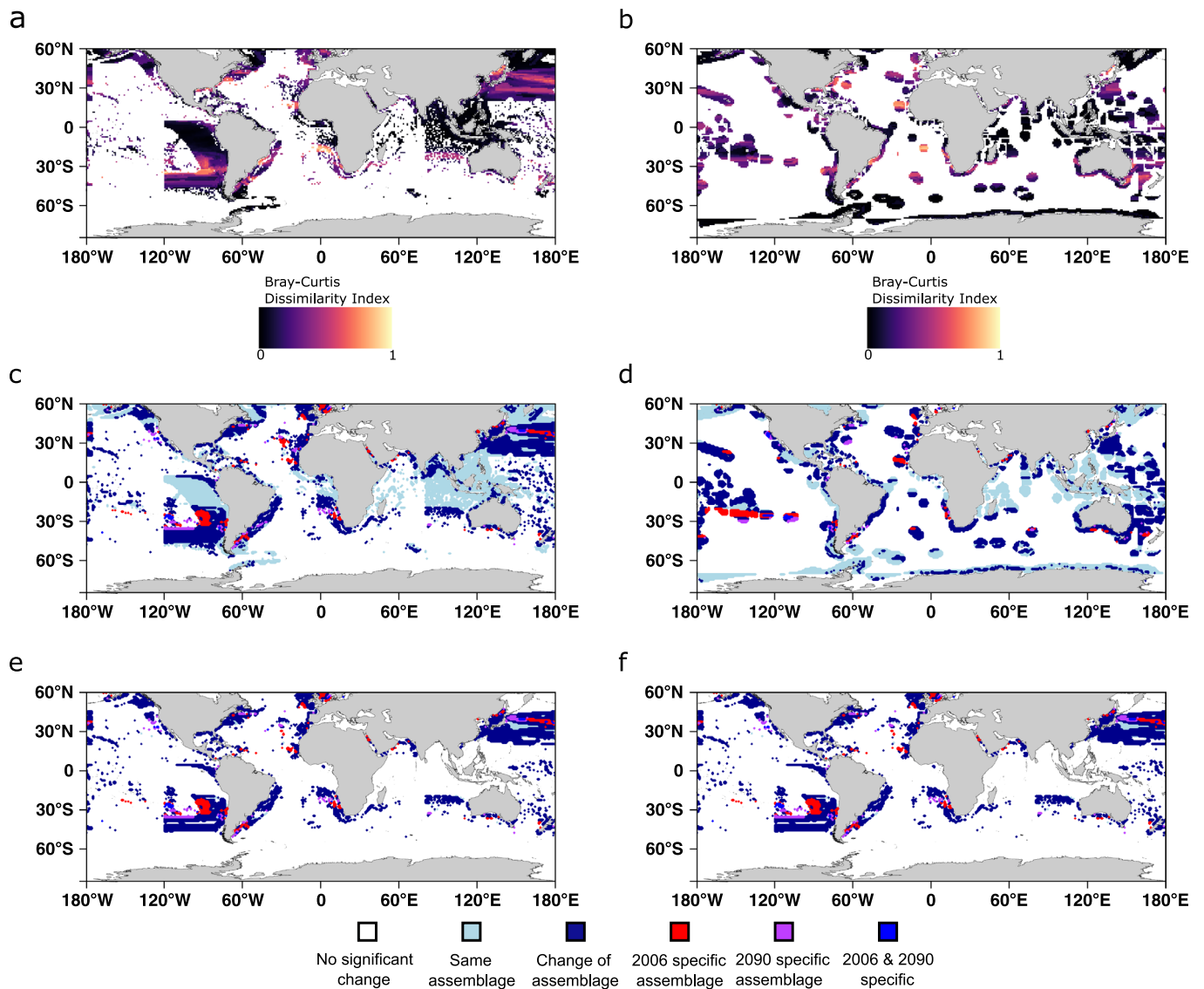


Extended Data Fig. 3 | Distribution of deltas between future temperature at each sampling site (surface) minus either the mean or maximum temperature within their contemporary genomic province. For most of the sites and across size fractions the future temperature projected by the bias adjusted ESM ensemble model is higher than both the maximum and mean contemporary temperature of their genomic province.

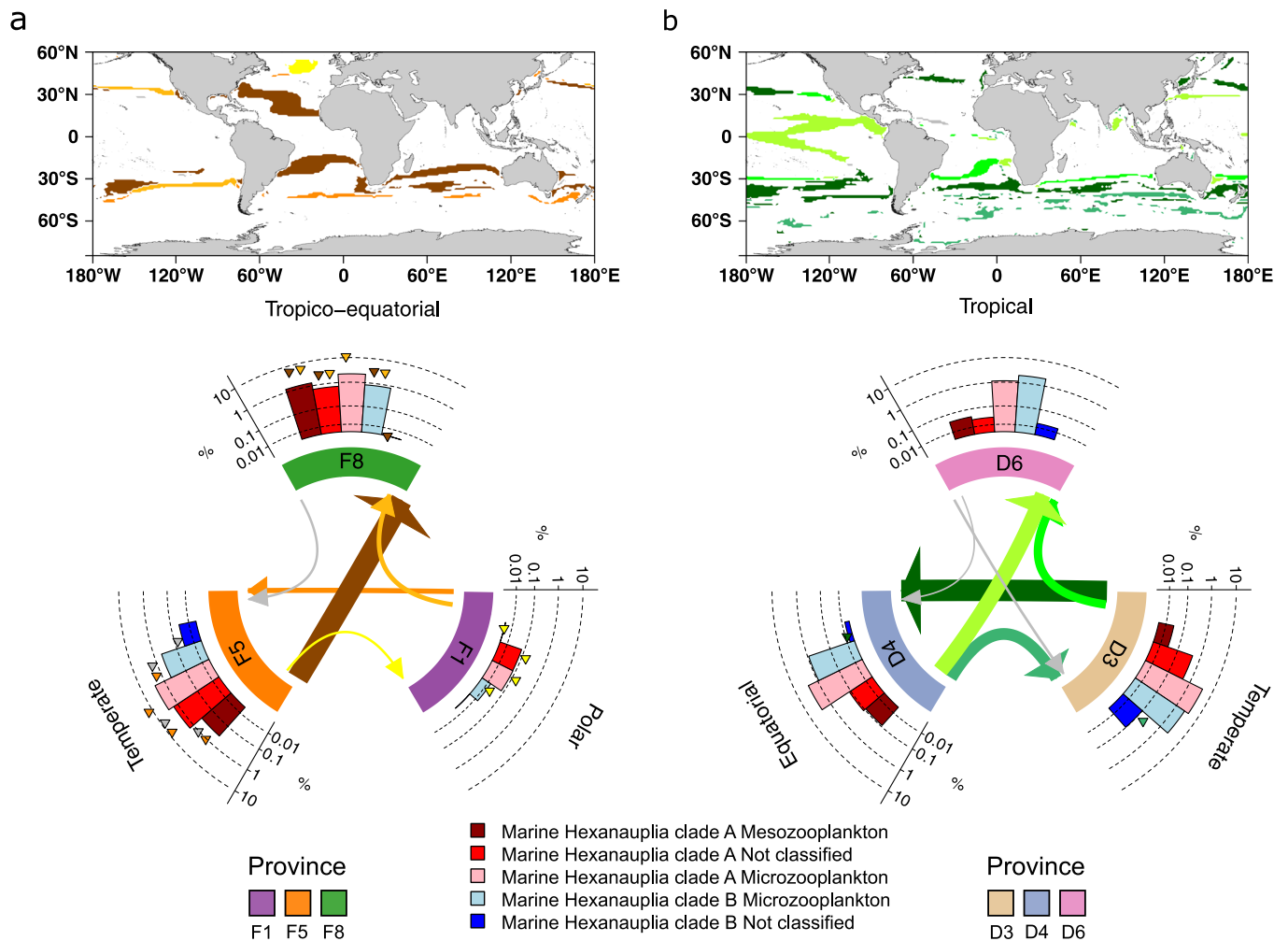


Extended Data Fig. 4 | Global geographical patterns for provinces of four plankton size fractions in present day and at the end of the century.

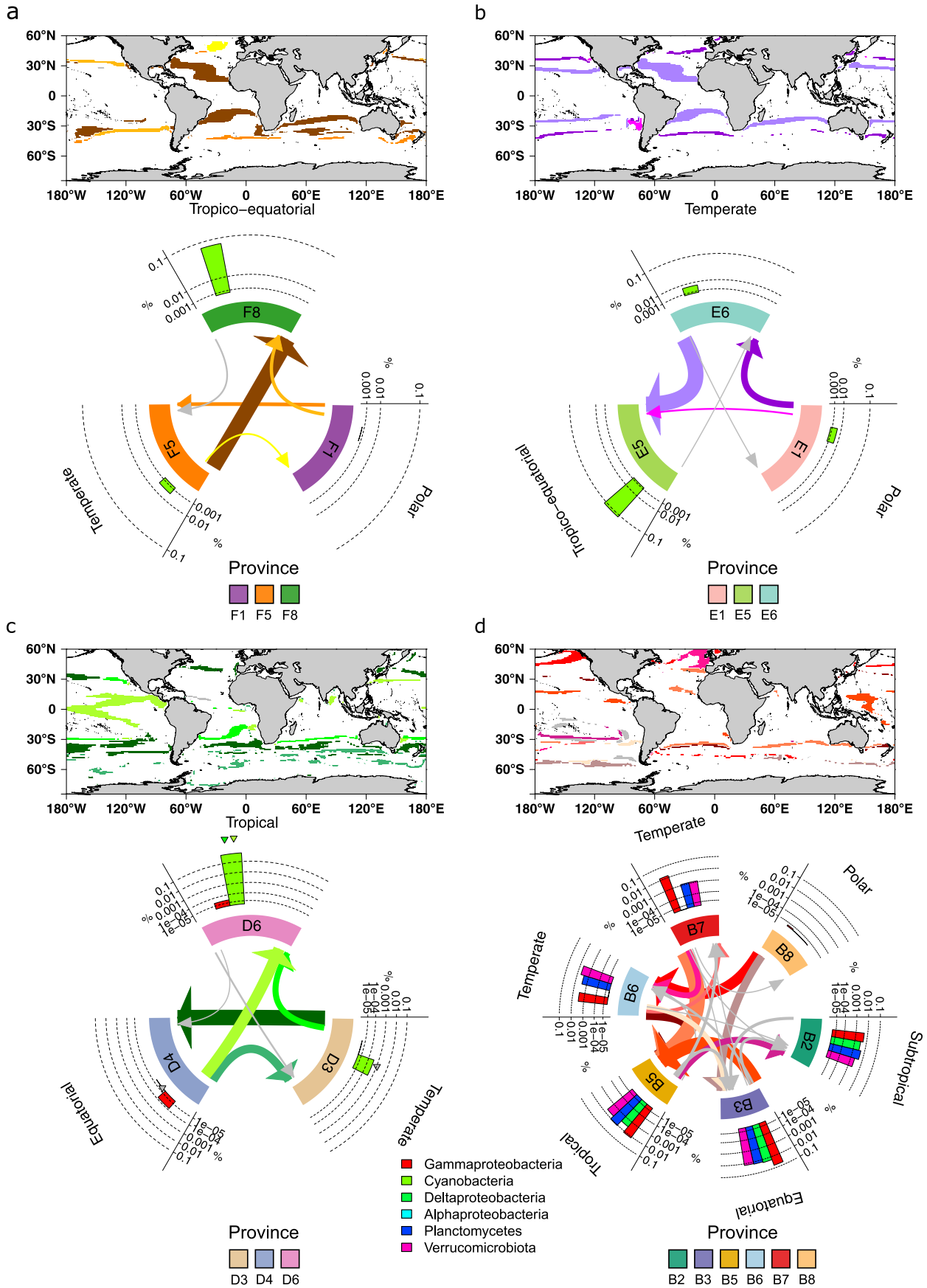
(a, c, e, g) Present day and (b, d, f, h) end of century biogeographies of size classes 180–2000, 5–20, 0.22–3 and 0–0.2 μm respectively. At each grid point of the maps the *dominant* province is represented using a darkness of colour proportional to its presence probability. Dots represent areas of uncertainty (where the delta of probability between the *dominant* and another province is inferior to 0.5). Expansion of tropical provinces and shrinkage of temperate provinces are consistently projected in all size fractions. We generated these map using R-package maps⁵¹.



Extended Data Fig. 5 | Bray-Curtis dissimilarity index and assemblage change maps comparing present day with end of the century projections of dominant provinces in principal fisheries⁵³ (4 last deciles) and Exclusive Economic Zones⁵⁴. Bray-Curtis dissimilarity index and assemblage changes in (a, c) Principal fisheries and (b, d) Exclusive Economic Zones. Assemblage changes in (e) Principal fisheries and (f) Exclusive Economic Zones in areas projected to encounter an important change (Bray-Curtis dissimilarity index superior to 1/6). We generated these map using R-package maps⁵¹.

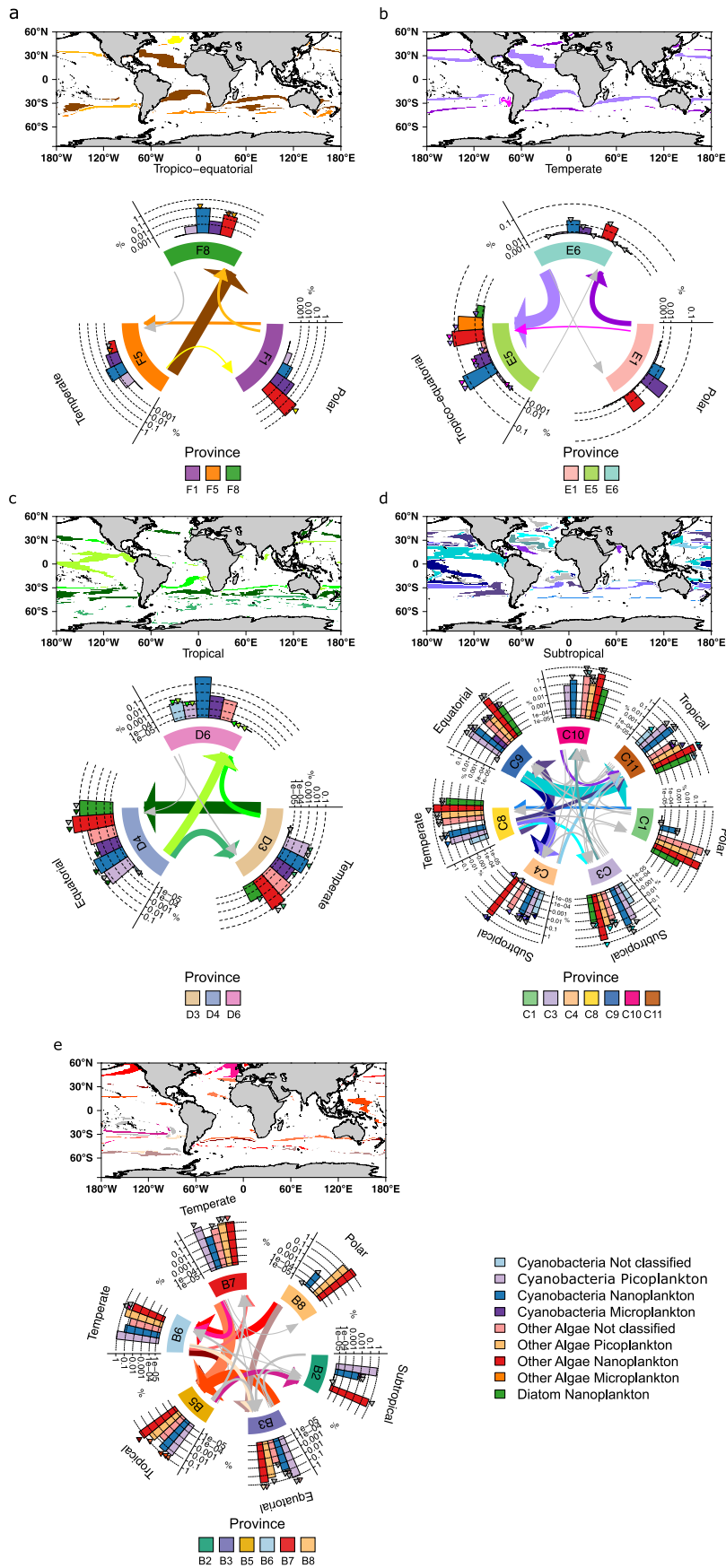


Extended Data Fig. 6 | Projected compositional shifts in marine hexanauplia in areas of *dominant* province change. (a) 180–2000 μm and (b) 5–20 μm . Top: Locations of *dominant* province change using colours corresponding to the type of province transition. Bottom: Circular plots summarizing significant compositional shifts in marine hexanauplia classified by size ('not classified' when no preferential size class is found). Each type of transition is represented by an arrow coloured according to the map and in grey if they represent less than 2% of the transitions. Barplots represent mean relative abundances of each group of organism. Arrows point towards the end of the century projected province and their widths are proportional to the area of change. Significant compositional changes in a type of organism are represented by triangles of the associated transition colour. We generated these map using R-package maps⁵¹.



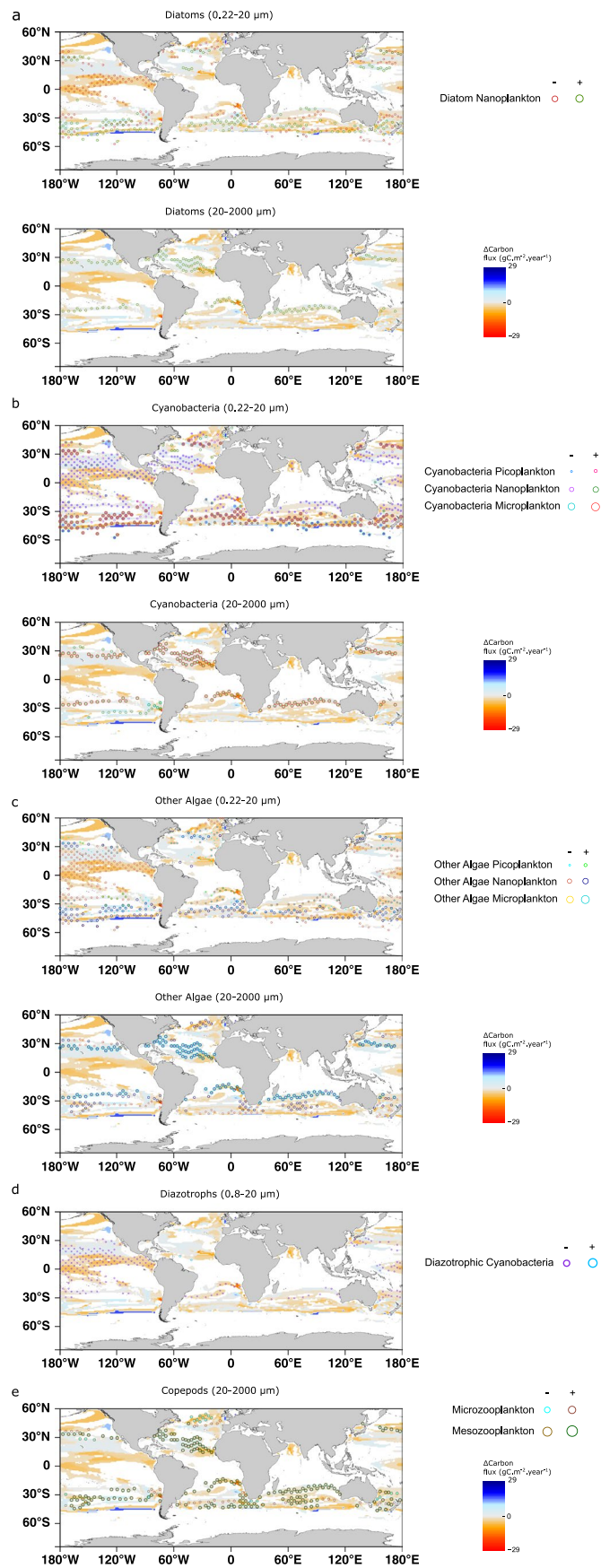
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Projected compositional shifts in bacterial diazotrophs in areas of dominant community change. (a) 180–2000 μm (b) 20–180 μm (c) 5–20 μm and (d) 0.22–3 μm . Top: Locations of *dominant* province change using colours corresponding to the type of province transition. Bottom: Circular plots summarizing significant compositional shifts in marine diazotrophs. Each type of transition is represented by an arrow coloured according to the map and in grey if they represent less than 2% of the transitions. Barplots represent mean relative abundances of each group of organism. Arrows points towards the end of the century projected province and their widths are proportional to the area of change. Significant compositional changes in a type of organism are represented by triangles of the associated transition. We generated these map using R-package maps⁵¹.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Projected compositional changes in phototrophs in areas of dominant community change. (a) 180–2000 μm (b) 20–180 μm (c) 5–20 μm (d) 0.8–5 μm and (e) 0.22–3 μm . Top: Locations of *dominant* province change using colours corresponding to the type of province transition. Bottom: Circular plots summarizing significant compositional shifts in phototrophs classified by size ('not classified' when no preferential size class is found). Each type of transition is represented by an arrow coloured according to the map and in grey if they represent less than 2% of the transitions. Barplots represent mean relative abundances of each group of organism. Arrows point towards the end of the century projected province and their widths are proportional to the area of change. Significant compositional changes in a type of organism are represented by triangles of the associated transition colour. We generated these map using R-package maps⁵¹.



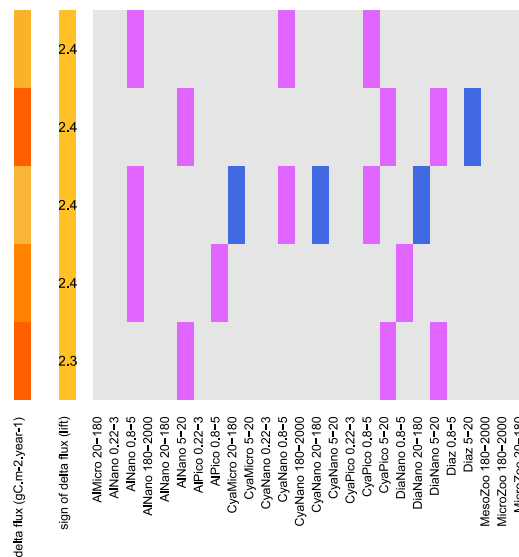
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Maps of carbon export flux changes in link with organisms' projected changes. Significant composition changes based on genomes relative abundances are represented for phototrophs, marine nitrogen fixers (Diazotrophic cyanobacteria) and copepods. For each map, transitions from several characteristic size classes are represented (**a**) Top: Diatoms 0.22–20 μm . Bottom: Diatoms 20–2000 μm (**b**) Top: Cyanobacteria 0.22–20 μm . Bottom: Cyanobacteria 20–2000 μm (**c**) Top: Other Algae 0.22–20 μm . Bottom: Other Algae 20–2000 μm (**d**) Diazotrophs 0.8–20 μm (**e**) Copepods 20–2000 μm . We generated these map using R-package maps⁵¹.

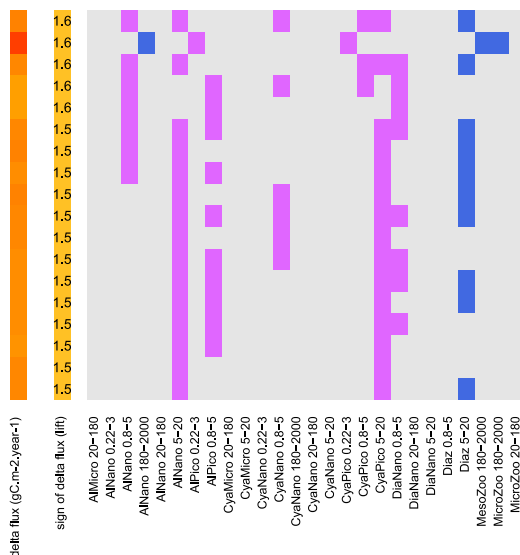
a Temperate/Subpolar



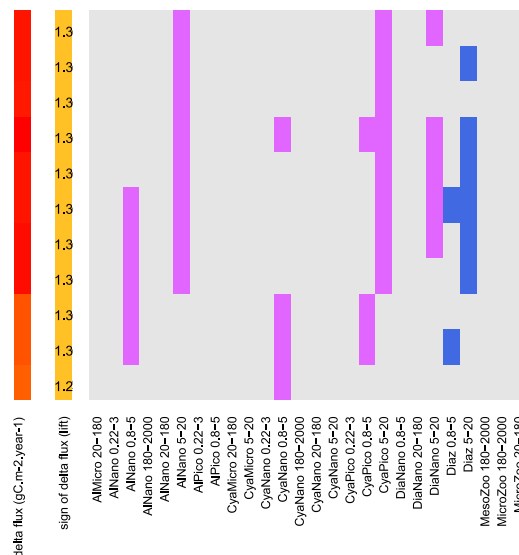
b Subtropical North



c Subtropical South



d Equatorial



Organisms' projected evolution:

Phototrophs

Types:
Al: Algae
Dia: Diatom
Cya: Cyanobacteria

Zooplankton

MicroZoo: Microzooplankton (20-180 μm)
MesoZoo: Mesozooplankton (180-2000 μm)

Diazotrophs

Diaz: Diazotrophs

Sizes:

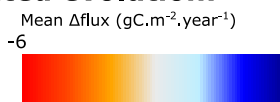
Pico: Picophytoplankton (0.22-3 μm)
Nano: Nanophytoplankton (0.8-20 μm)
Micro: Microphytoplankton (20-2000 μm)

Sign of relative abundance change



Carbon export projected evolution:

Sign of change Lift:
1.2 Increase
- + factor in
likeliness of
increase/decrease



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Association rules between changes in carbon flux and changes in organism relative abundances. Association rules in (a) temperate/subpolar (latitude $> 40^\circ$ ($< -40^\circ$)), (b) subtropical North (20° to 40° latitude), (c) subtropical South (-20° to -40° latitude) and (d) equatorial regions (-20° to 20° latitude). Each line represents an association rule between a change in carbon export found by the Apriori algorithm⁴⁵ (first column: mean change in carbon export and second column: sign of the change and lift of the rule (equation 5)). The other columns represent the changes in community composition (red: decrease of the given group, green: increase) associated with this change in carbon export.

Supplementary information

**Restructuring of plankton genomic
biogeography in the surface ocean under
climate change**

In the format provided by the
authors and unedited

Supplement of

**Restructuring of genomic biogeography of surface
ocean plankton under climate change**

Paul Frémont et al.

Correspondence to: Paul Frémont (pfremont@genoscope.cns.fr), Marion Gehlen (marion.gehlen@lsce.ipsl.fr) and Olivier Jaillon (ojailon@genoscope.cns.fr)

Supplementary discussion 1 | Niche models of genomic provinces

To compute and test the validity of realized environmental niches, we train four machine learning techniques to probabilistically associate genomic provinces with environmental data: sea surface temperature, salinity, three macronutrients (dissolved silica, nitrate and phosphate), one micronutrient (dissolved iron) plus a seasonality index of nitrate. The four machine learning techniques are Gradient Boosting Machine (gbm)¹, Random Forest (rf)², single hidden layer Neural Networks (nn)³ and Generalized Additive Models (gam)⁴. Following a cross-validation framework, a valid environmental niche is obtained for 27 out of 38 initial provinces (71%) comforting their definition and covering 529 samples out of 595 (89%, Supplementary Fig. 1). Rejected provinces contain relatively few stations (mean of 6 ± 2.6 versus 19 ± 15.3 for valid provinces, $p\text{-value} < 10^{-3}$ Wilcoxon test⁵). For spatial and temporal extrapolations of the provinces presented below, we use the ensemble model approach⁶ that considers mean predictions of machine learning techniques.

Supplementary discussion 2 | Integrated biogeographies using the PHATE algorithm

We use the PHATE dimension reduction algorithm⁷ to combine all provinces for all size classes into a single consensus biogeography revealing 4 or 7 robust clusters (Supplementary Fig. 3, *methods*). The 4 cluster consensus biogeography is mainly latitudinally organized distinguishing polar, subpolar, temperate and tropico-equatorial regions. The 7 cluster consensus biogeography distinguishes the equatorial pacific upwelling biome and three subpolar biomes that most likely reflect the chemico-physical structuring of the Southern Ocean and known polar fronts⁸ (red lines Fig. 2h). However, learning data are scarcer south of 60°S so these extrapolations need to be taken with caution.

PHATE is also used for modeled projections into two comparable consensus maps, one for present day and one for the end of the century (Supplementary Fig. 13). Some particularly visible patterns of geographical reorganization are common to several or even all size fractions and visible when comparing the two consensus maps (Supplementary Fig. 13 compared to Fig. 3a-d and Extended Data 4). For example, the

tropico-equatorial and tropical provinces expand in all size fractions and the provinces including the pacific equatorial upwelling shrink for size fractions smaller than 20 μm .

Supplementary discussion 3 | Comparison of the biogeographies with existing partitions of the oceans

Previous ocean partitioning either in biomes⁹⁻¹¹ or biogeochemical provinces (BGCPs)¹⁰ are based on physico-biogeochemical characteristics including SST⁹⁻¹¹, chlorophyll *a*⁹⁻¹¹, salinity⁹⁻¹¹, bathymetry^{9,10}, mixed layer depth¹¹ or ice fraction¹¹. Considering three of these partitions as examples we notice differences with our partitions (Supplementary Fig. 6-7) for example in terms of the number of regions in the considered oceans (56 for 2013 Reygondeau et al. BGCPs¹⁰, 17 for Fay and McKingley¹¹) and their structure (the coastal biome for 2013 Reygondeau et al. biomes¹⁰). Numerical comparison of our partitions with others (*Methods*) reveals low similarity between them, the highest being with Reygondeau biomes (Supplementary Figs. 6-8). However, biomes and BGCP frontiers closely match our province frontiers in many cases. Near the frontiers, *dominant* provinces have smaller probabilities in agreement with smooth transitions instead of sharp boundaries as already proposed⁹ and with a seasonal variability of the frontiers¹⁰ (Supplementary Fig. 6). Some of these transitions are very large and match entire BGCPs, for example in subtropical North Atlantic and subpolar areas where high annual variations are well known¹⁰.

Supplementary discussion 4 | Shifts of the *climato-genomic* provinces in response to climate change

Centroids of provinces with *dominance* areas larger than 10⁶ km² within a basin would be moved at least 200 km away for 77% of them, 96% of which move poleward (Supplementary Figs. 11 and 12). While a few longitudinal shifts larger than 1000 km are projected, the distribution of latitudinal shifts is largely concentrated around the mean (290 km) with no shifts superior to 1000 km (Supplementary Fig. 12b). These important longitudinal shifts corroborate existing projections¹²⁻¹⁴ and differ from trivial poleward shifts due to temperature increase, reflecting more complex spatial rearrangements of the other environmental drivers (Supplementary Fig. 10). The

average displacement speed of the provinces' centroids is $76 \pm 79 \text{ km.dec}^{-1}$ (latitudinally mean of $34 \pm 82 \text{ km.dec}^{-1}$, longitudinally $59 \pm 82 \text{ km.dec}^{-1}$).

Projected shifts in phytoplankton enriched provinces corroborate previously published shifts of North Pacific phytoplankton biomes: provinces C4 and C9 are projected to shift respectively at speeds of 118 km.dec^{-1} and 195 km.dec^{-1} comparable to 100 km.dec^{-1} and 200 km.dec^{-1} for the subtropical and tropical biomes of Polovina et al.¹⁴.

For all size fractions, climate change would lead to a poleward expansion of tropical and equatorial provinces at the expense of temperate provinces (Supplementary discussion 4, Supplementary Table 2, Extended Data 7, Supplementary Fig. 13). This is illustrated by the temperate province F5 of size fraction 180-2000 (Supplementary Fig. 12), which is projected to shrink in the five major basins. In the North Atlantic, its centroid would move approximately 800 km to the northeast (Supplementary Fig. 12c). Similar trends are found comparing present day and end of the century consensus maps (Supplementary discussion 2, Supplementary Fig. 13).

Supplementary discussion 5 | Expansion and shrinkage of provinces in response to climate change

To quantify patterns of expansion or shrinkage of the provinces, we calculate the surface covered by the *dominant* provinces weighted by probabilities of presence (Supplementary Fig. 14, Supplementary Table 2). In this way, *dominant* provinces are defined on 100% of the surface ocean (327 million km^2) but their presence probabilities correspond to the equivalent of 45 to 74% (due to sampling variability and niche overlaps) of the surface ocean depending on the plankton size fraction (Table 1). Overall, our results indicate expansions of the surface of tropical and tropico-equatorial provinces but in very different ways depending on the size fractions of organisms. The surface area of temperate provinces is $\sim 22 \text{ million km}^2$ on average (from 10 Mkm^2 for $0-0.2 \mu\text{m}$ to 49 Mkm^2 for $20-180 \mu\text{m}$) and should decrease by 36% on average (from -20 % for $5-20 \mu\text{m}$ up to -54% for $0.8-5 \mu\text{m}$, -12 million km^2 on average, +6% for $0.22-3 \mu\text{m}$). Tropical provinces cover $\sim 118 \text{ million km}^2$ on average (from 86 Mkm^2 for $0.8-5 \mu\text{m}$ up to 169 Mkm^2 for $180-2000 \mu\text{m}$) and their coverage should increase by 32% on

average (from +13% for 0-0.2 μm up to +75% for 0.8-5 μm , +25 million km^2 on average) (Supplementary Fig. 14 and Supplementary Table 2).

Supplementary discussion 6 | Changes in the distribution of grazers and phototrophs

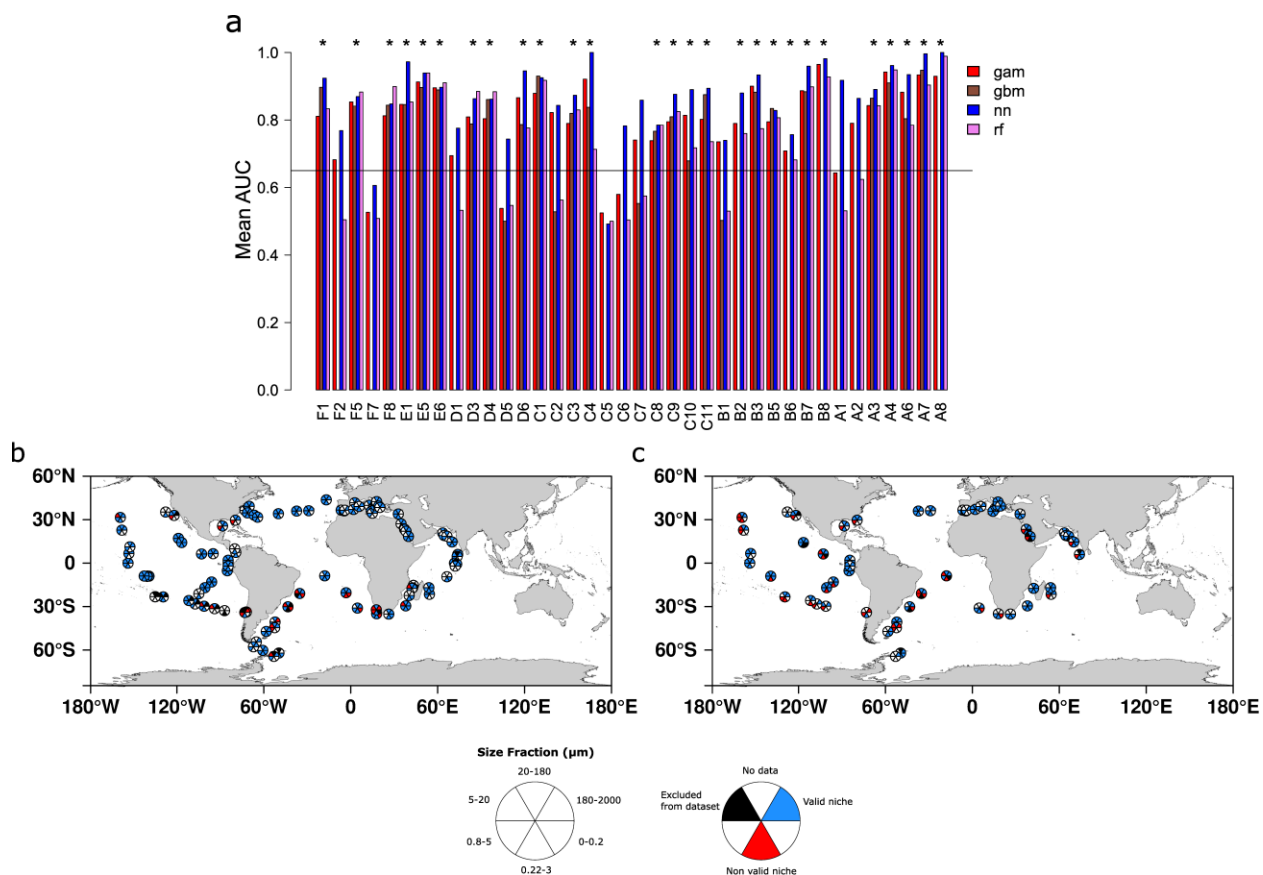
Copepods are cosmopolitan small crustaceans. These abundant grazers contribute to the biological pump and their body size is considered to be a key trait for carbon export¹⁵. They feed on smaller plankton¹⁵ and are prey to higher trophic levels. We characterize the composition of provinces using 198 environmental genomes detected in at least 5 samples and annotated as Marine Hexanauplia of clade A and B. They are further divided into five subgroups based on their differential abundances across the large size classes ($>5 \mu\text{m}$): 8 mesozooplankton clade A, 19 unclassified clade A, 71 microzooplankton clade A, 10 unclassified clade B and 90 microzooplankton clade B. Therefore, large copepods are preferentially from clade A in our data.

The relative abundances of unclassified clade B in size class 20-180 μm (Fig. 3e) and of mesozooplankton clade A in size class 180-2000 μm , (Extended Data 6a) are projected to increase in regions where the temperate province is replaced by the tropico-equatorial province. In areas where the polar province is replaced by the temperate province, a greater relative abundance in microzooplankton clade A (1% to 10%) (Fig. 3e) is projected. No significant compositional differences in the provinces of size fraction 5-20 μm are found (Supplementary Fig. 6b). These results highlight potential significant compositional shifts by the end of the century in the different clades and sizes of main grazers.

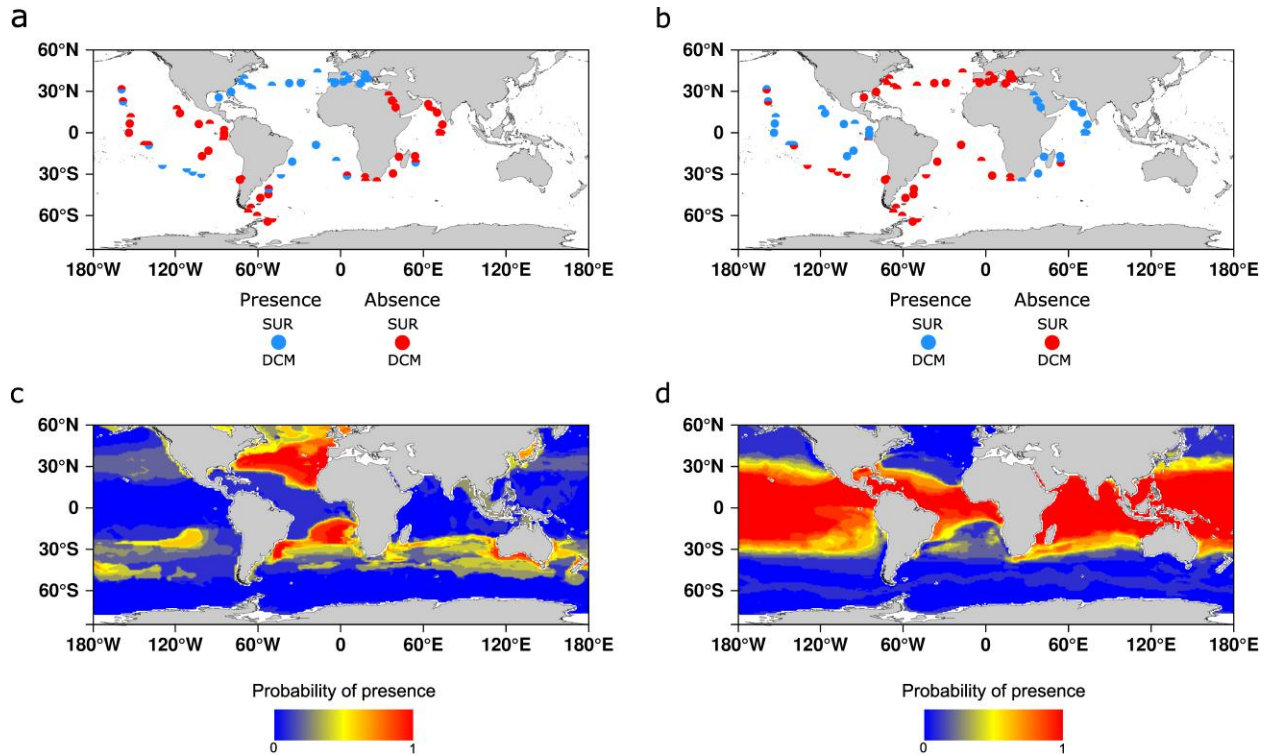
Phototrophs are the primary producer of the ocean that fix carbon from the atmosphere via photosynthesis and therefore sustain the whole trophic chain and carbon export to the deep oceans¹⁶⁻¹⁸. Their size is a key trait for carbon export: picophytoplankton (between 0.2 and 2 μm) and nanophytoplankton (2-20 μm) are believed to contribute relatively little to carbon export¹⁹, although the question is under debate²⁰, while larger cells of microphytoplankton ($>20 \mu\text{m}$) especially diatoms are large contributors²¹. We characterize our genomic provinces by their content in 231 phototrophs annotated by

their preferential size (nano to microphytoplankton) and taxonomy (diatoms, cyanobacteria and other algae (referred to as 'algae'))).

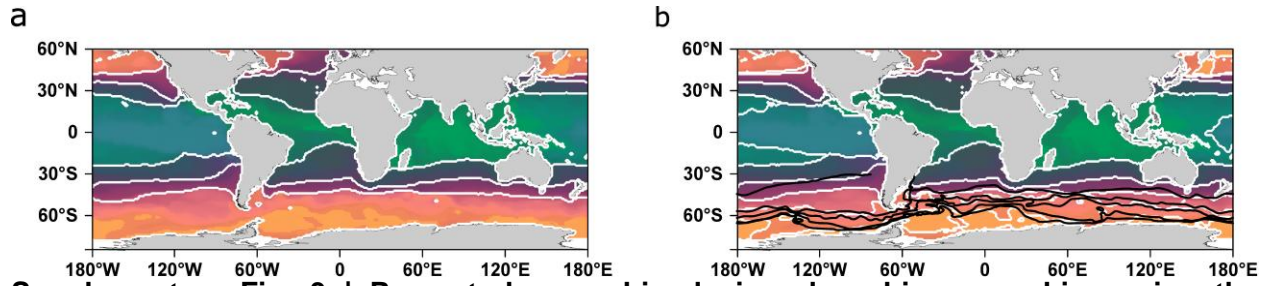
In regions of temperate to tropico-equatorial transition, nano-algae are projected to increase in size fraction 180-2000 μm (Extended Fig. 8a) while micro-algae, nano-diatoms and cyanobacteria (both nano- and micro-) are projected to increase in size fraction 20-180 μm (Extended Fig. 8b). In size fraction 5-20 μm , nano-diatoms, algae and cyanobacteria are all projected to increase in temperate to equatorial transitions while nano-diatoms and algae and pico-cyanobacteria decrease in equatorial to tropical transitions (Extended Fig. 8c). In size fraction 0.8-5 μm , nano-algae are decreasing in equatorial to tropical transitions. Nano-diatoms and cyanobacteria are decreasing in temperate to subtropical transitions (towards C4). Finally in size fraction 0.22-3 μm , in temperate to tropical transitions, pico-cyanobacteria and algae are projected to decrease. Overall, changes in phototroph composition are complex and our dataset is not sufficient to describe the entire diversity of phototrophs (especially large diatoms and dinoflagellates that can have bigger genomes).



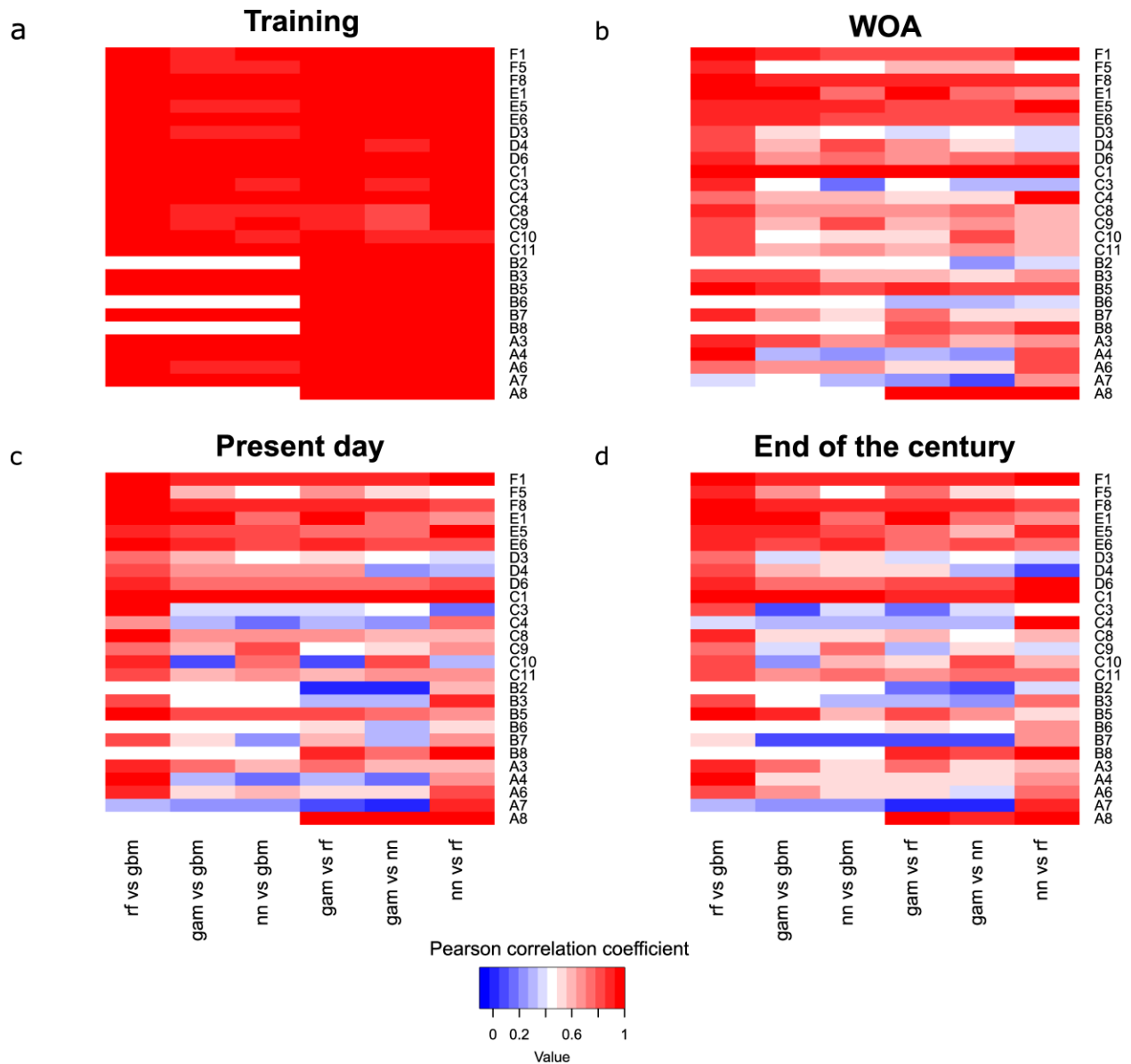
Supplementary Fig. 1 | Barplot of mean AUC over a random 30-fold cross validation process of the 38 initial metagenomic provinces and map of validated and non validated stations across the six size fractions in surface samples and DCM samples. (a) Mean AUC (Area Under the receiver operating Characteristic)²² is plotted for the best hyperparameter combination of the 4 machine learning techniques used for each of the 38 metagenomic provinces. A star for valid models is drawn at the top of each considered valid model. **(b-c)** For each Tara sample present in the dataset at surface **(b)** or Deep Chlorophyll Maximum (DCM) **(c)** and for each size fraction, the filter (one size fraction at one location and one depth) belongs either to a validated niche (green), a non validated niche (red), an excluded filter (see *Materials & Methods*) (black) or no data is available (white). Non validated niches represent only 11% of filters present in the dataset (66 out of 595). We generated these map using R-package *maps*³⁷.



Supplementary Fig. 2 | Sampling and projection maps on WOA13 climatological data of two example provinces from size fraction 180-2000 μm . (a) Sampling map of province F5. (b) Province F8. (c) Projection map of province F5 on WOA13²³. (d) Province F8. Qualitatively, projection maps are coherent with sampling maps of the two provinces with the highest probability of presence projected in the sampling regions. Other presence zones are also predicted by the projection. Sampling of these zones might be interesting to confirm our approach and projections such as South of Australia where a high probability of presence for province F5 is predicted. We generated these map using R-package *maps*³⁷.



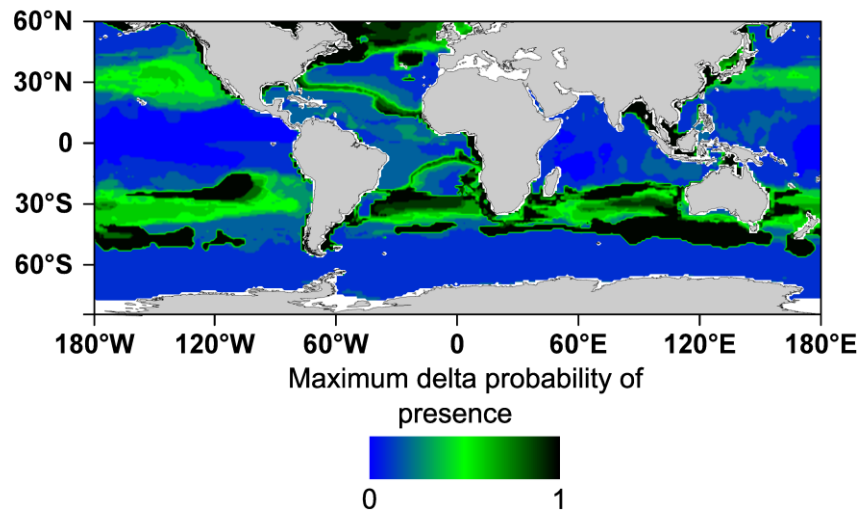
Supplementary Fig. 3 | Present day combined size class biogeographies using the PHATE algorithm on WOA13 dataset. (a) Combined size class biogeography using PHATE algorithm⁷ partitioned in 4 clusters. Each grid point is associated to a color depending on its PHATE coordinates (using three axes). Each coordinate is respectively assigned to a given degree of color between 0 and 255 according to equation (3) (either red green or blue depending on the axis).(b) Combined size class biogeography using PHATE algorithm partitioned in 7 clusters overlaid with Antarctic Circumpolar Current boundaries⁸ (black). Colors are the same as in (a). We generated these map using R-package *maps*³⁷.



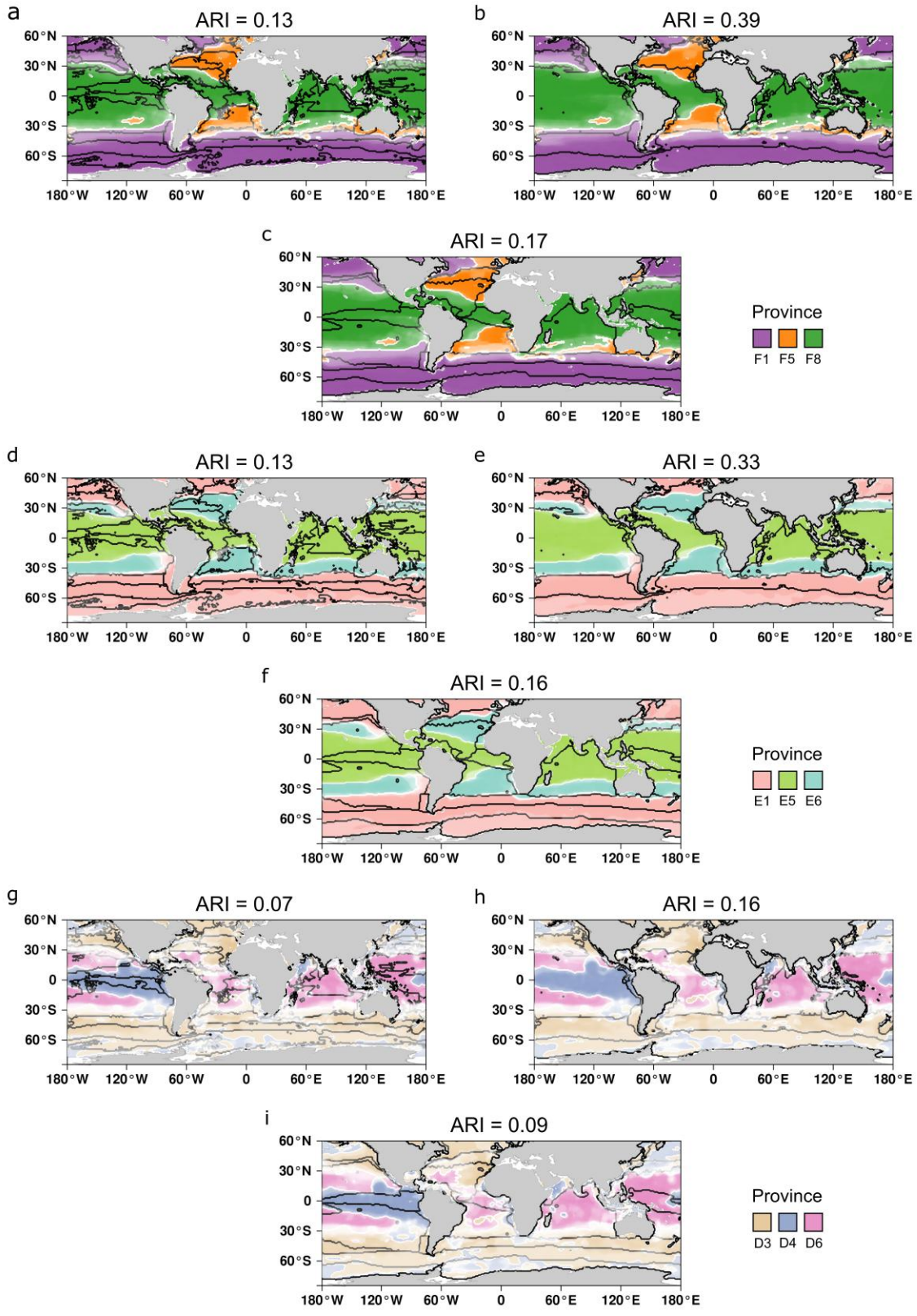
Supplementary Fig. 4 | Pairwise Pearson correlation coefficient heat maps between outputs of the 4 machine learning models. rf: Random Forest, gbm: gradient boosting machines, gam: general additive models, nn: neural networks. The rows are the provinces of the different size fractions. The columns are the pairwise comparisons of each machine learning technique. **(a)** Training set outputs. **(b)** WOA13 average data projection outputs (except for Iron, PISCES-v2²⁴ is used). **(c)** Present day data projection outputs (bias corrected mean model of 6 Earth System Models). **(d)** End of the century projection outputs. On the training set, models are in agreement with most of the correlation coefficients superior to 0.9. A drop in correlation is observed for modeled data (**c**, **d**) especially in small size fractions. This shows modeled data are more distant from the training set than WOA data. Random Forest and Gradient Boosting Machine are in very good agreement (first columns) which could be expected

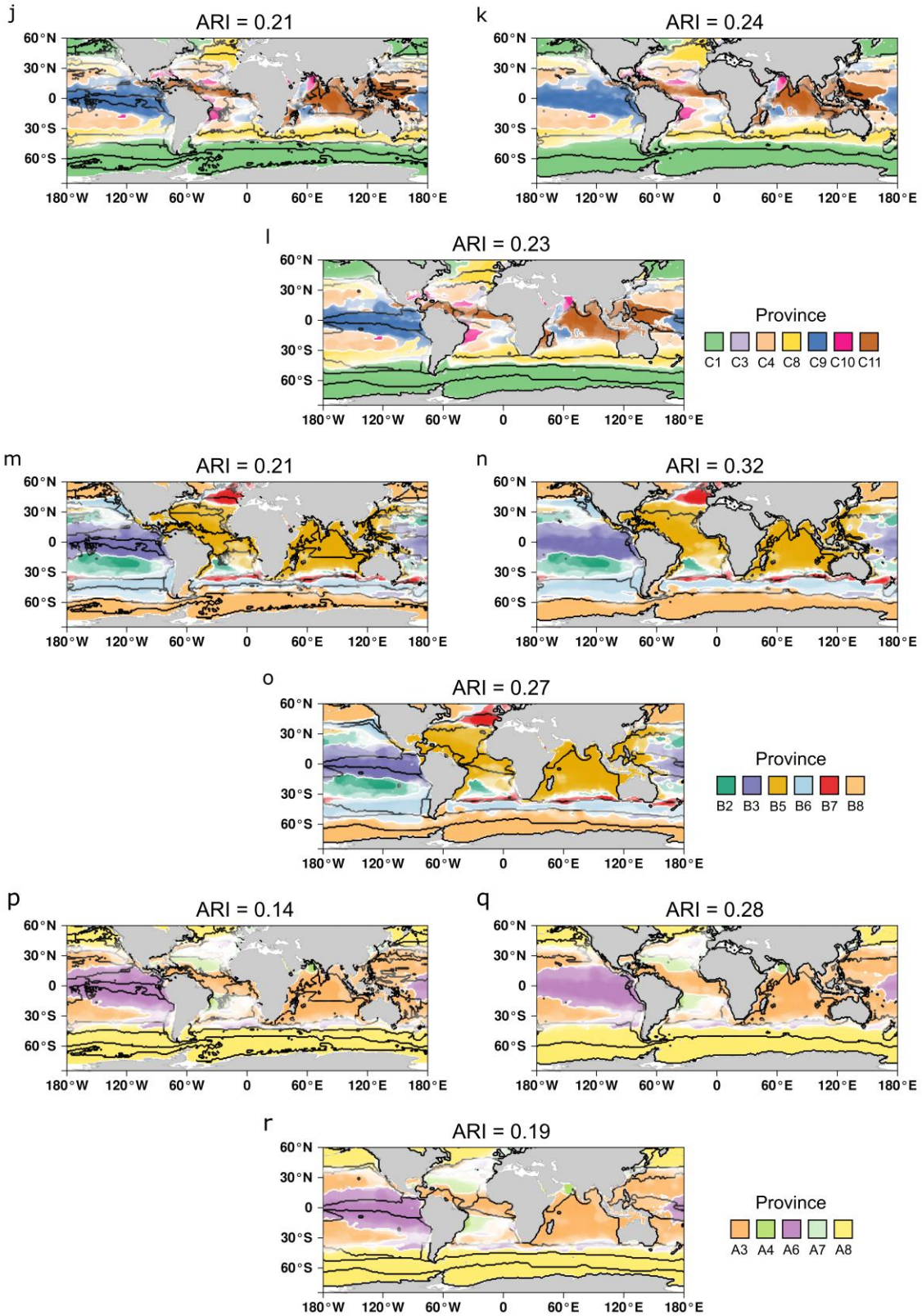
as they are
on multiple
trees.

both based
decision



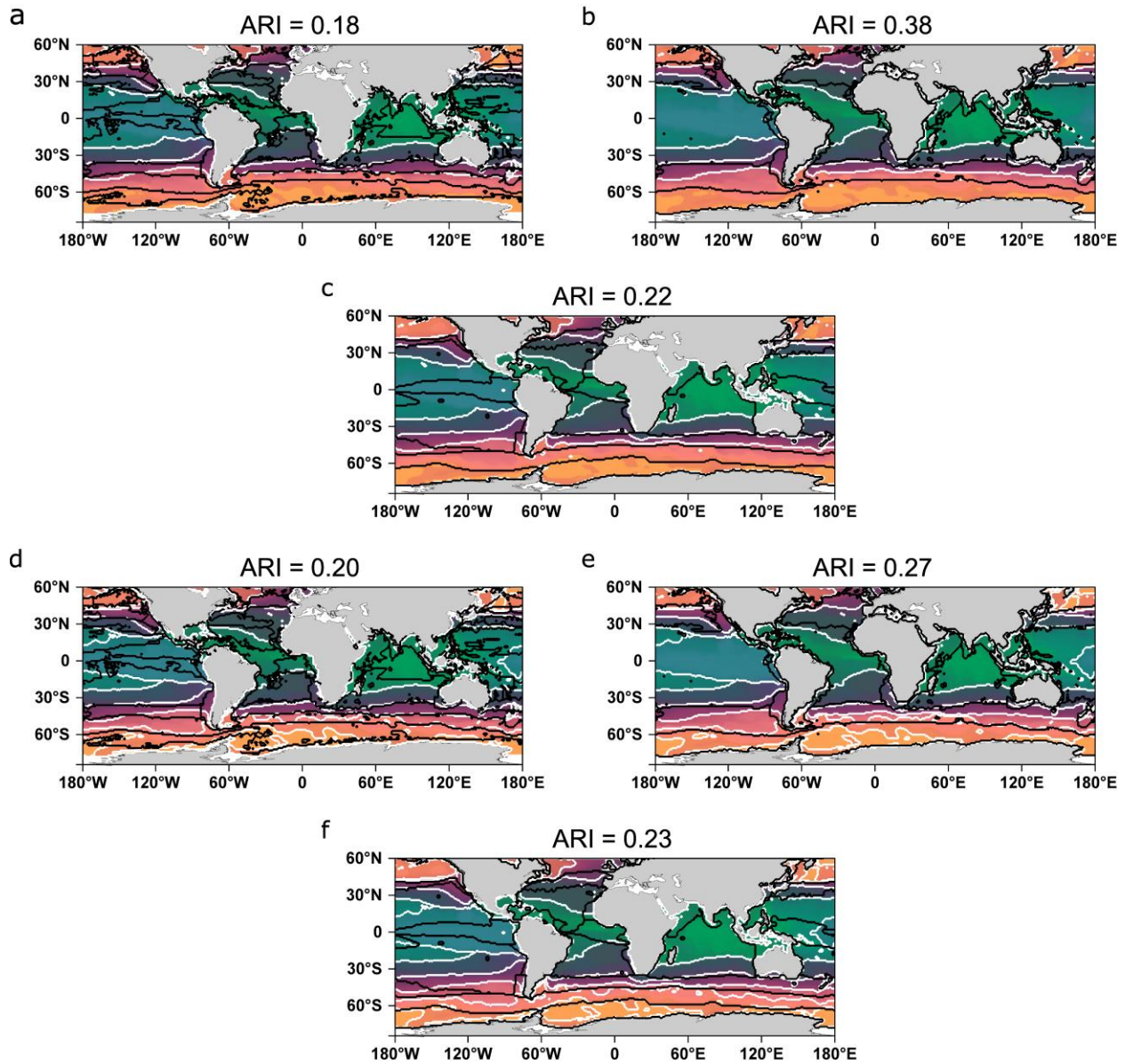
Supplementary Fig. 5 | Probability range map (WOA data) of province F5 of size fraction 180-2000 μm . At each point of the grid, the maximum delta probability of presence between the 4 machine learning projections is calculated. In black are the zones where two models completely disagree: one model predicts presence with certainty whereas the other predicts absence with certainty. Disagreement appears mostly in uncertain presence areas (Supplementary Fig. 2c) whereas models are generally in good agreement in absence areas (blue zones). We generated these map using R-package *maps*³⁷.



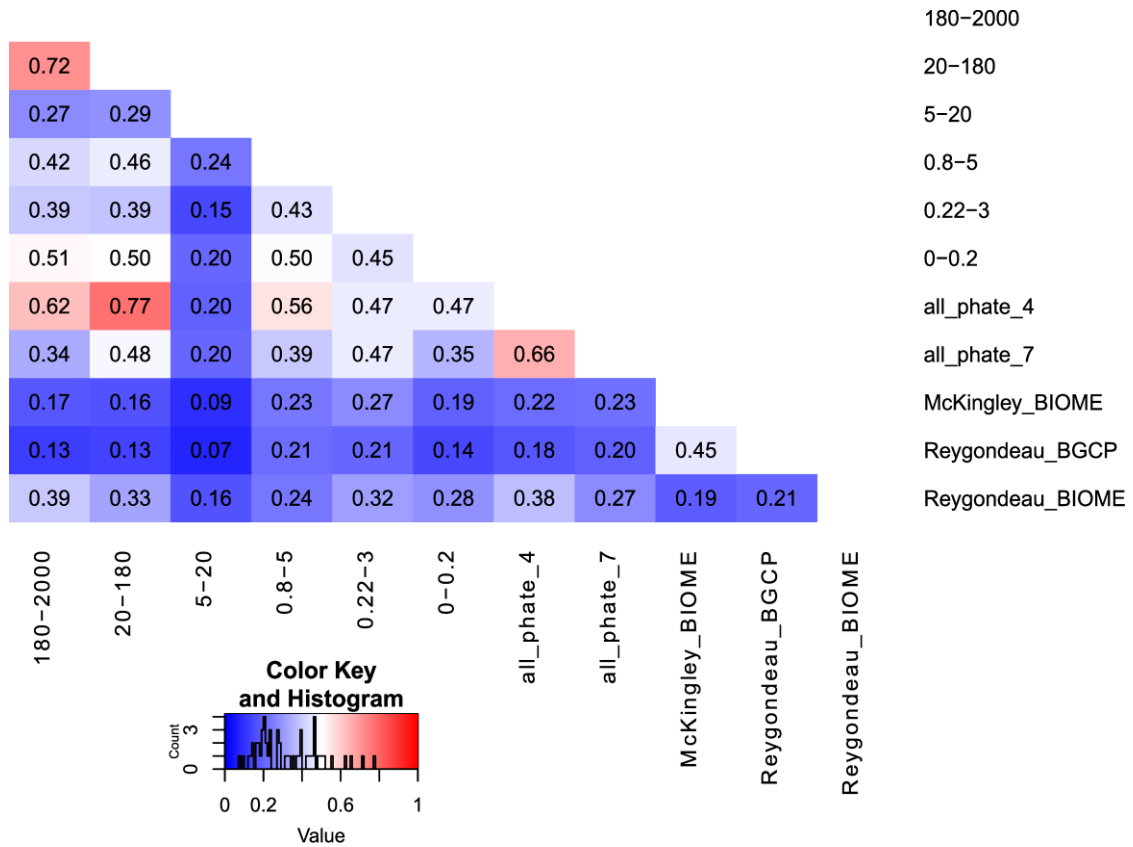


Supplementary Fig. 6 | Three existing oceanic partitions overlaid on top of plankton provinces for the six fractions and combined size fractions. The three oceanic partitions

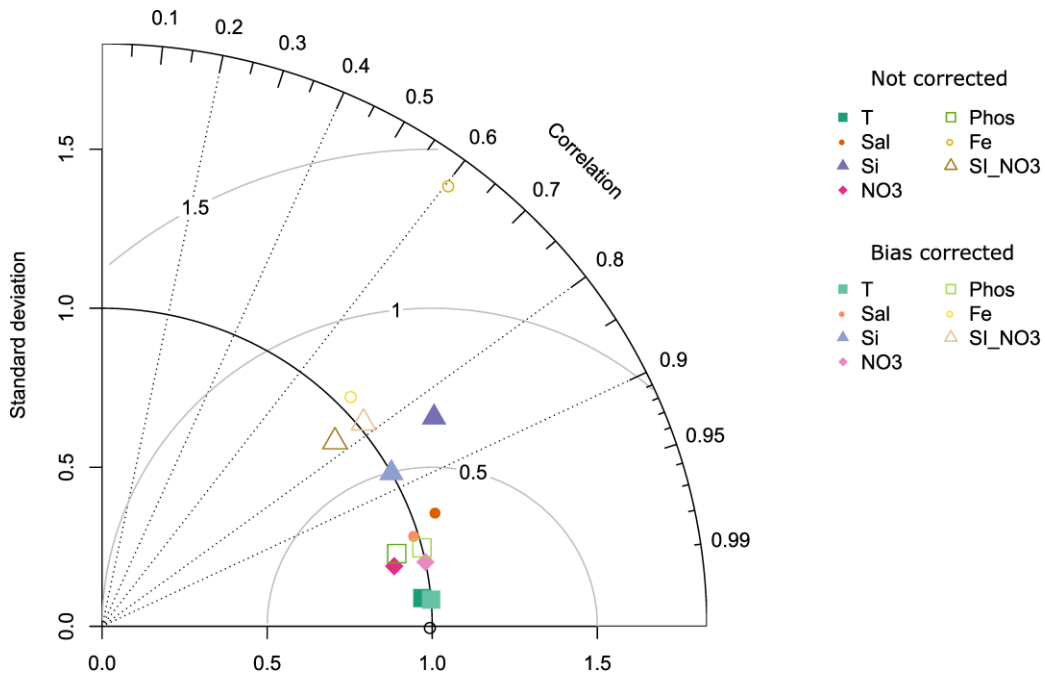
are Reygondeau et al. Biogeochemical Provinces (BGCP)¹⁰; Reygondeau et al. Biomes¹⁰; Fay and McKingley Biomes¹¹. Frontiers of each partitioning are overlaid in the above order on top of plankton provinces for the six size fractions. **(a-c)** 180-2000 μm **(d-f)** 20-180 μm **(g-i)** 5-20 μm **(j-l)** 0.8-5 μm **(m-o)** 0.22-3 μm and **(p-r)** 0-0.2 μm . Above all maps, the adjusted rand index (ARI, an index used to compare partitions), for the comparison of mapped biogeographies with the black line mask is shown. We generated these map using R-package *maps*³⁷.



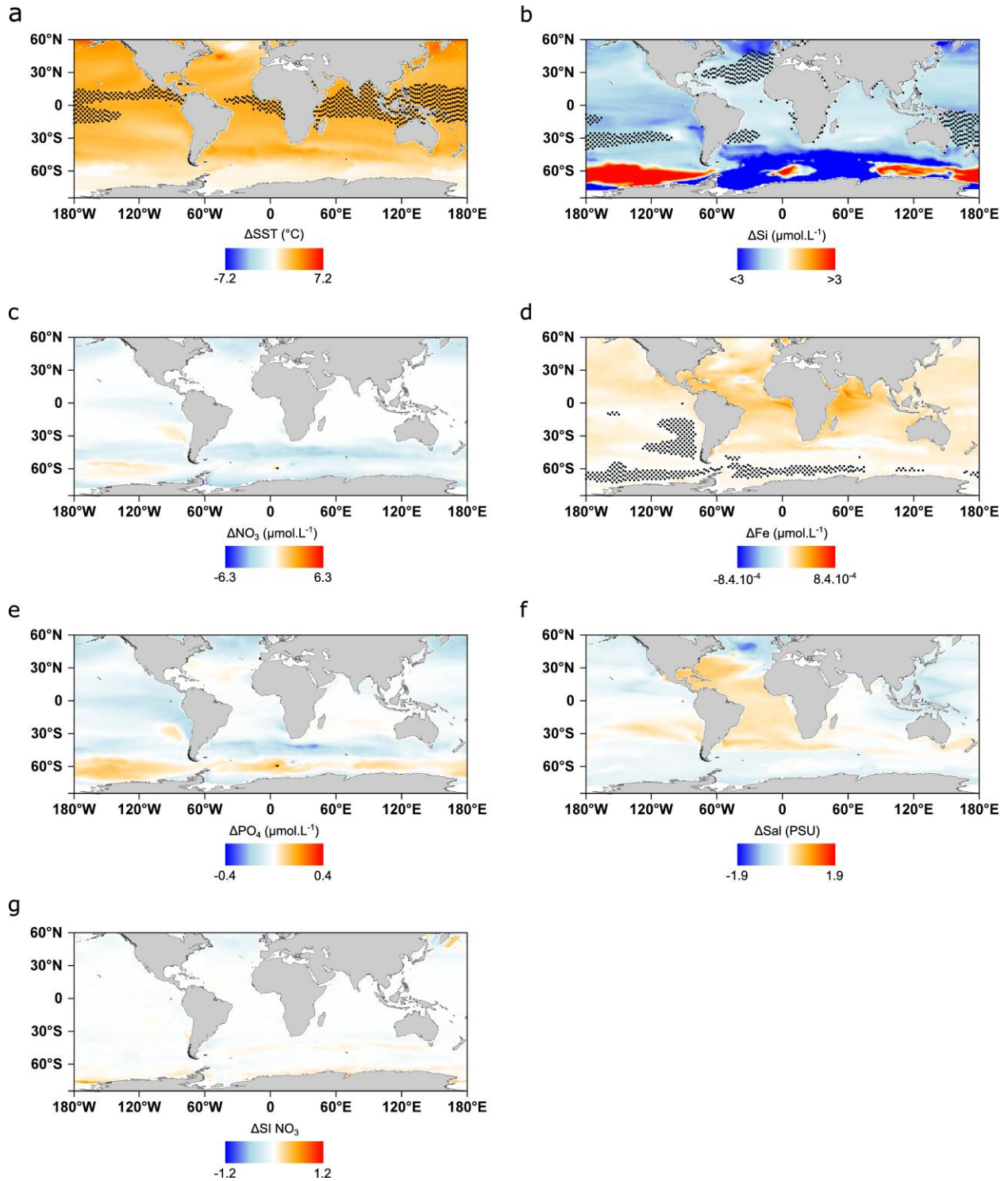
Supplementary Fig. 7 | Three existing oceanic partitions overlaid on top of plankton provinces for the combined size fractions. The three oceanic partitions are Reygondeau et al.¹⁰ Biogeochemical Provinces (BGCP); Reygondeau et al. Biomes¹⁰; Fay and McKingley Biomes¹¹. Frontiers of each partitioning are overlaid in the above order on top of size fraction plankton provinces combined by the PHATE algorithm in **(a-c)** 4 clusters and **(d-f)** 7 clusters. Each grid point is associated with a color depending on its PHATE coordinates (using three axes). Each coordinate is respectively assigned to a given degree of color between 0 and 255 according to equation (3) (either red green or blue depending on the axis). We generated these map using R-package *maps*³⁷.



Supplementary Fig. 8 | Pairwise comparisons of ocean partitions based on plankton provinces and existing biogeochemical partitions. The three oceanic partitions are Reygondeau et al.¹⁰ Biogeochemical Provinces (BGCP); Reygondeau et al. Biomes¹⁰; Fay and McKingley Biomes¹¹.

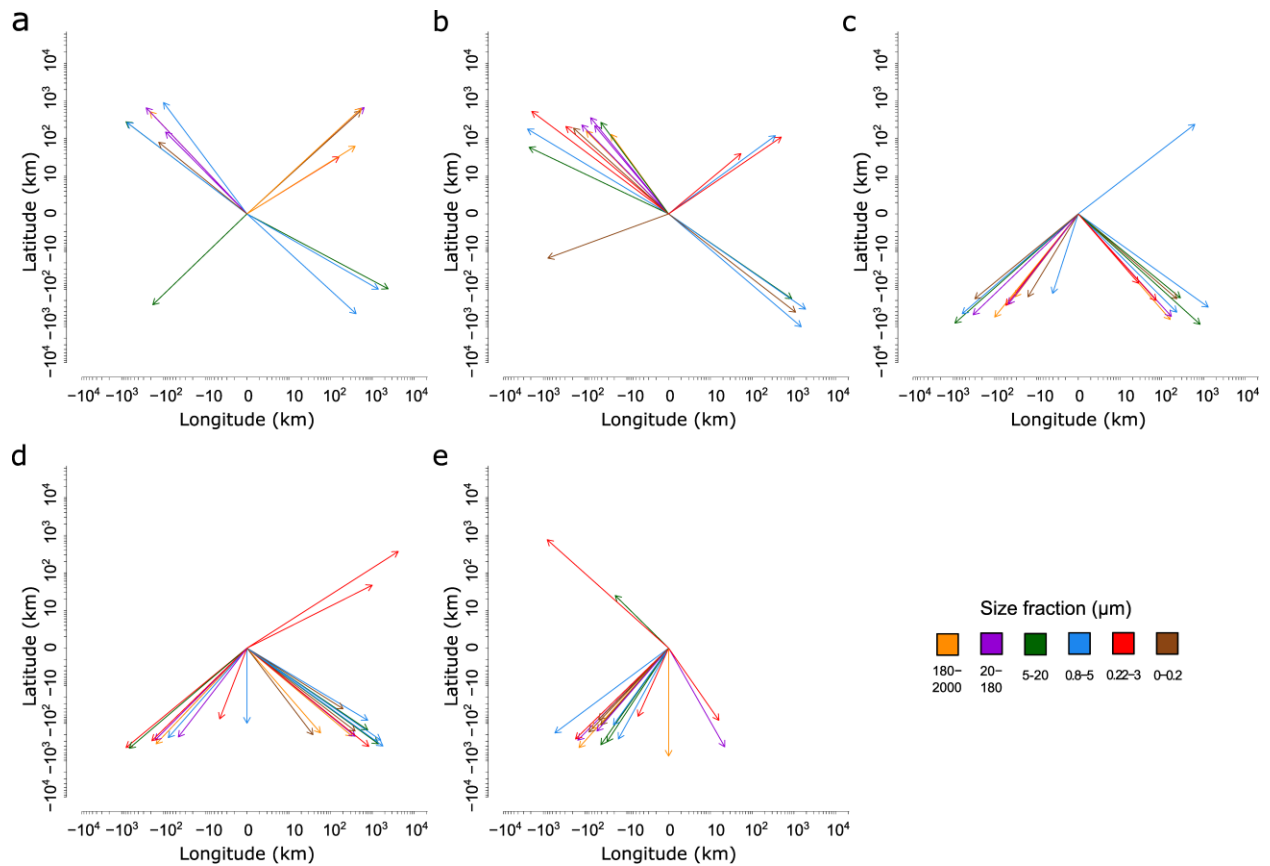


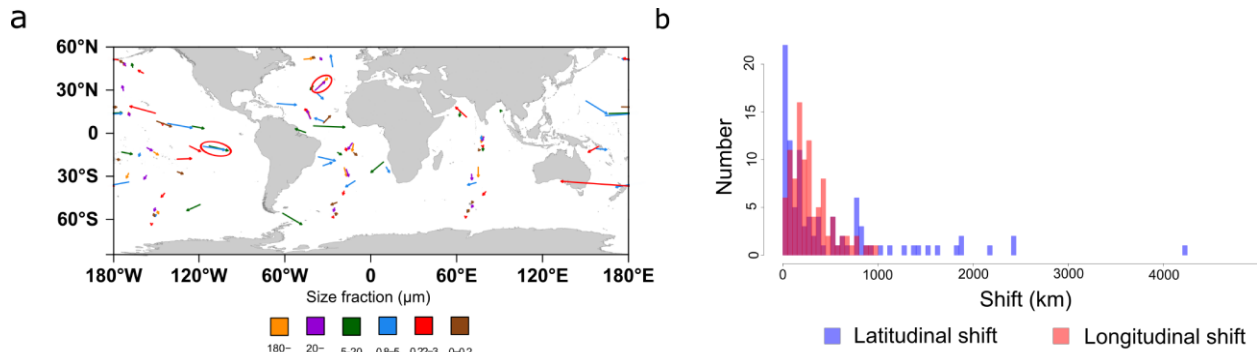
Supplementary Fig. 9 | Taylor diagram exhibiting statistical comparison of WOA 2013 observations and present day ESM model mean of the different drivers. Each variable is centered and scaled according to the mean and standard deviation of the observed variable (black circle point at standard deviation 1 on the x-axis). Dark color points are the ESM model mean without Cumulative Distribution Function transform (CDFt²⁵) bias correction. They are to be compared with light color points for which the bias correction is performed. Overall, good spatial correlations are found between the model mean and the observations. CDFt bias correction performs well by bringing standard deviations of the model to the observed standard deviations without decreasing correlations.



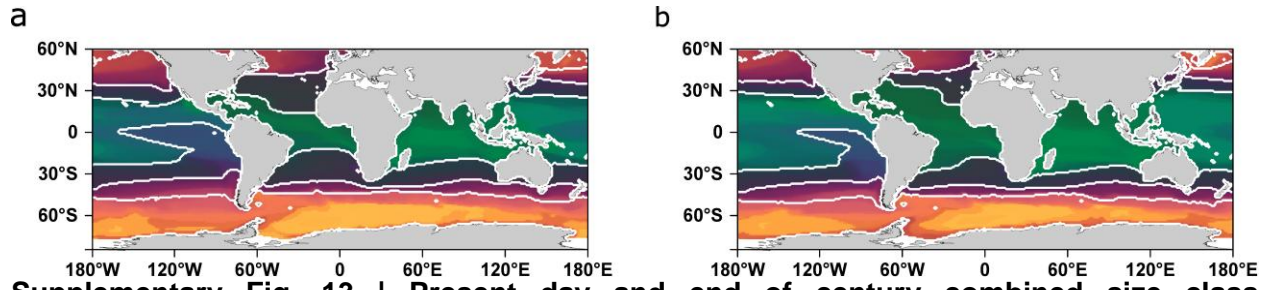
Supplementary Fig. 10 | Differences in drivers' intensity (2090/99-2006/15) in the bias corrected ESM model-mean under RCP8.5. (a) Sea Surface Temperature (SST). (b) Dissolved silica. (c) Nitrate. (d) Iron. (e) Phosphate. (f) Salinity (Sal). (g) Seasonality Index of Nitrate (SI NO₃). Note that the scale for dissolved silica variations is restricted to visualize small variations. Regions for which out of range values (*i.e.* inferior to the minimum or superior to the

maximum found in 2006-15) are reached at the end of the century are highlighted with small stars reflecting high uncertainty zones for machine learning approaches. We generated these map using R-package *maps*³⁷.



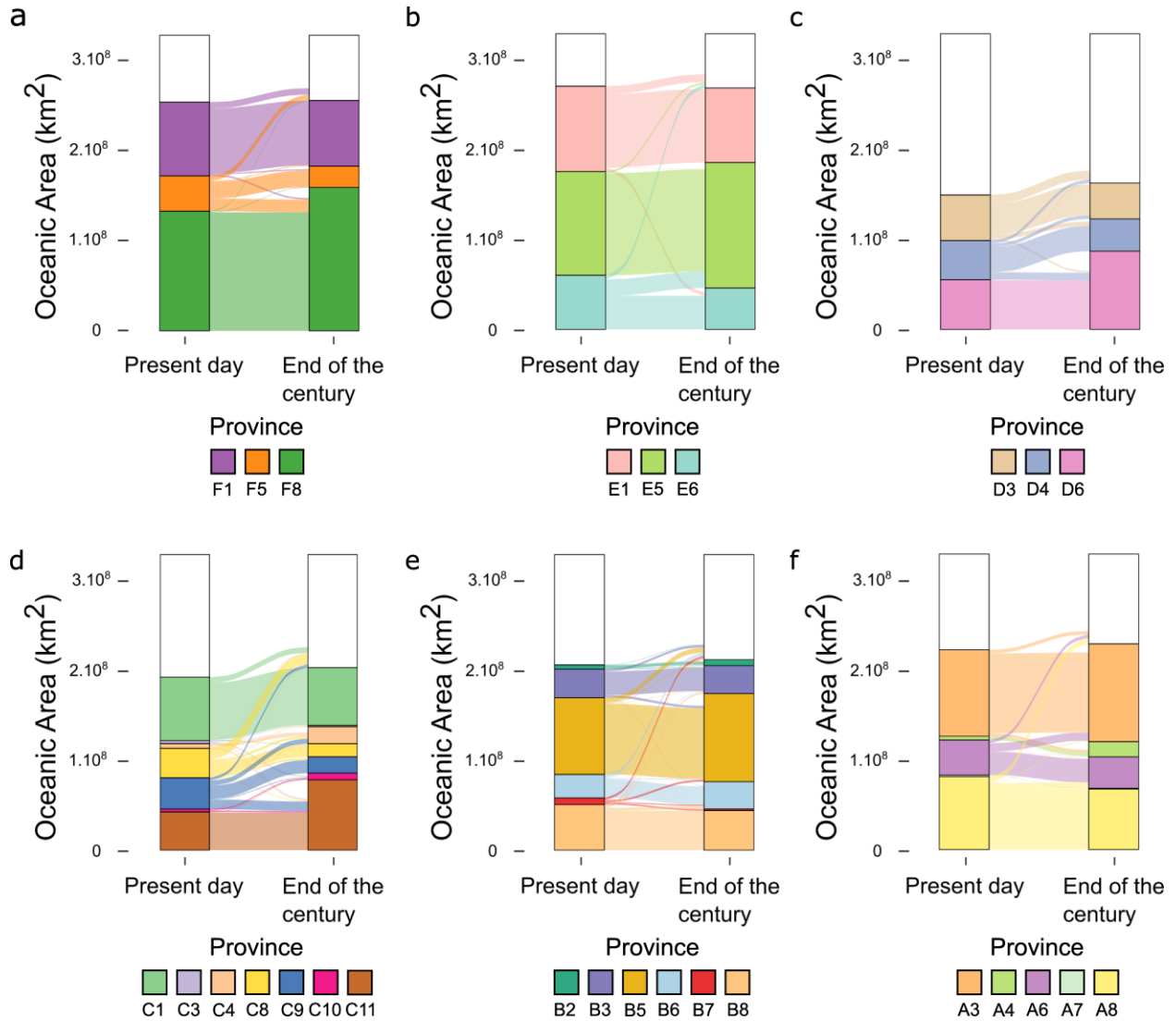


Supplementary Fig. 12 | Projected migration shifts of the provinces on the world map and their latitudinal and longitudinal shift distributions. (a) Migration shifts are represented as arrows pointing at the end of century centroid. Arrows are colored according to the size fraction. Some shifts correlate with each other (exemplified with circled arrows). For instance, parallel shifts are projected in the southern pacific equatorial communities of size fractions 0.8-5 μm and 5-20 μm (blue and green circled arrows). All non-poleward arrows belong to small size classes (<20 μm) showing differential responses to climate change depending on the size class. (b) Some longitudinal shifts are more important than latitudinal shifts with 14 longitudinal shifts superior to 1000 kms. Mean longitudinal shift (around 500 kms) is significantly higher (Student t-test²⁶ p-value<0.01) than mean latitudinal shift (around 290 kms) while medians (longitudinal 190 kms vs latitudinal 230 kms) are not significantly different (Wilcoxon test). The median migration speed is 47 $\text{km}\cdot\text{dec}^{-1}$, (latitudinally 23 $\text{km}\cdot\text{dec}^{-1}$, longitudinally 27 $\text{km}\cdot\text{dec}^{-1}$). We generated this map using R-package *maps*³⁷.

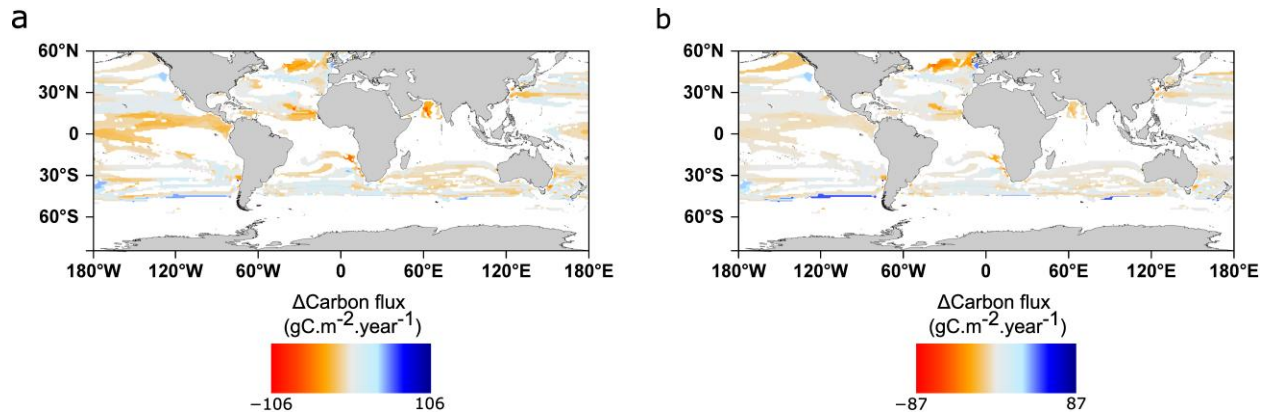


Supplementary Fig. 13 | Present day and end of century combined size class

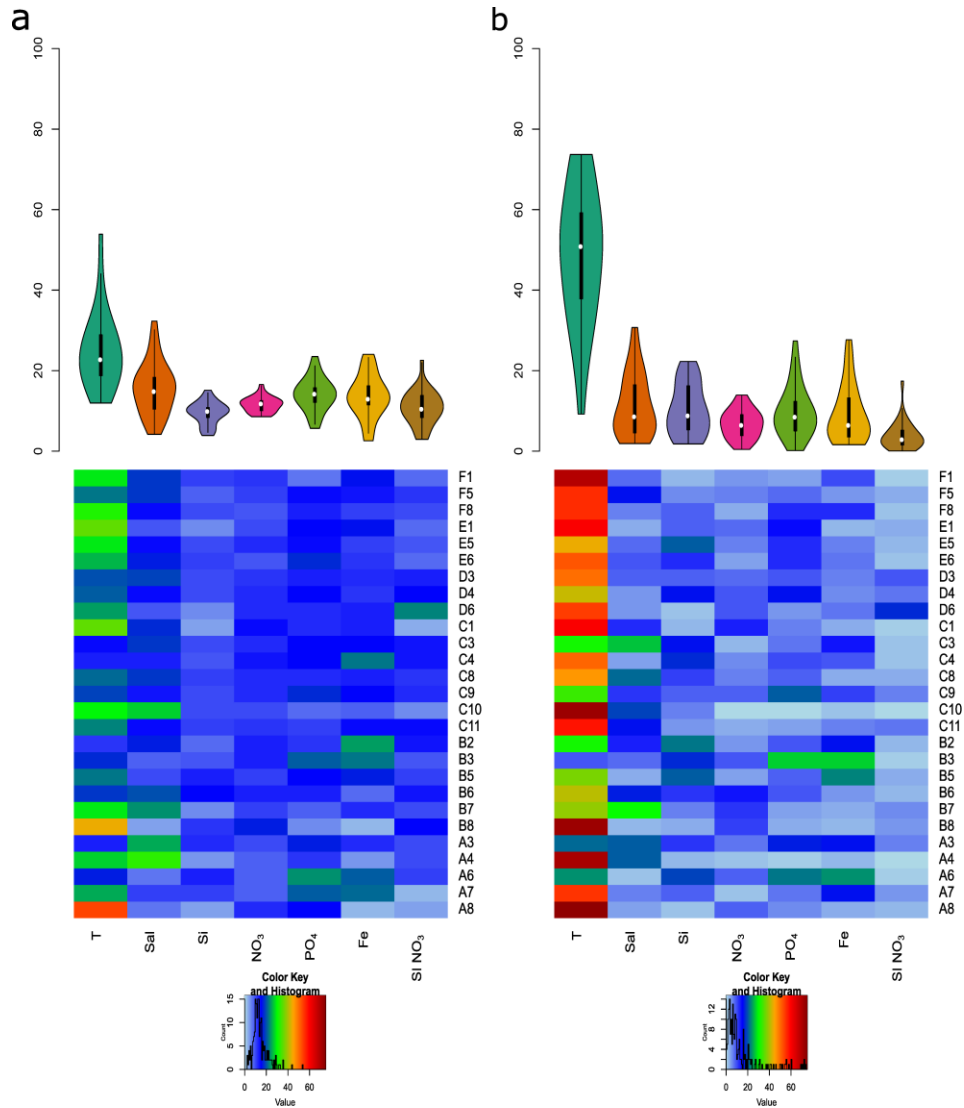
biogeographies using the PHATE algorithm⁷. (a) Present day (b) End of the century. Each grid point is associated with a color depending on its PHATE coordinates (using three axes). Each coordinate is respectively assigned to a given degree of color between 0 and 255 according to equation (3) (either red green or blue depending on the axis). Boundaries of clusters using k-medoids²⁷ (k=4) are represented in white. We generated these map using R-package *maps*³⁷.



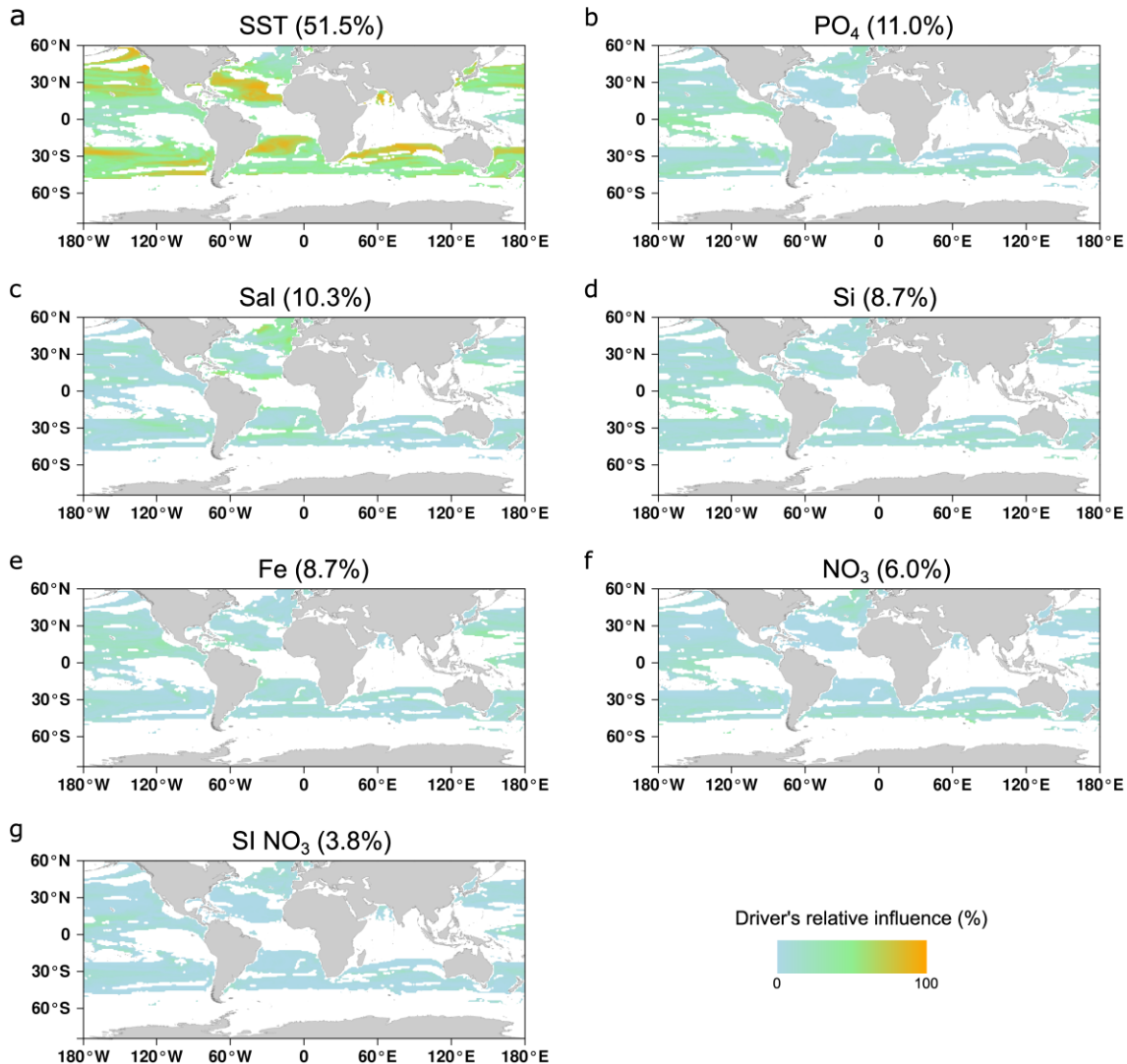
Supplementary Fig. 14 | Probabilistic covered areas of the provinces projected in present day and at the end of the century. (a) 180-2000 μm (b) 20-180 μm (c) 5-20 μm (d) 0.8-5 μm (e) 0.22-3 μm (f) 0-0.2 μm . The area covered by a province is defined as the area in which this province is dominant and weighted by its probability of presence at each point and grid cell area. Areas not covered by the provinces are represented in white.



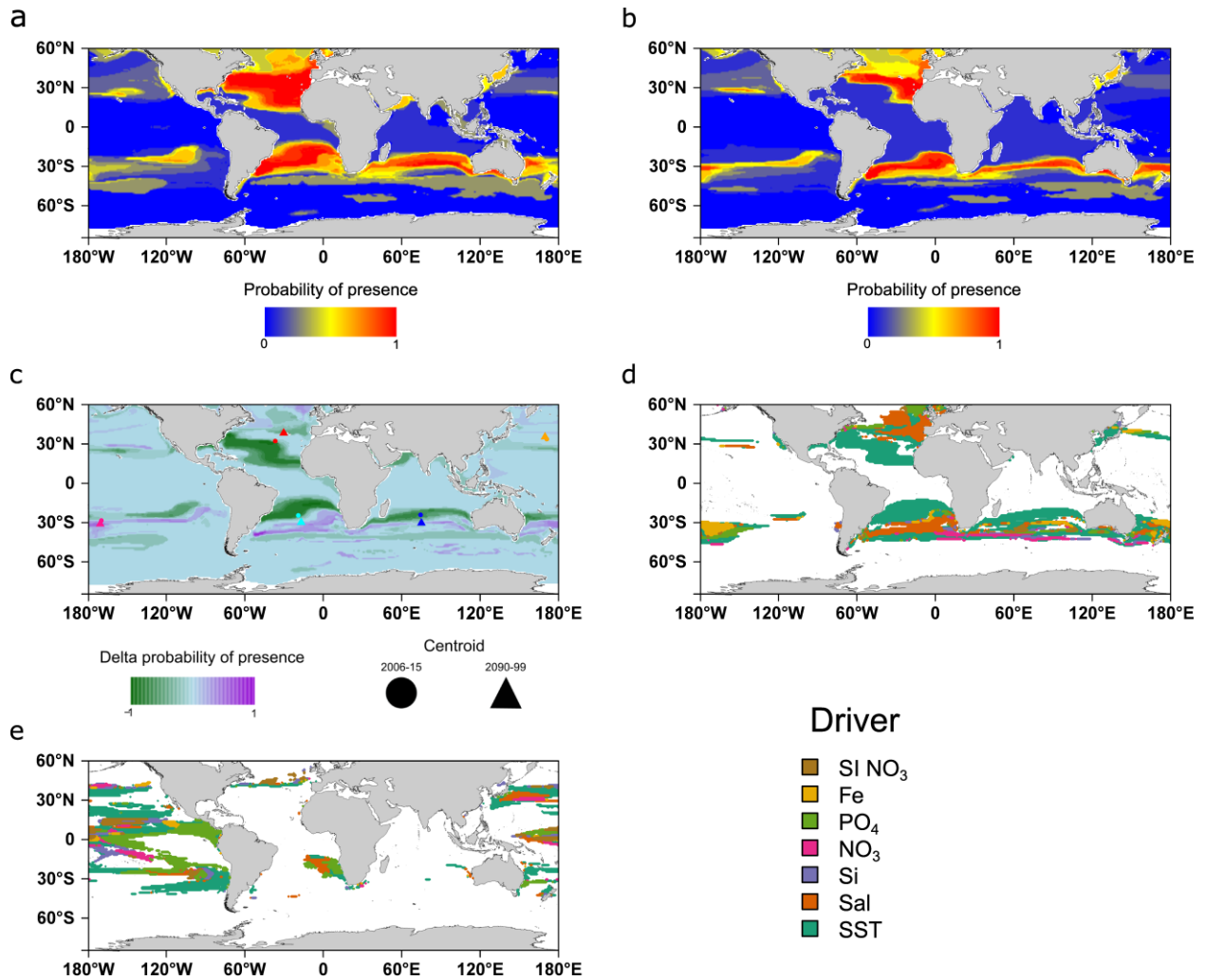
Supplementary Fig. 15 | Differences in carbon export by the end of the century using data from two present day extrapolative models of carbon export. Each assemblage of *climato-genomic* province is associated to its mean export flux from present day extrapolative models. Then the difference between present and future export is calculated by the difference between export fluxes from future and present assemblages at a given point. Means for assemblages are calculated based on **(a)** Eppley et al.²⁸ **(b)** Laws et al.²⁹. We generated these map using R-package *maps*³⁷.



Supplementary Fig. 16 | Distributions and heat maps of the relative influences of the different drivers in defining single niches associated with the provinces from DALEX R package³⁰ and in driving climate change associated reorganization of single provinces. Median relative influence of Temperature is significantly higher than for all other environmental parameters (Pairwise Wilcoxon test $p < 0.01$ for all parameters) in (a) defining the niches and (b) driving province reorganization. This is also the case within individual size fractions for (b) but not for (a). Respectively, Salinity (Sal) and Phosphate (PO_4), have second and third highest median relative influences far behind Temperature whereas dissolved Silica and seasonality index of Nitrate (SI NO_3) have the lowest median relative importance in (a). Numbers above violins are median (and mean) relative influences.



Supplementary Fig. 17 | Maps of environmental drivers' relative influences in driving province reorganization. Each environmental parameter's relative influence is quantified by considering only the variation of each parameter individually between present day and end of the century as defined in Barton *et al.*¹²(*Materials and Methods*) and where a significant change is projected. Importantly the mean impact of SST (51.5%) is to a great extent the highest. PO₄ is the second most impacting driver (11.0%). Contrary to Supplementary Fig. 18, relative influence is calculated here by combining all provinces together. Therefore, mean relative influences slightly differ especially for dissolved silica (Si). We generated these map using R-package *maps*³⁷.



Supplementary Fig. 18 | Projection maps of province F5 of size fraction 180-2000 μm in present day, at the end of the century and their difference and drivers of changes for province F5 and C9. At each grid point, the probability of presence of the province is computed as the average of the predicted probability of each of the four machine learning techniques (gbm, nn, rf and gam). Red color indicates a high probability of presence. **(a)** The projected province corresponds to the sampled province (North Atlantic and South Atlantic) but several other places have high probabilities of presence such as South Australia where no sampling is available. **(b)** At the end of the century, the province is projected to reduce significantly in size. **(c)** Delta probability of presence map (2006/15 – 2090/99) and core range shift in the 5 major oceanic basins of province F5 of size fraction 180-2000 μm . In all the basins, the centroid of the province is projected to be displaced poleward. **(d)** Main drivers associated with the projected changes. Changes are mainly driven by sea surface temperature (56%) followed by salinity (16%). **(e)** Main drivers associated with the shrinkage of the equatorial cluster C9 of size fraction 0.8-5. Considering only latitudes between the two tropics, changes are mainly driven by

decreases in PO_4 (24%) in addition to SST (27%) (overall 34% STT and 20% PO_4). We generated these map using R-package *maps*³⁷.

Model	Reference
CESM1-BGC	Gent et al., 2011 ³¹
GFDL-ESM2G	Dunne et al., 2013 ³²
GFDL-ESM2M	Dunne et al., 2013 ³²
HadGEM2-ES	Collins et al., 2011 ³³
IPSL-CM5A-LR	Dufresne et al., 2013 ³⁴
IPSL-CM5A-MR	Dufresne et al., 2013 ³⁴
MPI-ESM-LR	Giorgetta et al., 2013 ³⁵
MPI-ESM-MR	Giorgetta et al., 2013 ³⁵
NorESM1-ME	Bentsen et al., 2013 ³⁶

Supplementary Table 1 | Earth System models used to compute the mean model.

Fraction	Province	Climatic annotation	Area 2006-15 (MKm ²)	Area 2090-99 (MKm ²)	Delta area (MKm ² (%))
180-2000	F1	polar	82	73	-9 (-11%)
180-2000	F5	temperate	44	26	-18 (-41%)
180-2000	F8	tropico-equatorial	140	169	+29 (+21%)
20-180	E1	polar	97	84	-13 (-13%)
20-180	E5	tropico-equatorial	119	145	+26 (+22%)
20-180	E6	temperate	65	49	-15 (-23%)
5-20	D3	temperate	51	41	-10 (-20%)
5-20	D4	equatorial	46	37	-8 (-18%)
5-20	D6	tropical	65	97	+32 (+48%)
0.8-5	C1	polar	71	64	-6 (-8%)
0.8-5	C3	subtropical	3,6	1,4	-2,2 (-61%)
0.8-5	C4	subtropical	6	19	+13 (+217%)
0.8-5	C8	temperate	33	15	-18 (-54%)
0.8-5	C9	equatorial	35	19	-16 (-45%)
0.8-5	C10	subtropical	3.8	7.4	+3,6 (+95%)
0.8-5	C11	tropical	49	86	+37 (+75%)
0.22-3	B2	subtropical	5.1	6.7	+1,6 (+31%)
0.22-3	B3	equatorial	33	32	-1 (-0.3%)
0.22-3	B5	tropical	88	102	+13 (+15%)
0.22-3	B6	temperate	30	35	+5 (+17%)
0.22-3	B7	temperate	8	2	-6 (-75%)
0.22-3	B8	polar	51	44	-7 (-14%)
0-0.2	A3	tropical	101	114	+13 (+13%)
0-0.2	A4	subtropical	6	19	+13 (+216%)
0-0.2	A6	equatorial	40	36	-4 (-10%)
0-0.2	A7	temperate	1.7	1	-0.7 (-41%)
0-0.2	A8	polar	81	67	-14 (-17%)

Supplementary Table 2 | Genomic provinces climatic annotations and oceanic surfaces covered in present day (2006-15) and at the end of the century (2090-990).

	Fraction (μm)	SST	Sal	Si	NO3	PO4	Fe	SI NO3
Niche definition	180-2000	27.7	16.8	10.1	10.3	11.4	14	9.7
	20-180	29.7	13.3	8.9	10.7	16.3	13.2	7.9
	5-20	21.5	14.2	8.4	12.7	13.7	12.6	17
	0.8-5	22.6	17.3	9.2	12.3	13.4	14.5	10.8
	0.22-3	23.9	13.7	10.5	13	12.5	14.2	12.2
	0-0.2	27.1	16.4	9.3	9.5	17.9	12.1	7.6
	all	25	15.5	9.5	11.6	14.2	13.6	10.8
Climate change	180-2000	75.2	8.4	2.5	2.6	4.1	5.8	1.4
	20-180	71.7	4.2	7.3	3.9	8.7	3.3	1.0
	5-20	57.9	6.1	4.4	6.6	9.0	6.5	1.0
	0.8-5	54.9	13.4	7.2	5.5	9.7	5.9	3.6
	0.22-3	42.6	9.9	13.5	6.4	9.1	15.2	3.3
	0-0.2	42.3	18.9	6.6	6.4	12.1	10.8	2.9
	all	51.5	10.3	8.7	6	11	8.7	3.8

Supplementary Table 3 | Summary table of the relative importance of each environmental driver in niche definition and in driving geographical reorganization in response to climate change. Note that in both cases (niche definition and climate change), the row 'all' is not the mean over the size fractions. In the case of niche definition, this is due to a different number of niches in each size class. In the case of climate change, relative influence is either calculated for single provinces at a given grid point then recalculated for individual size class or calculated for all provinces together (row 'all').

References

1. Ridgeway, G. gbm: Generalized Boosted Regression Models. *R Packag. version 1.6-3.1* (2010).
2. Breiman, L. & Cutler, A. Breiman and Cutler's random forests for classification and regression. *Packag. 'randomForest'* (2012). doi:10.5244/C.22.54
3. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S Fourth edition* by. *World* (2002). doi:10.2307/2685660
4. Wood, S. N. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.* (2004). doi:10.1198/016214504000000980
5. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biometrics Bull.* **1**, (1945).
6. Jones, M. C. & Cheung, W. W. L. Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES J. Mar. Sci.* (2015). doi:10.1093/icesjms/fsu172
7. Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0336-3
8. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep. Res. Part I* (1995). doi:10.1016/0967-0637(95)00021-W
9. Longhurst, A. R. *Ecological Geography of the Sea. Ecological Geography of the Sea* (2007). doi:10.1016/B978-0-12-455521-1.X5000-1
10. Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem. Cycles* (2013). doi:10.1002/gbc.20089
11. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: Mean and temporal variability. *Earth Syst. Sci. Data* (2014). doi:10.5194/essd-6-273-2014
12. Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1519080113
13. Marinov, I. *et al.* North-South asymmetry in the modeled phytoplankton community response to climate change over the 21st century. *Global Biogeochem. Cycles* (2013). doi:10.1002/2013GB004599
14. Polovina, J. J., Dunne, J. P., Woodworth, P. A. & Howell, E. A. Projected expansion of the subtropical biome and contraction of the temperate and equatorial upwelling biomes in the North Pacific under global warming. *ICES J. Mar. Sci.* (2011). doi:10.1093/icesjms/fsq198
15. Richardson, A. J. In hot water: zooplankton and climate change. *ICES J. Mar. Sci.*

- 65, 279–295 (2008).
16. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* (80-.). (1998). doi:10.1126/science.281.5374.237
 17. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* (2016). doi:10.1038/nature16942
 18. Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochem. Cycles* (2012). doi:10.1029/2011GB004099
 19. Bopp, L., Aumont, O., Cadule, P., Alvain, S. & Gehlen, M. Response of diatoms distribution to global warming and potential implications: A global model study. *Geophys. Res. Lett.* **32**, (2005).
 20. Richardson, T. L. Mechanisms and pathways of small-phytoplankton export from the surface ocean. *Annual Review of Marine Science* **11**, (2019).
 21. Tréguer, P. *et al.* Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience* **11**, (2018).
 22. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* (2006). doi:10.1016/j.patrec.2005.10.010
 23. Boyer, T. P. *et al.* WORLD OCEAN DATABASE 2013, NOAA Atlas NESDIS 72. Sydney Levitus, Ed.; Alexey Mishonoc, Tech. Ed. (2013). doi:10.7289/V5NZ85MT
 24. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* (2015). doi:10.5194/gmd-8-2465-2015
 25. Michelangeli, P. A., Vrac, M. & Loukos, H. Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.* (2009). doi:10.1029/2009GL038401
 26. The probable error of a mean. *Biometrika* **6**, (1908).
 27. L., K. & P., R. Clustering by means of Medoids. in *Statistical Data Analysis Based on the L1 Norm and Related Methods* (1987).
 28. Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, (1979).
 29. Laws, E. A., Falkowski, P. G., Smith, W. O., Ducklow, H. & McCarthy, J. J. Temperature effects on export production in the open ocean. *Global Biogeochem. Cycles* **14**, (2000).
 30. Biecek, P. DALEX: explainers for complex predictive models. *J. Mach. Learn. Res.* **19**, 1–5 (2018).
 31. Gent, P. R. *et al.* The community climate system model version 4. *J. Clim.* (2011).

doi:10.1175/2011JCLI4083.1

32. Dunne, J. P. *et al.* GFDL's ESM2 global coupled climate-carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. *J. Clim.* (2012). doi:10.1175/JCLI-D-11-00560.1
33. Collins, W. J. *et al.* Development and evaluation of an Earth-System model - HadGEM2. *Geosci. Model Dev.* (2011). doi:10.5194/gmd-4-1051-2011
34. Dufresne, J. L. *et al.* Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Clim. Dyn.* (2013). doi:10.1007/s00382-012-1636-1
35. Giorgetta, M. A. *et al.* Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.* (2013). doi:10.1002/jame.20038
36. Bentsen, M. *et al.* The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geosci. Model Dev.* (2013). doi:10.5194/gmd-6-687-2013
37. maps: Draw Geographical Maps. <https://cran.r-project.org/web/packages/maps/index.html>

IV. Article 3: Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics

Dans cet article¹²¹ j'ai réalisé la partie de modélisation des niches écologiques des génomes assemblés par métagénomes (MAGs) (Figure 4 de l'article) ainsi que la projection de leur biogéographie au temps présent et à l'horizon 2100 selon le scénario important d'émission de gaz à effets de serre (RCP8.5). Ces résultats sont complémentaires des analyses du deuxième article³²¹ et inclut notamment les génomes signatures évoqués dans ce dernier, confirmant les structures biogéographiques observés. Étant co-auteur, je ne présente ici que l'abstract de l'article. Sa version intégrale est accessible en ligne à l'adresse suivante : <https://www.sciencedirect.com/science/article/pii/S2666979X22000477>.

J'ai compilé les résultats de ce travail dans un site interactif accessible à l'URL suivant : https://end.mio.osupytheas.fr/Ecological_Niche_database/.

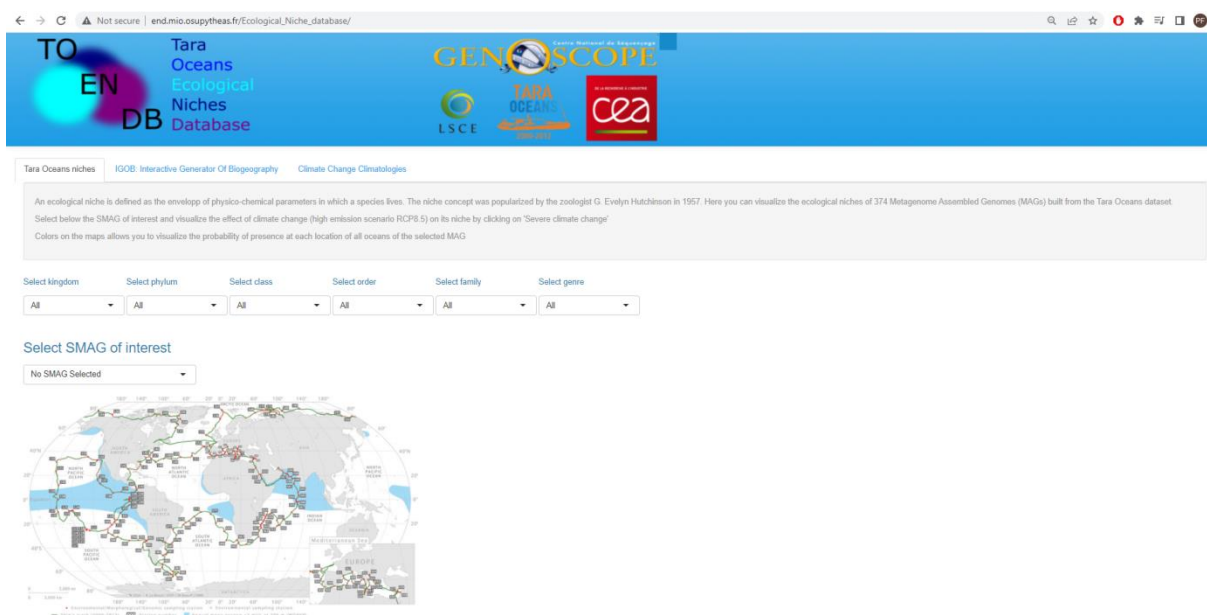


Figure 27 | Interface du site web nommé Tara Oceans Ecological Niches Database compilant les biogéographies de 374 MAGs eucaryotes issus de Delmont et al¹²¹.

Article

Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean

Tom O. Delmont,^{1,2,9,*} Morgan Gaia,^{1,2} Damien D. Hinsinger,^{1,2} Paul Frémont,^{1,2} Chiara Vanni,³ Antonio Fernandez-Guerra,⁴ A. Murat Eren,⁵ Artem Kourlaiev,^{1,2} Leo d'Agata,^{1,2} Quentin Clayssen,^{1,2} Emilie Villar,¹ Karine Labadie,^{1,2} Corinne Cruaud,^{1,2} Julie Poulain,^{1,2} Corinne Da Silva,^{1,2} Marc Wessner,^{1,2} Benjamin Noel,^{1,2} Jean-Marc Aury,^{1,2} Tara Oceans Coordinators, Colombar de Vargas,^{2,6} Chris Bowler,^{2,7} Eric Karsenti,^{2,6,8} Eric Pelletier,^{1,2} Patrick Wincker,^{1,2} and Olivier Jaillon^{1,2}

¹Génomique Métabolique, Genoscope, Institut François-Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, 91057 Evry, France

²Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GOSEE, 75016 Paris, France

³Microbial Genomics and Bioinformatics Research Group, Max Planck Institute for Marine Microbiology, Bremen, Germany

⁴Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

⁵Helmholtz Institute for Functional Marine Biodiversity at Oldenburg, Germany

⁶Sorbonne Université and CNRS, UMR 7144 (AD2M), ECOMAP, Station Biologique de Roscoff, Roscoff, France

⁷Institut de Biologie de l'ENS, Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France

⁸Directors' Research, European Molecular Biology Laboratory, Heidelberg, Germany

⁹Lead contact

*Correspondence: tom.delmont@genoscope.fr

<https://doi.org/10.1016/j.xgen.2022.100123>

SUMMARY

Marine planktonic eukaryotes play critical roles in global biogeochemical cycles and climate. However, their poor representation in culture collections limits our understanding of the evolutionary history and genomic underpinnings of planktonic ecosystems. Here, we used 280 billion Tara Oceans metagenomic reads from polar, temperate, and tropical sunlit oceans to reconstruct and manually curate more than 700 abundant and widespread eukaryotic environmental genomes ranging from 10 Mbp to 1.3 Gbp. This genomic resource covers a wide range of poorly characterized eukaryotic lineages that complement long-standing contributions from culture collections while better representing plankton in the upper layer of the oceans. We performed the first, to our knowledge, comprehensive genome-wide functional classification of abundant unicellular eukaryotic plankton, revealing four major groups connecting distantly related lineages. Neither trophic modes of plankton nor its vertical evolutionary history could completely explain the functional repertoire convergence of major eukaryotic lineages that coexisted within oceanic currents for millions of years.

CHAPITRE 2 : Changement d'expression génétique du plancton eucaryote dans la transition de l'Atlantique Nord à l'Océan Arctique

I. Introduction et résumé: la question de l'acclimatation et de l'atlantification

Après avoir exploré les possibilités migratoires des communautés de plancton face au réchauffement climatique, le second chapitre de cette thèse porte sur la possibilité des organismes à s'acclimater. L'acclimatation est ici définie comme la capacité des organismes à s'accoutumer à un milieu (défini par ses conditions environnementales) par modulation de leur physiologie via une transcription différentielle de leurs gènes³²². L'expression différentielle entre deux conditions d'un gène est sa transcription en plus grande abondance dans une des deux conditions c'est-à-dire que les cellules présentant une expression différentielle d'un gène dans un milieu donné aura un plus grand nombre de molécules d'ARN messager (ARNm) au sein de celle-ci. Une telle expression différentielle permettra potentiellement à la cellule de traduire un grand nombre d'ARNm en la protéine qu'il code. Il est à noter que, même si ce n'est pas nécessairement le cas³²³, les données protéomiques sont en général relativement bien corrélées avec les données protéomiques³²⁴. Nous utilisons ainsi dans l'article de ce chapitre les données métatranscriptomiques des expéditions *Tara Océans* de l'atlantique nord (Janvier-Mars 2012) et de l'expédition *Tara Polar Circle* (Mai à Septembre 2013) pour évaluer et comparer les physiologies des organismes planctoniques dans la transition entre l'Atlantique Nord (subtropical) et l'Océan Arctique (polaire). La transition entre ces deux océans constitue un excellent exemple de transition entre deux bassins ayant des conditions environnementales très différentes avec notamment d'importants fronts (fronts polaires) liés à la confrontation entre d'un côté des eaux chaudes et salines et de l'autre les eaux plus douces et froides polaires³²⁵. Ainsi l'Atlantique Nord est la porte d'entrée principales de courants océaniques tempérés (avec une contribution

moindre de l'Océan Pacifique) vers l'Arctique via la dérive nord atlantique (qui dérive du Gulf Stream) et entre dans l'arctique en se séparant en deux plus au nord. Ainsi les eaux atlantiques arrivent en Arctique soit par le détroit de Fram entre le Groenland et l'archipel du Svalbard soit par la mer de Barents plus à l'est³²⁵. La dérive nord atlantique constitue ainsi le courant de bord le plus important sur Terre¹⁷⁰. Dans le contexte du changement climatique cette transition apparaît particulièrement intéressante puisque la transition entre les deux océans subit un phénomène d'« Atlantification » lié au réchauffement des eaux atlantiques³²⁶, à la diminution de la couverture en glace³²⁷ et l'affaiblissement de l'halocline de ce bassin³²⁸. Ainsi certaines migrations vers le nord au sein de cette transition d'espèces planctoniques ont déjà été observées^{329,330}. De plus, les courants marins issus de l'atlantique semblent jouer un rôle très important dans ces migrations^{329,330} et dans le changement climatique de cette région³²⁶. Ainsi il est prévu qu'avec l'augmentation des températures, l'hétérotrophie augmente dans ce bassin et en conséquence au-delà de 6°C de réchauffement, l'arctique pourrait devenir une source de CO₂ au lieu d'un puits³³¹. Il apparaît ainsi important d'explorer les physiologies des organismes planctoniques qui s'acclimatent le long de cette transition et au sein de ces deux bassins.

Après une brève mise en contexte de la transition Atlantique Nord-Arctique, l'article de ce chapitre explore tout d'abord la connectivité océanique de la région et le lien entre la bêta-diversité des communautés eucaryotes du plancton (cf. chapitre 1), l'environnement et ces courants. Il est ainsi mis en évidence que la connectivité génomique Atlantique-Arctique *via* les courants est très différente de la connectivité mesurée surtout dans des océanstempérés du globe (Article 1, chapitre I). Ainsi le pic de corrélation dans la corrélation cumulative entre temps de transport et bêta-diversité se trouve sur un plateau entre 4 et 8 ans de transport (en opposition à 1,5 ans pour les autres océans, cf. Chapitre I, Article 1). Il en est de même pour la corrélation entre la distance environnementale et le temps de transport. Un léger contraste à noter est aussi une corrélation plus forte entre la bêta-diversité et les

temps de transport, provenant principalement des fortes différences de communautés de plancton entre les deux bassins, qu'entre les temps de transport et les paramètres environnementaux (ce qui était l'inverse pour le reste des océans, cf. Chapitre I, article 1).

Dans une deuxième partie, les gènes (unigenes du catalogue Marine Atlas of Tara Oceans Unigenes, MATOU), qui sont très principalement eucaryotes, sont classés selon leur appartenance propre à un des deux bassins ou leur partage entre les deux bassins. Chacun des deux bassins présente une forte endémicité avec un plus grand nombre de gènes uniques à l'atlantique (présentant néanmoins un plus grand nombre de gènes). Ainsi 55% des gènes considérés sont uniques à l'atlantique, 35% uniques à l'arctique et 10% sont partagés (mais représentent environ 25-30% des gènes arctiques et de l'abondance dans les métagénomés et métatranscriptomes). Une transition constituée de 3 stations (155, 158 et 163) apparaît enrichie en gènes communs (en nombre et en abondance relative) suggérant l'existence d'une zone de mélange importante au niveau de la transition et du front océanique polaire.

Un troisième résultat important de l'article est la mise en évidence de fonctions plus abondantes dans les métatranscriptomes et métagénomés arctiques que dans ceux de l'atlantiques. Il s'agit d'une analyse globale à l'échelle de l'ensemble des métatranscriptomes et métagénomés. Ainsi certaines fonctions liées à des fonctions post transcriptionnelles (méthylation de l'ARN notamment), traductionnelles (biogénèse des ribosomes) et post traductionnelles sont plus abondantes en arctiques. La méthylation de l'ARN est notamment connue pour jouer un rôle dans la stabilité des ARN et la résistance au froid³³². D'autre part, les ribosomes sont aussi connus pour être des indicateurs de la température pour la cellule³³³ et être altérés en condition de froid important³³⁴. L'ensemble de la machinerie cellulaire lié à l'expression génétique est ainsi différent dans l'arctique suggérant qu'il s'agit ici d'un mécanisme important d'acclimatation. D'autres fonctions liées à la fluidité des membranes sont aussi trouvées surabondantes en Arctique comme le métabolisme des sphingolipides et notamment des céramides critiques pour l'acclimatation au

froid^{335,336} ou encore des processus liés au catabolisme des lipides pouvant aussi jouer un rôle dans ces fonctions comme l'élimination d'acide gras libre en condition de froids extrêmes³³⁷. De manière intéressante, ces mécanismes semblent transcriptionnels et donc liés à une surexpression de ces fonctions et notamment dans certains groupes d'organismes fortement abondants (voir suppléments de l'article). Sans exclure un effet saisonnier (arctique échantillonné entre mai et septembre alors que l'atlantique a été échantillonné de janvier à mars), ces fonctions sont globalement plus abondantes sur l'ensemble des sites arctiques et de sa période d'échantillonnage (avec certaines variations internes à l'arctique) confortant l'idée qu'il s'agit de véritables différences fonctionnelles entre les deux bassins.

Dans une quatrième partie de l'article, une analyse plus fine à l'échelle de l'expression des gènes eucaryotes (unigenes de MATOU) a été réalisée le long des gradients environnementaux de la transition entre les deux bassins océaniques. Cette étude est réalisée sur les gènes communs aux deux bassins et présents dans un nombre suffisant de stations. De manière intéressante, la température est la variable environnementale physique qui est corrélée négativement aux profils d'expression d'un plus grand nombre de gènes (par rapport à la salinité et la durée du jour). L'analyse des fonctions de ces gènes négativement corrélés à la température montre une réponse commune des principaux clades d'algues marines unicellulaires (Bacillariophyta, Mamiellales, Phaeocystales et Pelagophyceae) suggérant une stratégie commune à l'entrée dans le bassin arctique à l'échelle de ces clades. Ce résultat est appuyé par le fait que ce n'est pas le cas pour les gènes corrélant positivement avec la température. En écho et en soutien à la deuxième partie de l'article de nombreuses fonctions allant de la transduction de signaux cellulaires jusqu'à la modification post traductionnelles des protéines, la synthèse d'acides aminés et en passant par les modifications post transcriptionnelles des ARNs et la synthèse des ribosomes sont enrichies parmi ces fonctions corrélant négativement avec le froid. L'hypothèse d'une grande réorganisation de la machinerie cellulaire par des changements transcriptionnels de l'expression génétique lors de l'entrée dans

l'arctique est émise. En plus de ces mécanismes, de nombreuses fonctions connues pour être liées à l'acclimatation au froid, au stress oxydatif ou encore à la durée du jour sont enrichies parmi les fonctions négativement corrélées à la température. Ces résultats sont montrés à la fois à l'échelle des clades d'algues cités précédemment mais aussi (encore plus clairement) à l'échelle de transcriptomes reconstruits par la métagénomique (MetaGenome based Transcriptomes en anglais, MGTs). Ces fonctions aux mécanismes variées sont souvent connues chez les plantes terrestres et mêmes dans d'autres règnes du vivant montrant à la fois leur universalité et diversité. D'autres fonctions liées au fonctionnement des mitochondries (métabolismes de stabilité énergétique de la cellule), au repliement protéique sont enrichies dans de nombreux organismes parmi les fonctions négativement corrélées à la température. Enfin dans certains organismes des fonctions liées à la division cellulaire, la biosynthèse de nucléotides et la réparation de l'ADN sont aussi enrichies.

Pour résumer, cet article montre un lien unique entre la connectivité océanique par les courants marins et les différences métagénomiques et métatranscriptomiques entre le bassin atlantique nord et l'arctique se différenciant du reste des océans mondiaux (chapitre I). Deuxièmement, des différences à la fois à l'échelle du fonctionnement global des communautés de phytoplanctons ainsi qu'à l'échelle plus fine de l'expression des gènes sont décrites et mettent en évidence des phénomènes de changements d'expression génétique chez un nombre important d'algues unicellulaires le long du gradient environnemental reliant les deux bassins. Ces mécanismes incluent notamment de nombreuses fonctions de régulation de l'expression génétique et de résistance et d'acclimatation au froid et au stress.

II. Article préliminaire: Gene expression changes of surface ocean eukaryotic plankton in the North Atlantic to Arctic Ocean transition

Cet article préliminaire a fait intervenir de nombreux laboratoires du Genoscope pour la création des jeux de données utilisées : le Laboratoire de Séquençage (LS, partie séquençage), le laboratoire de Bioinformatique pour la Génomique et la Biodiversité (LBGB, Corinne Da Silva, Artem Kourlaïev, Jean Marc Aury) et le Laboratoire d'Analyse Génomique des Eucaryotes (LAGE, Eric Pelletier). J'ai pour ma part réalisé l'ensemble des analyses statistiques et des figures de l'article. Nous avons aussi eu de nombreuses discussions sur la conceptualisation et les résultats du projet avec nos collaborateurs de Naples dont Daniele Iudicone et Lucia Campese. Chris Bowler (ENS) et Lee Karp Boss (Université du Maine) ont aussi apporté leurs commentaires et participé à la rédaction. L'article est disponible sur biorxiv à l'url suivant : <https://www.biorxiv.org/content/10.1101/2022.11.01.514737v1>.

1 **Changes in functional composition and gene** 2 **expression in eukaryotic plankton at the Atlantic-Arctic** 3 **Polar front**

4 Paul Frémont^{*1,2}, Corinne Da Silva¹, Lucia Campese³, Émilie Villar¹, Achal Rastogi^{1,4}, Jean-Marc
5 Aury¹, Chris Bowler^{5,2}, Lee Karp Boss⁶, Patrick Wincker^{1,2}, Eric Pelletier^{1,2}, Marion Gehlen⁷,
6 Daniele Iudicone³, Olivier Jaillon^{*1,2}

7 ^{*}Corresponding authors: Paul Frémont (pfremont@genoscope.cns.fr) and Olivier Jaillon (ojailon@genoscope.cns.fr)

8 ¹Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay,
9 Evry, France. ²Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara
10 Oceans, Paris, France. ³Stazione Zoologica Anton Dorn, Naples, Italy. ⁴Current address: Corteva Agriscience, The V
11 Ascendas, Atria Block, 12th Floor, Madhapur, Hyderabad 500081, India. ⁵Institut de Biologie de l'Ecole Normale
12 Supérieure (IBENS), Ecole Normale Supérieure, Centre National De La Recherche Scientifique (CNRS), Institut
13 National De La Santé Et De La Recherche Médicale (INSERM), Université Paris Sciences et Lettres (PSL), Paris,
14 France. ⁶School of Marine Sciences, University of Maine, Orono, Massachusetts, USA. ⁷LSCE-IPSL, CEA, CNRS,
15 Université Paris-Saclay, Gif-sur-Yvette, France.

16 **Abstract**

17 **Planktonic communities are shaped by ocean currents and participate in multiple**
18 **global biogeochemical cycles. Unraveling how the functions of communities**
19 **respond to strong environmental gradients will improve our understanding of the**
20 **interactions between climate change and marine ecosystems. Here, we investigate**
21 **changes in functions and gene expression of eukaryotic plankton transiting**
22 **between the North Atlantic Ocean (NAO) in winter and the Arctic Ocean (AO) in**
23 **spring/summer using metatranscriptomes and metagenomes from *Tara* Oceans. In**
24 **Arctic communities, functions involved in maintaining the protein pool and**
25 **translation machinery appear to be more active. Four major phylogenetically**
26 **distant algal groups are abundant in both basins and show similar strategies at**
27 **transcriptional level, including increased expression of some functions related to**
28 **cold acclimation. These results shed lights on gene expression strategies shared**
29 **by cosmopolitan phototrophs of widely separated lineages.**

30

31 **Introduction**

32 Planktonic microorganisms thrive in all oceans of the Earth under diverse conditions ranging from
33 highly oligotrophic, warm saline tropical gyres to highly seasonal, low-salinity and cold polar
34 waters. They are key players in many biogeochemical processes such as primary production via
35 photosynthesis (1), carbon export to the deep ocean (2–4) and essentially sustain the entire

36 oceanic trophic system (5). Many studies based either on satellite (6, 7) or genomic data (4, 8, 9)
37 revealed a mainly latitudinal and temperature driven biogeographical partitioning of the oceans.
38 While some species have strictly defined environmental niches (4, 10), some display cosmopolitan
39 biogeographies, such as *Bathycoccus prasinus* which is present in both polar and temperate
40 oceans (11) or *Pelagomonas calceolata* that is cosmopolitan in all except polar oceans (12).
41 Understanding how microorganisms acclimate and adapt (or not) to highly differentiated
42 environments as they are passively transported by currents to new physico-chemical conditions
43 is a critical question in the context of climate change. As an example, the transition between the
44 North Atlantic and Arctic waters is made through the North Atlantic Current (NAC) that is the most
45 important boundary current on Earth, with more than 40 Sv (40 million m³/s) flow rate of water in
46 the south and 20 Sv as it crosses the more eastward Mid-Atlantic Ridge (13). Major inflows of the
47 Arctic Ocean, mainly a landlocked sea, are therefore the warm and saline Atlantic Waters (AW)
48 that enter via the Barents Sea Opening and eastern Fram Strait. The warming of AW, the arctic
49 sea-ice loss (14, 15), along with the shoaling of the intermediate-depth Atlantic Water layer and
50 the increased winter ventilation in the eastern Eurasian Basin of the Arctic are making it
51 structurally similar to the Western Eurasian Basin (16, 17). This suite of processes has been called
52 “atlantification” of the Eurasian Basin and it is bringing the Arctic into a new climate state,
53 characterized by a much greater role of Atlantic inflows (18). Documented range expansions
54 attributed to warming include new appearances of temperate dinoflagellates and
55 coccolithophores in high-latitude subpolar regions (18, 19), while other planktonic taxa are
56 predicted to extend their ranges poleward under future climate conditions (20, 21).
57 This context raises the question of physiological acclimation to this new environment of incoming
58 species and consequences on this ecosystem. While the physiological response of plankton to
59 various environmental changes has been relatively well explored *in vitro* (10, 22, 23), studies are
60 only just beginning to explore key metabolisms to explore photosynthesis, nitrogen metabolism
61 (24), nutrient limitation (25) as well as community-level responses to iron availability (26) using *in*
62 *situ* omics data. Genomic adaptation and acclimation to cold environments has also been
63 investigated in the diatom *Fragilariopsis cylindrus* whose genome is notably enriched in metal-
64 binding proteins, as well as ice-binding proteins and other stress related proteins associated with
65 photosynthesis (27). In this diatom, many genes were also found to have bi-allelic differential
66 expression in the Southern Ocean (27).
67 As the Arctic warms, atlantification might turn the Arctic into a CO₂ source instead of a sink, which
68 would have major consequences on the global carbon budget (28, 29). For example in some

69 autotrophic plankton, when temperatures increase, the metabolism switches to heterotrophy and
70 its rate increases (30) in agreement with the metabolic theory of ecology (MTE) (31). In the context
71 of the atlantification, it appears critical to explore functional characterization of North Atlantic
72 plankton communities entering the Arctic Ocean and to compare the functioning of both basins.
73 Here we study metagenomic and metatranscriptomic data from eukaryotic plankton communities
74 to characterize their differences in composition and function along their transport through the
75 North Atlantic and into the Arctic Ocean basin. Significant changes in gene expression levels of
76 certain functions, consistent across various lineages, emerge in light of the strong environmental
77 clines along this route.

78 **Results**

79 **Oceanographic context of the NAO/AO transition**

80 The *Tara* schooner sampled the North Atlantic Ocean between January and March 2012 and the
81 Arctic Ocean between May and September 2013 (*Tara* Oceans sampling sites 143 to 196). While
82 the sampling route did not sample both basins the same year, there is a continuum in the seasons
83 sampled. Both basins have a relatively high seasonality, yet recent studies (4, 8, 9) suggest that
84 relatively stable plankton biogeographies exist at the global scale, which is confirmed here (see
85 below). In terms of connectivity, values of minimum transport time by currents (T_{min}) computed
86 from previous studies (8, 9) between each proximal station (defined by this minimum transport
87 time) confirm that the flow of currents is directed from the Atlantic into the Arctic (Figure 1). To
88 enter the Arctic Ocean the boat first followed the western route via the Greenland Sea south of
89 the Fram strait (*Tara* Oceans stations 158 to 163, T_{min} =3.2 months). The boat then left the main
90 current and bifurcated to the Barents Sea route (*Tara* Oceans stations 163 to 168, T_{min} =7.5
91 months, stations 158 to 168 orange dashed arrow, T_{min} =6.1 months). From station_168 onward,
92 the boat sampled different Arctic regions under the influence of Atlantic inflows with the exceptions
93 of stations 191 to 196 which are impacted by Pacific inflows (Figure 1a and Supplementary Figure
94 1b). The polar front is identified by a steep temperature gradient located at the time of sampling
95 (June 2013) between station 158 and 163 (orange arrow Figure 1b). We further discuss
96 environmental differences between the two oceanic basins and in light of current connectivity in
97 Supplementary Discussion 1. Considering this strong environmental change along the NAC we
98 first studied the genomic differentiation of the eukaryote communities of the two basins.

99 **Genomic biogeography of the NAO/AO transition**

100 We updated a collection of ~158 million eukaryotic unigenes (longest assembled gene
101 sequences) constructed from metatranscriptomic data (32) that included the *Tara* Oceans Polar
102 Circle stations. For this study, we selected 69.2 million unigenes from the 0.8-2000 μ m size

103 fraction that is enriched in unicellular eukaryotes and small zooplankton. To explore
104 biogeographical structure, we studied changes in plankton community composition by estimating
105 2 types of β diversity (β D) among stations in metagenomes and metatranscriptomes (see
106 Methods, equations 4 and 5). These β diversity estimates were highly correlated (see Methods).
107 A strong differentiation was found between the Arctic and the Atlantic although each basin
108 appears to be relatively heterogeneous (Figure 2a). In the Arctic, this heterogeneity doesn't
109 correspond to the different oceanographic regions (33). We also note that, despite being
110 influenced by Pacific waters (34) stations 193 to 196 do not cluster together (Figure 2a). From the
111 β -diversity matrix clustering (Figure 2a), we classify stations 143-155 as NAO and 158-196 as
112 AO. We can then define and analyze unigenes that are either specific to a single basin (NAO or
113 AO) or common to both basins (Methods). The taxonomic composition of these specific and
114 common unigenes, according to their relative abundances (Methods) is presented (Figure 2c and
115 2d) and discussed in the Supplementary Discussion 2 (Figure 2 and Supplementary Figure 5). In
116 total, 10% (n=6.9 million) of expressed unigenes were common to both basins, whereas 55%
117 (n=38.2 million) are unique to the North Atlantic and 35% (n=24.1 million) are unique to the Arctic.
118 The relative abundances of common unigenes reveal a clear transition between the two basins
119 corresponding to the polar front (Figure 2d). Stations 155, 158 (NAO side of the front) and 163
120 (AO side of the front) are found enriched with common unigenes (51.2% of the read abundance
121 on average and 42.3 % of the unigenes in number, Figure 2d, Supplementary Figures 3-4). In the
122 Arctic, the common unigenes represent a larger proportion of both reads and number of unigenes
123 (Figure 2d and Supplementary Fig 3-4). This higher relative proportion of genes found only in the
124 Atlantic compared to the Arctic is likely due to a selective environmental bottleneck at the polar
125 front and lower species diversity in the Arctic (Supplementary Figure 4). Overall, these results
126 confirm the mixing of different communities from different water masses, as is well known in
127 oceanic fronts (35, 36). They also illustrate the existence of a cline of genomic composition of
128 communities at the polar front. This implies the existence of species that successfully cross the
129 environmental front (studied below).

130 **Linking genomic biogeography with physical connectivity**

131 To assess the role of ocean circulation on this biogeography, we then compared the minimum
132 transport times (T_{min}) separating sample sites with the corresponding β -diversity estimates
133 (Figure 2b, Supplementary Figure 1a). In contrast to the results of Richter et al. (8), a maximum
134 correlation is found between 4 and 8 years of transport time instead of 1.5 years. These are typical
135 times that separate Atlantic and Arctic stations (light blue dots Figure 2b). The strong correlation
136 with beta diversity indicates that the influence of currents on community composition previously

137 identified in the North Atlantic (8) extends far north beyond the Arctic Circle. However, this
138 maximum correlation is probably also due to the strong differentiation of the Atlantic and polar
139 communities.

140 We also found that the calculated environmental distance (ED) between sampling sites shows
141 the same maximum correlation of around 4-8 years with transport time (Supplementary
142 Discussion 1 and Supplementary Figure 1c) but also with β -diversity (β D) (Supplementary Figure
143 2b-c). These data indicate that β -diversity and currents are most strongly correlated (maximum
144 correlation at 0.75 vs. 0.67 for T_{min} vs. ED and 0.5 for ED vs. β D), in contrast to Richter et al. (8)
145 in which environmental distance is better correlated with β -diversity. These results highlight the
146 strong and not readily decomposable interactions between the physicochemical environment,
147 currents, and planktonic communities that characterize the seascape (37).

148 **Functional divergences between the eukaryotic communities of NAO and AO**

149 The major differences in the composition of the Atlantic and polar eukaryotic communities suggest
150 that they function differently. To investigate this question we used Gene Ontology (GO) functional
151 annotations of unigenes. Our unigene set contains 4,695 different Gene Ontology (GO) terms
152 (35). Statistical analysis reveals that 66 terms are significantly more abundant at the community
153 level in the Arctic and 11 in the Atlantic (examples in Figure 3 and comprehensive list in
154 Supplementary Table 1, t-test, Hommel correction $p < 0.05$). Out of these 66 GO terms in Arctic
155 communities, 20 are related to RNA processing and 11 are related to RNA methylation (6 to tRNA
156 methylation) (Figure 3a, Supplementary Table 1). RNA methylation has been reported to be
157 important in cold stressed plants (36). As an example among these 11 GO terms, 'tRNA (adenine-
158 N1-)-methyltransferase activity' is known to be important for tRNA stability (37) and has also been
159 reported to be involved in mRNA methylation and transcriptome protection in granules that are
160 formed within the cytoplasm of stressed cells (38). Finally, 'tRNA methylation' is also well
161 documented as being important in translational fidelity (39). In particular, stressful conditions have
162 been reported to change the balance between free tRNA and aa-tRNA (40) which is critical for the
163 accuracy of protein synthesis (41).

164 Many other significantly more abundant GO terms in the Arctic are related to regulation and
165 downstream processes of gene expression, such as regulation of gene expression ('snoRNA
166 binding' (Figure 3b), 'ncRNA processing', 'mitochondrial RNA metabolic process') and ribosome
167 biogenesis ('ribosome biogenesis' (Figure 3c), 'preribosome, small subunit precursor', and
168 'ribonucleoprotein complex biogenesis'). ncRNA, snoRNA (a type of ncRNA) and mitochondrial
169 RNA metabolism are known to play important roles in regulation of gene expression (42, 43).

170 snoRNAs are involved in chemical modification of RNAs, including by methylation, while ncRNA
171 (small RNAs) have been shown to play a critical role in cold acclimation in *Arabidopsis* (44).
172 Ribosomes are known to be cold and heat shock sensors (45) and their dynamics and stability
173 are altered in cold conditions (46). Functions associated with protein post-translational
174 modifications were also found to be significantly more abundant in the Arctic ('protein peptidyl-
175 prolyl isomerization', 'peptidyl-proline modification', 'serine-type exopeptidase activity' and
176 'serine-type carboxypeptidase activity' (Figure 3d), 'intracellular protein transmembrane
177 transport').

178 Taken together, these abundant Arctic functions raise the hypothesis that an important set of
179 reversible epigenetic processes at the transcriptional and translational level (RNAs), as well as
180 other post translational modifications are more active in the Arctic. At the community level, these
181 gene expression regulatory processes appear to be important to thrive in the Arctic during spring
182 and summer. In the Arctic, we find other abundant functions that are known to be important in
183 cold environments, such as maintenance of chromosome integrity as well as many functions
184 related to lipid catabolism (Figure 3e), membrane fluidity, and organization of organelle
185 membranes (Supplementary Discussion 3). In addition to these more abundant functions, our
186 results suggest a more rapid protein turnover in spring/summer Arctic plankton communities
187 (Supplementary Discussion 3).

188 While the functions we describe appear as being significantly more abundant in the Arctic when
189 we conduct the analysis at the community scale (all taxa together), this was not the case at the
190 individual taxon level (with the exception of some of the differentially more abundant taxa in the
191 Arctic) (Supplementary Figure 6). Furthermore, taxa contribute very unevenly to the abundance
192 of these functions (Figure 3). For example, the Ciliophora appear to be important contributors to
193 'lipid catabolic processes' (Figure 3) whereas the Dinophyceae are important contributors of the
194 GO term 'RNA methylation'. These differences are probably due to differences in the gene
195 repertoire of the different taxa. This community effect could be due to higher statistical power
196 when all taxa are grouped together. In any case, it may correspond to a similar transcriptional
197 response that occurs in many groups and becomes significant at the community level and to
198 expression regulation.

199 We then examined expression changes at a community level, which would be expected to be
200 shared by many taxa. In this case, a higher ratio of metatranscriptomic to metagenomic
201 abundance should be observed in the Arctic. We found most of these functions more expressed
202 at the community level, however this signal seems to come mainly from a few dominant groups
203 (Mamiellales, Dinophyceae, Haptophyta, Opisthokonta and unclassified Eukaryotes)

204 (Supplementary Fig. 6). On the other hand we note the absence of Bacillariophyta (diatoms) in
205 these groups. These functions, which are particularly important in the Arctic environment, are
206 overexpressed in a large number of taxa that therefore likely share the same strategy. In addition,
207 many of the associated GO terms are functionally consistent with each other (see Methods,
208 Supplementary Figures 6 and 7). Without excluding a seasonality effect and despite some
209 variations within the Arctic (Figure 3 d,b,e,f for instance), these functions appear more abundant
210 over the whole spatio-temporal sampling of the Arctic basin (from May to September, see
211 distribution plots of Figure 3).

212 Given the spatiotemporal continuity of the sampled sites, our results therefore indicate that we
213 can detect real functional differences between eukaryotic communities before and after the polar
214 front. We also detected 11 more abundant GO terms in Arctic metagenomes that would reflect an
215 enrichment of these functions in the genetic repertoire of Arctic species. Eight of these are also
216 functions related to gene expression (Supplementary Table 1). Only two are related to lipid
217 metabolism ("Sphingolipid metabolic process" and "Ceramide metabolic process") and are part of
218 the set of more abundant functions in the Arctic metatranscriptomes.

219 To complete the GO term analysis, we used the ALDEx2 package (47, 48) to compare PFAM
220 abundances between the AO and NAO (Supplementary Discussion 4 and 5). Interestingly, almost
221 all PFAMs associated with cold stress and cold acclimation were more abundant (but not
222 necessarily more highly expressed) in the Arctic (Supplementary Fig. 8 and Supplementary
223 Discussion 4).

224 Conversely, we also searched for GO terms that were more abundant in the Atlantic. Considering
225 all taxa, 11 GO terms were significantly more abundant in the metatranscriptomes and 85 in the
226 metagenomes (3 in common with all 11) (Supplementary Table 1 for the exhaustive list,
227 Supplementary Discussion 5, Supplementary Fig. 16). Overall, they relate to functions in central
228 carbon metabolism and lipid biosynthesis (contrasting with lipid catabolism in the Arctic). We also
229 found specific functions of photosynthesis (cyclic electron transfer chain and light harvesting
230 system) that contrast with the more abundant central functions of photosynthesis (PFAMs) in the
231 Arctic (Supplementary Discussion 5).

232 As discussed above, AO and NAO are strongly connected, and our data indeed confirm the
233 presence of many cosmopolitan taxa. The functional differences that emerged in our
234 metatranscriptomic data may correspond in these species to altered gene expression of some
235 key functions.

236 **Transcriptional changes in major eukaryotic plankton groups in the NAO/AO transition**

237 We then investigated gene expression changes in plankton groups that were found in abundance
238 in both basins. We hypothesize that in these groups the expression of certain genes may be
239 modified along the environmental gradient around the polar front. We specifically searched for
240 gene abundance patterns significantly correlated with different environmental parameters and
241 analyzed their functional and taxonomic annotation. We selected a set of ~485,000 unigenes that
242 were present in both metatranscriptomes and metagenomes from at least 5 stations covering both
243 basins. We determined an expression value for each unigene at each station from a
244 transformation of the metatranscriptomic abundance values and normalization by metagenomics
245 (see Methods and Supplementary Figs. 8-9 for reasoning).

246 We found ~88,500 instances of significant correlations, positive or negative, between gene
247 expression profiles and physical parameters (temperature, sunlight duration (SSD), and salinity).
248 These physical environmental parameters were highly correlated with each other, however
249 temperature is the parameter significantly correlated with the largest number of gene expression
250 profiles (Supplementary Figure 10, ~19,500 negative and ~12,700 positive, FDR=0.05,
251 Supplementary Figures 11-12). Among the unigenes negatively correlated with temperature, we
252 found a particularly large number annotated as Mamiellales, diatoms and Pelagophyceae (Figure
253 4a, Supplementary Figure 13). Therefore, we further analyzed the taxonomic and functional
254 distribution of these unigenes correlated negatively or positively with temperature by a clustering
255 and dimension reduction approach. For all algal taxa tested (Bacillariophyta, Mamiellales,
256 Phaeocystales and Pelagophyceae), all genes whose expression level was negatively correlated
257 with temperature overlapped strongly (Figure 4b) and clustered together (Figure 4c), apart from
258 copepods. These results suggest that the temperature cline encountered by plankton in this ocean
259 system leads to a transcriptomic response of a core set of functions in a similar way in eukaryotic
260 micro- and picophytoplankton, independently of their phylogeny. Nevertheless, beyond this set of
261 core functions, other functions appeared to follow the same response only in certain lineages
262 (Figure 4b). These were numerous for the Mamiellales, Bacillariophyta and Pelagophyceae, but
263 few in the Phaeocystales. We note that Mamiellales and diatoms have the most similar repertoire
264 of cold-responsive genes, in agreement with the functional convergence between these groups
265 recently proposed (49).

266 To better characterize this transcriptional response, we analyzed whether enrichments could be
267 detected in certain functions. To do so, we applied a rationale of metagenomic co-variation of
268 metatranscriptomic data to reconstitute *in silico* more or less partial gene repertoires of a species
269 according to a procedure we had already used (Methods). We thus obtained a new set of 1,000
270 Metagenomic-based transcriptomes, or MGTs. Of these, six algal MGTs (probably corresponding

271 to 5 species) with a sufficient number of unigenes significantly negatively correlated with
272 temperature contained a significant number of enriched functions (see Methods). Two of these
273 MGTs are annotated as *Bathycoccus* (Mamillales), one as *Micromonas* (Mamiellales), one as
274 *Phaeocystis*, one as *Aureococcus* (Pelagophyceae), and one as *Minutocellus* (a pennate diatom).
275 In these MGTs, we found enriched functions known to be directly related to cold acclimation, light
276 or osmotic stress, as well as a large number of functions related to the regulation of gene
277 expression (Figure 5). In addition, the enriched functions contain processes related to energy
278 consumption, DNA division/replication, motility, nucleotide biosynthesis, protein folding and DNA
279 repair (Figure 5). We discuss these functions in more detail in Supplementary Discussion 6.
280 Overall, our results suggest that some eukaryotic lineages of unicellular planktonic algae have
281 common mechanisms of transcriptional changes when moving from the Atlantic to the Arctic basin
282 during the seasons we studied. However, lineage-specific properties emerge, particularly with
283 respect to cold and stress response functions that are more diverse. In addition, regulation and
284 other processes related to gene expression, from signal transduction to post-translational
285 modification, are found to be highly active at the community and species levels, highlighting their
286 likely critical roles in the AO and NAO/AO transition.

287 **Discussion**

288 Using meta-omic approaches, we have characterized a differentiation of eukaryotic plankton into
289 two types of communities that occurs during its transport by currents from the North Atlantic in
290 winter to the Arctic Ocean in spring and summer. These comparisons show that the two
291 communities are functionally and compositionally divergent but share certain lineages that persist
292 despite the strong environmental front they encounter. We compared the rates of change in
293 community composition with that of gene expression in polar and non-polar samples as was done
294 on the prokaryotic component of the same plankton samples (24). Similar to what has been
295 reported for prokaryotes (24), we observe a very clear modification around 15°C with a greater
296 contribution of community turnover in polar samples compared to temperate samples
297 (Supplementary Fig. 19). However, community turnover is still significantly lower than gene
298 expression changes, unlike in prokaryotes (24). This could suggest that gene expression changes
299 are higher in eukaryotes, but this may be a consequence of the large proportion of non-coding
300 regions in some eukaryotic genomes resulting in less sampling of eukaryotic communities at the
301 genomic than transcriptomic level (many zeros). In our analysis, the functional divergences are
302 explained both by changes in gene abundance, and thus taxa, but also by the level of gene
303 expression. The functions that differentiate the two types of communities are related to gene
304 expression (from signal transduction to post-translational modification), lipid metabolism

305 (catabolism and membrane organization), protein trafficking, and telomere maintenance. Life in
306 the polar environment appears to require tight regulation of post-transcriptional mechanisms,
307 including RNA methylation, which is essential for RNA stability (37, 50) and protein translation
308 fidelity (40, 41, 44). For those species that successfully cross the polar front, which we refer to as
309 persistent species, the genes for which transcriptional changes are observed carry functions also
310 related to gene expression (from signal transduction to post-translational modification) as well as
311 cold/stress responses, energy consumption processes, protein folding/ubiquitination, cell
312 division/DNA replication, nucleotide biosynthesis, and DNA repair. To explain gene expression
313 variations, we tested an important set of environmental parameters including physical factors and
314 nutrient concentrations. However, nutrient concentration values tested were not sufficiently
315 homogeneous and representative. Among the remaining physical factors we tested (temperature,
316 day length and salinity), temperature was the most important factor in explaining variations in
317 gene expression.

318 A similar transcriptional response is described across the major algal lineages; Mamiellales,
319 Bacillariophyta, Dinophyceae and Phaeocystales all show a transcriptional response to cold that
320 is similar among them but clearly different from that of copepods. This algal transcriptional
321 response is characterized by an increase in all of these groups in the expression of certain
322 functions and by an increase in the expression of other functions only in specific groups. Functions
323 known to be directly related to cold stress acclimation in many land plants are enriched in these
324 data. This suggests that these cold acclimation mechanisms are shared across photosynthetic
325 lineages. It is difficult at this stage to infer the nature of the evolutionary events from which these
326 properties are derived and which are now shared by phylogenetically distant lineages. It is likely
327 to be a combination of convergence and ancient adaptations.

328 Overall our study suggests that a strong transcriptional response involving key universal and life-
329 necessary but also more specific functions such as stress and cold responsive genes is critical
330 for plankton to enter and keep on living in the Arctic environment (in spring and summer here).
331 Importantly the temperature gradient we investigate is concentrated in the few stations of the
332 transition and no more than two years of transport time (Figure 1a) showing that persisting
333 plankton have the time to respond. Indeed acclimation responses are believed to be quite fast, in
334 order of days to week (27, 51). While the transcriptional response we report seem critical to adapt
335 it might not be sufficient to explain the survival of persisting algae. Indeed organisms might also
336 need to adapt by directly changing the sequence of their proteins as recently suggested in
337 *Bathycoccus* (11) and in the SAR supergroup (52) or by mechanisms such as allele specific
338 expression recently reported in the copepod *Oithona similis* (53) and in the diatom *Fragilariopsis*

339 *cylindrus* (27). Therefore, the signal we observe is probably due to a mix of acclimation and
340 adaptation responses that we can't disentangle here. Critically, adaptation might require more
341 time (54) than acclimation although adaptation can occur relatively fast in large populations (55,
342 56). Investigating both transcriptional and sequence changes at a broader scale and for more
343 organisms seem now critical to better understand how plankton can thrive in so many climates,
344 what differentiate specialists from more ubiquitous species, changes from one to another (55) and
345 their response to a changing climate as well as consequences on ecosystems (28). Particularly,
346 in the context of atlantification it might be critical to assess which species have the potential to
347 adapt and acclimate to the new environment that is being formed. The strong connectivity allowed
348 by the Gulf Stream followed by the North Atlantic Current (13) is believed to play a critical role in
349 these processes (18) in the context of global warming (57, 58). Nevertheless, this circulation might
350 collapse with global warming (58–60).

351

352 **Materials and Methods**

353 **Marine Atlas of *Tara* Oceans Unigenes (MATOU-v1.2)**

354 We updated the original version of the Marine Atlas of *Tara* Oceans Unigenes (32) by including
355 samples encompassing the Arctic seas. This extension includes a total of 581 samples (+140)
356 from 6 different size fractions divided between 0.8 and 2000 μm . In total all samples are collected
357 at 88 locations across the oceans (previously 68) at three depths (Surface (SRF), Deep
358 Chlorophyll Maximum (DCM) and Mesopelagic zones (MES)). We followed the same
359 methodology as previously described for the sequencing, assembly, mapping, taxonomic and
360 functional annotation of the updated atlas (32). The new catalog totals ~158.5 million unigenes
361 increasing by ~41.7 million unigenes the previous version (+35.7%).

362 **Metagenomics-based Transcriptomes (MGT-v1.2)**

363 We generated an updated version of the Metagenomics-based Transcriptomes (MGTs) that
364 cluster unigenes based on their relative abundance covariations across *Tara* Oceans samples
365 into relevant transcriptomic and more or less taxonomically coherent units. As for the MATOU-
366 v1.2 catalog, *Tara* Oceans Polar Circle samples were added to rebuild the collection from scratch.
367 Clustering and taxonomic assignment of MGTs were computed using the same methodology as
368 Vorobev et al. (61). Briefly, the metagenomic RPKM-based (reads per kilo base per million of
369 mapped reads) abundance matrix of sufficiently occurring unigenes was submitted to the canopy
370 clustering algorithm described in Nielsen et al. (62). This is a density based clustering that does
371 not take into account the sequence composition, as opposed to most binning tools. We used a

372 max Pearson's correlation difference of 0.1 to define clusters and then clusters were merged if
373 canopy centroids' distances were smaller than 0.05 (250k iterations, default parameters). This
374 updated version extends the collection to a total of 1000 MGTs counting at least 1000 unigenes
375 each (previously 924 MGT with min. 500 unigenes, +8.2%) encompassing 9,547,691 unigenes
376 (previously 6,946,068 +37.5%).

377 **Environmental parameters and distance**

378 We use a set of 11 environmental parameters from three different sources (*In situ*, World Ocean
379 Atlas 2018, PISCES-v2 model and the U.S. Naval Observatory), all believed to be relevant
380 ecologically. We use three *physical* variables: Temperature and Salinity (*in situ Tara*
381 measurements (63)) and Sea Sunshine Duration (SSD, Calculated from Astronomical
382 Applications of the U.S. Naval Observatory). Two macronutrients are from *in situ Tara Oceans*
383 measurements: dissolved Silica (SiO_2), and phosphate (PO_4^{3-}) (64). Nitrate (NO_3^-) is extracted from
384 the World Ocean Atlas 2018 (65). From these we derive N^* and Si^* calculated as follows:

$$385 \quad N^* = [NO_3^-] - 16 * [PO_4^{3-}] \quad (1)$$

$$386 \quad Si^* = [SiO_2] - [NO_3^-] \quad (2)$$

387 N^* reflect the excess of Nitrate compared to phosphate and inversely the excess of phosphate
388 (expressed in the order or magnitudes of nitrate concentration according to a constant ratio of
389 1/16 (66–68)). This metric was shown to be importantly correlated with the habitat of diazotrophic
390 organisms (68). The second Si^* was shown to reflect competition between diatoms and other
391 organisms. We use one micronutrient (Iron) simulated from the PISCES-v2 model (69, 70). Finally
392 two seasonality indexes (temperature and nitrate) are calculated from the World Ocean Atlas
393 2018, reflecting seasonal maximal seasonal fluctuation over the years at the sampling locations:

$$394 \quad SI(X(x, y)) = \frac{\max_t(X(x, y, t)) - \min_t(X(x, y, t))}{\max_S(\max_t(X(x, y, t)) - \min_t(X(x, y, t)))} \quad (3)$$

395 X is either the nitrate concentration or the temperature. The indices t and S respectively refer to
396 time (months here) and spatial (*Tara* samples here) coordinates. All variables are extracted at the
397 month and sampling location of the *Tara* stations considered here.

398 An environmental distance is calculated as the Euclidean distance (71) of standardized
399 parameters (z-values (72)):

$$400 \quad ED(i, j, P) = \sqrt{\sum_{p \in P} (x_{ip} - x_{jp})^2}$$

401 $ED(i, j, P)$ is the environmental distance between station i and j for a set of parameters P (here
402 we used the whole set).

403 **Calculating -omic community dissimilarity**

404 Following Richter et al. (8), metagenomic and metatranscriptomic community distance between
405 pairs of samples was estimated using shotgun metagenomes and metatranscriptomes for size
406 fraction 0.8-2000 μm . We used an improved -omic comparison method (SimkaMin (73)) that
407 computes standard ecological distances by replacing species counts by counts of DNA sequence
408 k -mers (segments of length k). SimkaMin is less computationally intensive than the previous
409 Simka thanks to an efficient data subsampling scheme used to estimate Bray-Curtis and Jaccard
410 dissimilarities (73). β diversity for -omic reads is defined with the following equation within Simka:

$$411 \quad \beta \text{ diversity} = \frac{b+c}{2a+b+c} \quad (4)$$

412 Where a is the number of distinct k -mers shared between two samples, and b and c are the
413 number of distinct k -mers specific to each sample. We represented the distance between each
414 pair of samples on a heatmap using the heatmap.2 function of the R-package gplots_2.17.0 (74).
415 We also calculate β -diversity between station i and j based on MATOU-v2 unigenes abundances
416 as follow:

$$417 \quad \beta \text{ diversity}_{ij} = \frac{\sum_k |RA_k^i - RA_k^j|}{\sum_k |RA_k^i + RA_k^j|} \quad (5)$$

418 Where RA_k^i is the relative abundance of unigenes k in station i . We find high correlation between
419 unigenes based distance and simkaMin based distance ($r^2=0.96$ for metatranscriptomes-based
420 distances) as well as between metatranscriptomes and metagenomes distances ($r^2=0.99$ and
421 0.97 respectively for simkaMin and unigenes Bray-Curtis).

422 **Taxonomic description of the metatranscriptomes and metagenomes**

423 Each sample of our dataset was described taxonomically by taxonomical annotation of unigenes.
424 We harmonized unigene taxonomic annotation by selecting a set of ecologically relevant and
425 sufficiently abundant groups: Hexanauplia, Bacillariophyta, Pelagophyceae, Crysohyceae,
426 Coccochaerales, Phaeocystales, Ciliophora, Tunicata, Mamiellales, Dinophyceae, Bacteria,
427 Streptophyta, Cnidaria, Rhizaria, Insecta, Cryptophyta, Euglenozoa, Amoebozoa, Craniata,
428 Fungi, Viruses, Archaea, Eukaryota (unclassified), root, and respectively other Haptophyta,
429 Opisthokonta, Stramenopiles, Alveolata and Viridiplantae. Each unigene was then assigned a
430 geographical annotation: either unique to the Atlantic basin, unique to the Arctic basin, or shared
431 between the two (when a unigene was found in at least one station of the two basins). To find
432 taxonomical groups/clades that were significantly relatively more abundant in one of the two

433 basins, we applied a Wilcoxon test on the mean of the logarithm of the relative abundance of each
434 taxon across the station of the two basins and applied a Hommel correction (75) on the p-values
435 for multiple hypothesis correction testing. Note that for this analysis we used the logarithms of
436 abundances and not the centered log ratio (used for the unigene expression analysis; see below)
437 as the former might bias some abundances as it is centered on zero in the case of skewed
438 distributions (which is the case here).

439 **Functional characterization: comparing the Arctic and Atlantic basin**

440 For each unigene of the MATOU-v1.2 catalog we converted the PFAM (76) (protein family domain
441 annotation) to the Gene Ontology annotation (35, 77). We used the full GO dataset including the
442 three main categories of the catalog: molecular function, cellular component and biological
443 processes. Unigenes can have several PFAMs or PFAMs can map on several GOs. To take this
444 into account we duplicated the functional annotation and therefore each unigene is assigned to
445 each level of its Gene Ontology hierarchy and can have several annotations. For each GO term
446 found across our dataset (a total of 4,695 terms from a total of 44,000 in the database), we defined
447 its relative abundance within each sample of our data by summing the relative abundance of all
448 the unigenes annotated as such. We applied a Student test (78) on the means of the logarithms
449 of the relative abundances of each GO term between the two basin means, applied Hommel
450 multiple comparison (75) correction, and found significantly relatively more abundant functions (p-
451 val<0.05 after correction). Here we used only the logarithm of the relative abundance and not the
452 centered-log-ratio as the multiplicity and unevenness of levels in the GO hierarchy did not allow
453 us to define an average logarithm consistent across the hierarchy.

454 We also used ALDEx2 R package (47, 48) to compute significantly more abundant PFAMs
455 between North Atlantic and Arctic metagenomes. ALDEx2 uses standard Bayesian techniques
456 (79) to infer the posterior distribution of $[p_1, p_2, \dots, p_i, \dots, p_n]$ as the product of the multinomial
457 likelihood (where p_i is the proportions of gene i (here we use PFAMs) in a sample) with a
458 $Dirichlet(\frac{1}{2}, \frac{1}{2}, \dots)$ prior. ALDEx2 then uses the full Dirichlet posterior probabilities distribution to
459 draw multiple estimates (Monte-Carlo realizations) of each p_i such that
460 $[p_1, p_2, \dots, p_i, \dots, p_n] \sim Dirichlet([n_1, n_2, \dots, n_i, \dots, n_n] + \frac{1}{2})$. Drawing p_i from the full posterior allows
461 accounting for the large variance and extreme non-normality of the marginal distributions of p_i
462 when the associated n_i are small. p_i are then transformed using the centered-log-ratio (see next
463 paragraph for this transformation). ALDEx2 therefore models the abundance of each gene
464 (PFAMs for us) as a random-effect ANOVA model accounting for the fact that within-condition
465 abundance/expression is a non-negligible random variable:

466
$$q_{ijk} = \mu_{ij} + v_{ijk} + \tau_{ijk} + \varepsilon_{ijk} \quad (6)$$

467 Where q_{ijk} is the adjusted log-abundance/expression, μ_{ij} is the expected expression of gene i
468 within each condition j , v_{ijk} is the sample specific expression change for replicate k ,
469 τ_{ijk} is the sampling variation from inferring abundance from read counts and ε_{ijk} the remaining
470 nonspecific error, independent and identically distributed within k . To find significantly more
471 abundant PFAMs, ALDEx2 then performs a Welch t-test on each q_{ijk} estimates (drawn from the
472 Monte Carlo realization) and corrected for multiple comparison by the Benjamini-Hochberg
473 method (80)(80)7, we plotted the median of the following metrics to visualize between condition
474 differences as well as within condition variability:

475 1. The absolute fold difference between-conditions

476
$$\Delta_A(i, j, j') = W(i, j) - W(i', j') \quad (7)$$

477 Where $W(i, j) = \sum_{k=1}^n q_{ijk}$ is the within condition mixture. Index i indicates the PFAMs index, index
478 j indicates the condition index (Atlantic or Arctic) and finally k indexes the replicate of a given
479 condition.

480 2. The between sample, within condition difference

481
$$\Delta_W(i, j) = \max_{k'=k'} |q_{ijk} - q_{ijk'}| \quad (8)$$

482 **Unigene expression analysis and response to environmental parameters**

483 We defined expression of each of unigene as the difference between the centered-log-ratio of the
484 relative abundance of the unigene from metatranscriptomic reads and the centered-log-ratio of its
485 relative abundance from metagenomic reads:

486
$$E = clr(MetaT) - clr(MetaG) \quad (9)$$

487 The clr was used as it was shown to be stable to subsampling of the dataset (81). It is defined as
488 follows:

489
$$q_i = \log(p_i) - \frac{1}{m} \sum_{k=1}^m \log(p_k) \quad (10)$$

490 Where p_i is the relative abundance of unigene i and we have $\sum_{k=1}^m p_k = 1$. Only unigenes which
491 were found in at least five stations and present in both basins (at least one expression value
492 available in the two basins) were considered, totaling ~485,000 unigenes. This considerably
493 decreased the number of unigenes on which we could do the analysis but we believe this is the
494 most robust way to analyze unigene expression within metaomic samples by taking into account
495 both the metatranscriptomic and metagenomic abundances and having sufficient statistical power
496 (at least 5 stations). To validate this approach, we computed for each unigene the Erb rho

497 correlation coefficient (82) between $clr(MetaT)$ and $(MetaG)$. Importantly, they correlated well
498 with each other although more or less strongly depending on the taxa, justifying the validity of this
499 approach (Supplementary Fig. 8-9).

500 For each unigene expression profile we computed two Pearson correlation coefficients with each
501 environmental parameter: one correlating the ‘real’ data and one correlating ‘randomized’ data
502 with the given environmental parameter. The distribution of the values of the second test was
503 used as a null model. To find unigenes whose expression was significantly correlated to a
504 variable, we compared the distribution of correlation between ‘real’ data and ‘shuffled’ data. We
505 set a false discovery rate (FDR (83, 84)) of 0.05, *i.e.*, a unigene was considered significantly
506 differently expressed if its correlation was superior to the maximum correlation of the first 2.5%
507 (at each tail of the distribution) of the ordered correlations of the ‘randomized’ dataset. Then, to
508 have a more stringent view of the functional landscape of significantly correlated or significantly
509 differentially expressed (basin comparison), we performed an enrichment statistical test
510 (hypergeometric test (85), Hochberg correction, p -value=0.05). For this test, we used only the
511 deepest GO annotation of each unigene found within significantly correlated unigenes in
512 comparison to the functional GO annotations of all the unigenes from the given taxonomical group
513 or MGT. All these steps were performed for each taxonomical group and for the set of MGTs.

514 **Functional clustering of the GO catalog using the Levenshtein distance**

515 Extensive efforts have been made to cluster the Gene Ontology catalog using multiple Semantic
516 similarity Measures Approaches (86–88). We propose here a classification based on Levenshtein
517 distance (89). To do so, each GO term found within our dataset (4695 GOs) was represented by
518 a huge integer describing its position and ontology in the hierarchy of the Gene Ontology atlas.
519 Let go be the GO with the following hierarchy annotation:
520 GO:000XXX;GO00XXXX;GO0000XX;...;GO000XX (in the order of the hierarchy) the descriptive
521 number of the GO will be 000XXX00XXX0000XXX...000XX. Then we computed the Levenshtein
522 distance between two GOs $go1$ and $go2$ as follows:

$$523 \quad lev(go1, go2) = \begin{cases} |go1| & \text{if } |go1| = 0 \\ |go2| & \text{if } |go2| = 0 \\ lev(tail(go1, go2)) & \text{if } go1[0] = go2[0] \\ 1 + \min \begin{cases} lev(tail(go1), go2) \\ lev(go1, tail(go2)) \\ lev(tail(go1), tail(go2)) \end{cases} & \text{otherwise} \end{cases} \quad (11)$$

524 We then used UPGMA (90) hierarchical clustering to cluster the distance matrix using correlation
525 as distance. To optimize the hierarchical clustering, we found a seed that minimized the Silhouette

526 criteria for the set of clusters retrieved. After clustering, groups were defined as the largest clusters
527 with the bootstrap probability superior to 0.95. Clustering with only 1 bootstrap generally performs
528 better than doing 1,000 bootstraps to maximize the Silhouette criteria. After finding the optimal
529 seed (the one with the highest Silhouette), we added a cleaning step of the clusters by allocating
530 wrongly clustered GOs (negative silhouette) to clusters containing unique GOs. This second step
531 decreases slightly the average silhouette coefficient although it also decreases the number of
532 misclassified GOs (relatively to the Levenshtein distance). Therefore we privileged accuracy over
533 average efficiency of the clustering.

534 **Community composition and expression based distances**

535 We used estimated centered log ratios from the ALDEx2 package (47, 48) based on the PFAM
536 abundances to compute pairwise community composition distances and expression distances
537 between each pair of stations. Expression is defined as the difference between centered log ratios
538 of metatranscriptomic abundances and metagenomic abundances (Equation 9). Then the
539 distance between each pair of station is defined as the euclidean distance over all PFAMs :

$$540 \quad D_{ij} = \sqrt{\sum_{PFAMs} (x_i^{PF} - x_j^{PF})^2}$$

541 Where x_i^{PF} is either the abundance (centered log ratio) or the expression value of PFAM PF in
542 station i . Following Salazar et al. (24), we then computed, for temperature bins each containing
543 5 stations, the median ratio between abundance based distances and expression based distances
544 of all pairwise distances from stations found in each bin. A Wilcoxon test and holm correction for
545 multiple comparison ($p < 0.05$) to test whether this ratio is significantly different than 1 is then
546 applied for each bin.

547 **Author Contributions**

548 P.F., O.J., M.G. and D.I. conceived the study. P.F. computed the results, compiled, and analyzed
549 the data. P.F. wrote the initial draft of the paper. C.D.S. and E.P.: MATOU and unigene computing.
550 J.M.A.: Bioinformatics Control Quality of Sequencing. E.P.: MGT computing. L.C. and A.R.
551 helped with preliminary analysis and data compilation. E.V.: SimkaMin computing. P.F., O.J., D.I.,
552 M.G., E.P., L.K.B., C.B. and L.C. discussed the results and wrote the paper.

553 **References**

- 554 1. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere:
555 Integrating terrestrial and oceanic components. *Science (1979)* (1998),
556 doi:10.1126/science.281.5374.237.

- 557 2. L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R.
558 Brum, L. P. Coelho, J. C. I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M.
559 Picheral, J. Poulain, S. Searson, L. Stemann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-
560 Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. de Vargas,
561 D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving
562 carbon export in the oligotrophic ocean. *Nature* (2016), doi:10.1038/nature16942.
- 563 3. S. A. Henson, R. Sanders, E. Madsen, Global patterns in efficiency of particulate organic carbon
564 export and transfer to the deep ocean. *Global Biogeochem Cycles* (2012),
565 doi:10.1029/2011GB004099.
- 566 4. P. Frémont, M. Gehlen, M. Vrac, J. Leconte, T. O. Delmont, P. Wincker, D. Iudicone, O. Jaillon,
567 Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat*
568 *Clim Chang* (2022), doi:10.1038/s41558-022-01314-8.
- 569 5. F. Azam, T. Fenchel, J. Field, J. Gray, L. Meyer-Reil, F. Thingstad, The Ecological Role of Water-
570 Column Microbes in the Sea. *Mar Ecol Prog Ser* (1983), doi:10.3354/meps010257.
- 571 6. G. Reygondeau, W. W. L. Cheung, C. C. C. Wabnitz, V. W. Y. Lam, T. Frölicher, O. Maury, Climate
572 Change-Induced Emergence of Novel Biogeochemical Provinces. *Front Mar Sci*. **7** (2020),
573 doi:10.3389/fmars.2020.00657.
- 574 7. A. R. Fay, G. A. McKinley, Global open-ocean biomes: Mean and temporal variability. *Earth Syst*
575 *Sci Data* (2014), doi:10.5194/essd-6-273-2014.
- 576 8. D. J. Richter, R. Watteaux, T. Vannier, J. Leconte, P. Frémont, G. Reygondeau, N. Maillet, N.
577 Henry, G. Benoit, O. da Silva, T. O. Delmont, A. Fernández-Guerra, S. Suweis, R. Narci, C. Berney,
578 D. Eveillard, F. Gavory, L. Guidi, K. Labadie, E. Mahieu, J. Poulain, S. Romac, S. Roux, C. Dimier, S.
579 Kandels, M. Picheral, S. Searson, T. O. Coordinators, S. Pesant, J.-M. Aury, J. R. Brum, C. Lemaitre,
580 E. Pelletier, P. Bork, S. Sunagawa, F. Lombard, L. Karp-Boss, C. Bowler, M. B. Sullivan, E. Karsenti,
581 M. Mariadassou, I. Probert, P. Peterlongo, P. Wincker, C. de Vargas, M. Ribera d'Alcalà, D.
582 Iudicone, O. Jaillon, Genomic evidence for global ocean plankton biogeography shaped by large-
583 scale current systems. *Elife*. **11**, e78129 (2022).
- 584 9. G. Sommeria-Klein, R. Watteaux, F. M. Ibarbalz, J. Pierella Karlusich Juan, D. Iudicone, C. Bowler,
585 H. Morlon, Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science*
586 (1979). **374**, 594–599 (2021).
- 587 10. A. C. Palmisano, C. W. Sullivan, PHYSIOLOGY OF SEA ICE DIATOMS. I. RESPONSE OF THREE POLAR
588 DIATOMS TO A SIMULATED SUMMER-WINTER TRANSITION. *J Phycol.* **18** (1982),
589 doi:10.1111/j.1529-8817.1982.tb03215.x.
- 590 11. J. Leconte, Y. Timsit, T. O. Delmont, M. Lescot, G. Piganeau, P. Wincker, O. Jaillon, *bioRxiv*, in
591 press, doi:10.1101/2021.07.13.452163.
- 592 12. N. Guérin, M. Ciccarella, E. Flamant, P. Frémont, S. Mangenot, B. Istace, B. Noel, S. Romac, C.
593 Bachy, M. Gachenot, E. Pelletier, A. Alberti, O. Jaillon, C. Cruaud, P. Wincker, J.-M. Aury, Q.
594 Carradec, *bioRxiv*, in press, doi:10.1101/2021.10.25.465678.

- 595 13. T. Rossby, The North Atlantic Current and surrounding waters: At the crossroads. *Reviews of*
596 *Geophysics*. **34** (1996), doi:10.1029/96RG02214.
- 597 14. M. Steele, T. Boyd, Retreat of the cold halocline layer in the Arctic Ocean. *J Geophys Res Oceans*.
598 **103** (1998), doi:10.1029/98jc00580.
- 599 15. I. v. Polyakov, L. A. Timokhov, V. A. Alexeev, S. Bacon, I. A. Dmitrenko, L. Fortier, I. E. Frolov, J. C.
600 Gascard, E. Hansen, V. v. Ivanov, S. Laxon, C. Mauritzen, D. Perovich, K. Shimada, H. L. Simmons,
601 V. T. Sokolov, M. Steele, J. Toole, Arctic ocean warming contributes to reduced polar ice cap. *J*
602 *Phys Oceanogr*. **40** (2010), doi:10.1175/2010JPO4339.1.
- 603 16. I. v. Polyakov, A. v. Pnyushkov, M. B. Alkire, I. M. Ashik, T. M. Baumann, E. C. Carmack, I.
604 Goszczko, J. Guthrie, V. v. Ivanov, T. Kanzow, R. Krishfield, R. Kwok, A. Sundfjord, J. Morison, R.
605 Rember, A. Yulin, Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the
606 Arctic Ocean. *Science* (1979). **356** (2017), doi:10.1126/science.aai8204.
- 607 17. I. v. Polyakov, T. P. Rippeth, I. Fer, M. B. Alkire, T. M. Baumann, E. C. Carmack, R. Ingvaldsen, V. v.
608 Ivanov, M. Janout, S. Lind, L. Padman, A. v. Pnyushkov, R. Rember, Weakening of cold halocline
609 layer exposes sea ice to oceanic heat in the eastern arctic ocean. *J Clim*. **33** (2020),
610 doi:10.1175/JCLI-D-19-0976.1.
- 611 18. L. Oziel, A. Baudena, M. Ardyna, P. Massicotte, A. Randelhoff, J. B. Sallée, R. B. Ingvaldsen, E.
612 Devred, M. Babin, Faster Atlantic currents drive poleward expansion of temperate phytoplankton
613 in the Arctic Ocean. *Nat Commun*. **11** (2020), doi:10.1038/s41467-020-15485-5.
- 614 19. G. Neukermans, L. Oziel, M. Babin, Increased intrusion of warming Atlantic water leads to rapid
615 expansion of temperate phytoplankton in the Arctic. *Glob Chang Biol*. **24** (2018),
616 doi:10.1111/gcb.14075.
- 617 20. D. A. Hutchins, F. Fu, Microorganisms and ocean global change. *Nat Microbiol* (2017),
618 doi:10.1038/nmicrobiol.2017.58.
- 619 21. J. Goldsmit, C. W. McKindsey, R. W. Schlegel, D. B. Stewart, P. Archambault, K. L. Howland, What
620 and where? Predicting invasion hotspots in the Arctic marine realm. *Glob Chang Biol*. **26** (2020),
621 doi:10.1111/gcb.15159.
- 622 22. M. Moulager, A. Monnier, B. Jesson, R. Bouvet, J. Mosser, C. Schwartz, L. Garnier, F. Corellou, F.
623 Y. Bouget, Light-dependent regulation of cell division in *Ostreococcus*: Evidence for a major
624 transcriptional input. *Plant Physiol*. **144** (2007), doi:10.1104/pp.107.096149.
- 625 23. J. S. O'Neill, G. van Ooijen, L. E. Dixon, C. Troein, F. Corellou, F. Y. Bouget, A. B. Reddy, A. J. Millar,
626 Circadian rhythms persist without transcription in a eukaryote. *Nature*. **469** (2011),
627 doi:10.1038/nature09654.
- 628 24. G. Salazar, L. Paoli, A. Alberti, J. Huerta-Cepas, H. J. Ruscheweyh, M. Cuenca, C. M. Field, L. P.
629 Coelho, C. Cruaud, S. Engelen, A. C. Gregory, K. Labadie, C. Marec, E. Pelletier, M. Royo-Llonch, S.
630 Roux, P. Sánchez, H. Uehara, A. A. Zayed, G. Zeller, M. Carmichael, C. Dimier, J. Ferland, S.
631 Kandels, M. Picheral, S. Pisarev, J. Poulain, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G.
632 Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O.

- 633 Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes,
634 C. Sardet, S. Speich, L. Stemann, M. B. Sullivan, S. Sunagawa, P. Wincker, Gene Expression
635 Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell*
636 (2019), doi:10.1016/j.cell.2019.10.014.
- 637 25. L. J. Ustick, A. A. Larkin, C. A. Garcia, N. S. Garcia, M. L. Brock, J. A. Lee, N. A. Wiseman, J. Keith
638 Moore, A. C. Martiny, Metagenomic analysis reveals global-scale patterns of ocean nutrient
639 limitation. *Science (1979)*. **372** (2021), doi:10.1126/science.abe6301.
- 640 26. L. Caputi, Q. Carradec, D. Eveillard, A. Kirilovsky, E. Pelletier, J. J. Pierella Karlusich, F. Rocha
641 Jimenez Vieira, E. Villar, S. Chaffron, S. Malviya, E. Scalco, S. G. Acinas, A. Alberti, J. M. Aury, A. S.
642 Benoiston, A. Bertrand, T. Biard, L. Bittner, M. Boccara, J. R. Brum, C. Brunet, G. Busseni, A.
643 Carratalà, H. Claustre, L. P. Coelho, S. Colin, S. D’Aniello, C. da Silva, M. del Core, H. Doré, S.
644 Gasparini, F. Kokoszka, J. L. Jamet, C. Lejeune, C. Lepoivre, M. Lescot, G. Lima-Mendez, F.
645 Lombard, J. Lukeš, N. Maillet, M. A. Madoui, E. Martinez, M. G. Mazzocchi, M. B. Néou, J. Paz-
646 Yepes, J. Poulain, S. Ramondenc, J. B. Romagnan, S. Roux, D. Salvagio Manta, R. Sanges, S. Speich,
647 M. Sprovieri, S. Sunagawa, V. Taillandier, A. Tanaka, L. Tirichine, C. Trottier, J. Uitz, A. Veluchamy,
648 J. Veselá, F. Vincent, S. Yau, S. Kandels-Lewis, S. Searson, C. Dimier, M. Picheral, P. Bork, E. Boss,
649 C. de Vargas, M. J. Follows, N. Grimsley, L. Guidi, P. Hingamp, E. Karsenti, P. Sordino, L.
650 Stemann, M. B. Sullivan, A. Tagliabue, A. Zingone, L. Garczarek, F. d’Ortenzio, P. Testor, F. Not,
651 M. R. d’Alcalà, P. Wincker, C. Bowler, D. Iudicone, Community-Level Responses to Iron Availability
652 in Open Ocean Plankton Ecosystems. *Global Biogeochem Cycles*. **33** (2019),
653 doi:10.1029/2018GB006022.
- 654 27. T. Mock, R. P. Otilar, J. Strauss, M. McMullan, P. Paajanen, J. Schmutz, A. Salamov, R. Sanges, A.
655 Toseland, B. J. Ward, A. E. Allen, C. L. Dupont, S. Frickenhaus, F. Maumus, A. Veluchamy, T. Wu, K.
656 W. Barry, A. Falcatore, M. I. Ferrante, A. E. Fortunato, G. Glöckner, A. Gruber, R. Hipkin, M. G.
657 Janech, P. G. Kroth, F. Leese, E. A. Lindquist, B. R. Lyon, J. Martin, C. Mayer, M. Parker, H.
658 Quesneville, J. A. Raymond, C. Uhlig, R. E. Valas, K. U. Valentin, A. Z. Worden, E. V. Armbrust, M.
659 D. Clark, C. Bowler, B. R. Green, V. Moulton, C. van Oosterhout, I. v. Grigoriev, Evolutionary
660 genomics of the cold-Adapted diatom *Fragilariopsis cylindrus*. *Nature*. **541** (2017),
661 doi:10.1038/nature20803.
- 662 28. C. M. Duarte, S. Agustí, P. Wassmann, J. M. Arrieta, M. Alcaraz, A. Coello, N. Marbà, I. E. Hendriks,
663 J. Holding, I. García-Zarandona, E. Kritzberg, D. Vaqué, Tipping elements in the Arctic marine
664 ecosystem. *Ambio*. **41** (2012), doi:10.1007/s13280-011-0224-7.
- 665 29. P. Friedlingstein, M. O’Sullivan, M. W. Jones, R. M. Andrew, J. Hauck, A. Olsen, G. P. Peters, W.
666 Peters, J. Pongratz, S. Sitch, C. le Quéré, J. G. Canadell, P. Ciais, R. B. Jackson, S. Alin, L. E. O. C.
667 Aragão, A. Arneeth, V. Arora, N. R. Bates, M. Becker, A. Benoit-Cattin, H. C. Bittig, L. Bopp, S.
668 Bultan, N. Chandra, F. Chevallier, L. P. Chini, W. Evans, L. Florentie, P. M. Forster, T. Gasser, M.
669 Gehlen, D. Gilfillan, T. Gkritzalis, L. Gregor, N. Gruber, I. Harris, K. Hartung, V. Haverd, R. A.
670 Houghton, T. Ilyina, A. K. Jain, E. Joetzjer, K. Kadono, E. Kato, V. Kitidis, J. I. Korsbakken, P.
671 Landschützer, N. Lefèvre, A. Lenton, S. Lienert, Z. Liu, D. Lombardozzi, G. Marland, N. Metzl, D. R.
672 Munro, J. E. M. S. Nabel, S. I. Nakaoka, Y. Niwa, K. O’Brien, T. Ono, P. I. Palmer, D. Pierrot, B.
673 Poulter, L. Resplandy, E. Robertson, C. Rödenbeck, J. Schwinger, R. Séférian, I. Skjelvan, A. J. P.
674 Smith, A. J. Sutton, T. Tanhua, P. P. Tans, H. Tian, B. Tilbrook, G. van der Werf, N. Vuichard, A. P.

- 675 Walker, R. Wanninkhof, A. J. Watson, D. Willis, A. J. Wiltshire, W. Yuan, X. Yue, S. Zaehle, Global
676 Carbon Budget 2020. *Earth Syst Sci Data*. **12** (2020), doi:10.5194/essd-12-3269-2020.
- 677 30. J. M. Holding, C. M. Duarte, J. M. Arrieta, R. Vaquer-Suyner, A. Coello-Camba, P. Wassmann, S.
678 Agustí, Experimentally determined temperature thresholds for Arctic plankton community
679 metabolism. *Biogeosciences*. **10** (2013), doi:10.5194/bg-10-357-2013.
- 680 31. J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, G. B. West, "Toward a metabolic theory of
681 ecology" in *Ecology* (2004), vol. 85.
- 682 32. Q. Carradec, E. Pelletier, C. da Silva, A. Alberti, Y. Seeleuthner, R. Blanc-Mathieu, G. Lima-
683 Mendez, F. Rocha, L. Tirichine, K. Labadie, A. Kirilovsky, A. Bertrand, S. Engelen, M. A. Madoui, R.
684 Méheust, J. Poulain, S. Romac, D. J. Richter, G. Yoshikawa, C. Dimier, S. Kandels-Lewis, M.
685 Picheral, S. Searson, S. G. Acinas, E. Boss, M. Follows, G. Gorsky, N. Grimsley, L. Karp-Boss, U.
686 Krzic, S. Pesant, E. G. Reynaud, C. Sardet, M. Sieracki, S. Speich, L. Stemmann, D. Velayoudon, J.
687 Weissenbach, O. Jaillon, J. M. Aury, E. Karsenti, M. B. Sullivan, S. Sunagawa, P. Bork, F. Not, P.
688 Hingamp, J. Raes, L. Guidi, H. Ogata, C. de Vargas, D. Iudicone, C. Bowler, P. Wincker, A global
689 ocean atlas of eukaryotic genes. *Nat Commun* (2018), doi:10.1038/s41467-017-02342-1.
- 690 33. M. Royo-Llonch, P. Sánchez, C. Ruiz-González, G. Salazar, C. Pedrós-Alió, M. Sebastián, K. Labadie,
691 L. Paoli, F. M. Ibarbalz, L. Zinger, B. Churchward, M. Babin, P. Bork, E. Boss, G. Cochrane, C. de
692 Vargas, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels, F. Not, H.
693 Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Setmmann, M. B. Sullivan, S.
694 Chaffron, D. Eveillard, E. Karsenti, S. Sunagawa, P. Wincker, L. Karp-Boss, C. Bowler, S. G. Acinas,
695 T. O. Coordinators, Compendium of 530 metagenome-assembled bacterial and archaeal genomes
696 from the polar Arctic Ocean. *Nat Microbiol*. **6**, 1561–1574 (2021).
- 697 34. P. Wassmann, K. N. Kosobokova, D. Slagstad, K. F. Drinkwater, R. R. Hopcroft, S. E. Moore, I.
698 Ellingsen, R. J. Nelson, E. Carmack, E. Popova, J. Berge, The contiguous domains of Arctic Ocean
699 advection: Trails of life and death. *Prog Oceanogr*. **139** (2015),
700 doi:10.1016/j.pocean.2015.06.011.
- 701 35. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S.
702 S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J.
703 E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of
704 biology. *Nat Genet*. **25** (2000), , doi:10.1038/75556.
- 705 36. Y. Wang, C. Pang, X. Li, Z. Hu, Z. Lv, B. Zheng, P. Chen, Identification of tRNA nucleoside
706 modification genes critical for stress response and development in rice and Arabidopsis. *BMC*
707 *Plant Biol*. **17** (2017), doi:10.1186/s12870-017-1206-0.
- 708 37. C. Zhang, G. Jia, Reversible RNA Modification N1-methyladenosine (m1A) in mRNA and tRNA.
709 *Genomics Proteomics Bioinformatics*. **16** (2018), , doi:10.1016/j.gpb.2018.03.003.
- 710 38. M. Alriquet, G. Calloni, A. MartÄ±ñez-LimoÄ±n, R. Delli Ponti, G. Hanspach, M. Hengesbach, G. G.
711 Tartaglia, R. M. Vabulas, The protective role of m1A during stress-induced granulation. *J Mol Cell*
712 *Biol*. **12** (2020), doi:10.1093/jmcb/mjaa023.

- 713 39. C. Valadon, O. Namy, The importance of the Epi-transcriptome in translation fidelity. *Noncoding*
714 *RNA*. **7** (2021), , doi:10.3390/ncrna7030051.
- 715 40. Y. L. J. Pang, R. Abo, S. S. Levine, P. C. Dedon, Diverse cell stresses induce unique patterns of tRNA
716 up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. *Nucleic Acids*
717 *Res*. **42** (2014), doi:10.1093/nar/gku945.
- 718 41. A. Fluitt, E. Pienaar, H. Viljoen, Ribosome kinetics and aa-tRNA competition determine rate and
719 fidelity of peptide synthesis. *Comput Biol Chem*. **31** (2007),
720 doi:10.1016/j.compbiolchem.2007.07.003.
- 721 42. T. J. Nicholls, J. Rorbach, M. Minczuk, Mitochondria: Mitochondrial RNA metabolism and human
722 disease. *Int J Biochem Cell Biol*. **45**, 845–849 (2013).
- 723 43. G. Dieci, M. Preti, B. Montanini, Eukaryotic snoRNAs: A paradigm for gene expression flexibility.
724 *Genomics*. **94**, 83–88 (2009).
- 725 44. B. Tiwari, K. Habermann, M. A. Arif, H. L. Weil, A. Garcia-Molina, T. Kleine, T. Mühlhaus, W. Frank,
726 Identification of small RNAs during cold acclimation in *Arabidopsis thaliana*. *BMC Plant Biol*. **20**
727 (2020), doi:10.1186/s12870-020-02511-3.
- 728 45. R. A. VanBogelen, F. C. Neidhardt, Ribosomes as sensors of heat and cold shock in *Escherichia*
729 *coli*. *Proc Natl Acad Sci U S A*. **87** (1990), doi:10.1073/pnas.87.15.5589.
- 730 46. D. O. Bayles, M. H. Tunick, T. A. Foglia, A. J. Miller, Cold shock and its effect on ribosomes and
731 thermal tolerance in *Listeria monocytogenes*. *Appl Environ Microbiol*. **66** (2000),
732 doi:10.1128/AEM.66.10.4351-4355.2000.
- 733 47. A. D. Fernandes, J. M. Macklaim, T. G. Linn, G. Reid, G. B. Gloor, ANOVA-Like Differential
734 Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One*. **8** (2013),
735 doi:10.1371/journal.pone.0067019.
- 736 48. A. D. Fernandes, J. N. S. Reid, J. M. Macklaim, T. A. McMurrough, D. R. Edgell, G. B. Gloor,
737 Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA
738 gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*.
739 **2** (2014), doi:10.1186/2049-2618-2-15.
- 740 49. T. O. Delmont, M. Gaia, D. D. Hinsinger, P. Frémont, C. Vanni, A. Fernandez-Guerra, A. M. Eren, A.
741 Kourlaiev, L. d’Agata, Q. Clayssen, E. Villar, K. Labadie, C. Cruaud, J. Poulain, C. da Silva, M.
742 Wessner, B. Noel, J.-M. Aury, S. Sunagawa, S. G. Acinas, P. Bork, E. Karsenti, C. Bowler, C. Sardet,
743 L. Stemmann, C. de Vargas, P. Wincker, M. Lescot, M. Babin, G. Gorsky, N. Grimsley, L. Guidi, P.
744 Hingamp, O. Jaillon, S. Kandels, D. Iudicone, H. Ogata, S. Pesant, M. B. Sullivan, F. Not, K.-B. Lee,
745 E. Boss, G. Cochrane, M. Follows, N. Poulton, J. Raes, M. Sieracki, S. Speich, C. de Vargas, C.
746 Bowler, E. Karsenti, E. Pelletier, P. Wincker, O. Jaillon, Functional repertoire convergence of
747 distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics*. **2**,
748 100123 (2022).
- 749 50. T. Ohira, T. Suzuki, Precursors of tRNAs are stabilized by methylguanosine cap structures. *Nat*
750 *Chem Biol*. **12** (2016), doi:10.1038/nchembio.2117.

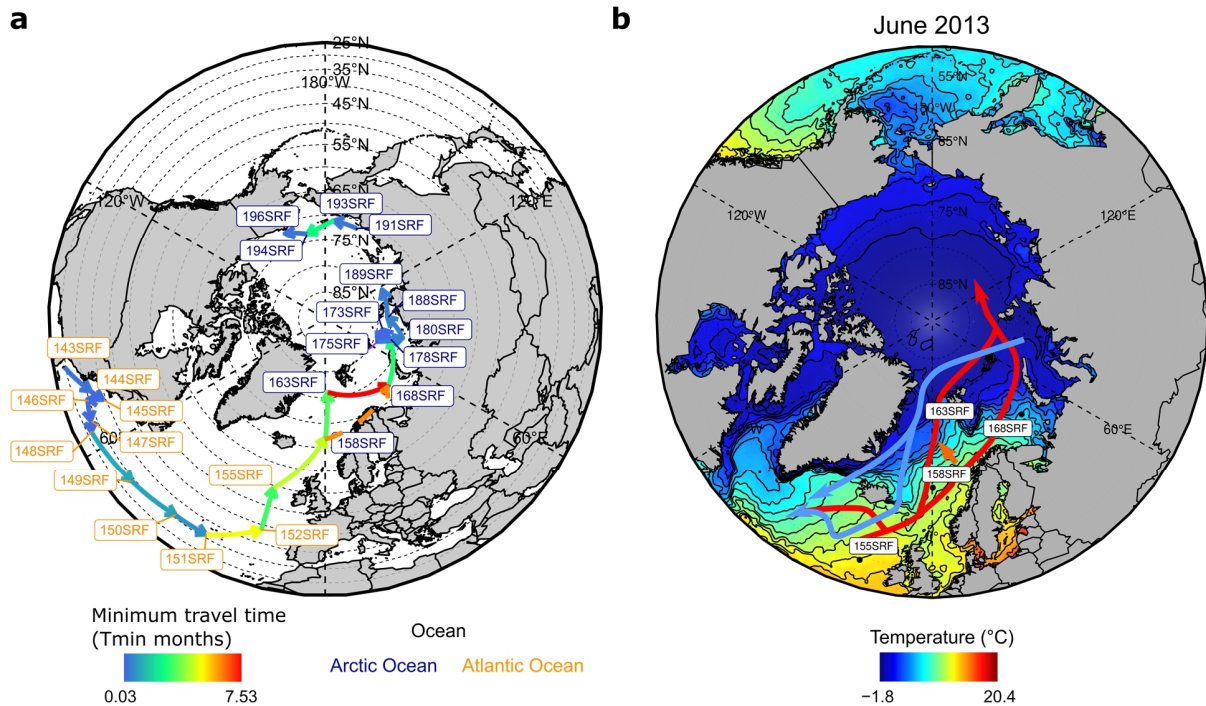
- 751 51. S. Abu-Ghosh, Z. Dubinsky, D. Iluz, Acclimation of thermotolerant algae to light and temperature
752 interaction. *J Phycol.* **56** (2020), doi:10.1111/jpy.12964.
- 753 52. T. O. Delmont, E. Kiefl, O. Kilinc, O. C. Esen, I. Uysal, M. S. Rappé, S. Giovannoni, A. M. Eren,
754 Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global
755 SAR11 subclade. *Elife* (2019), doi:10.7554/eLife.46497.
- 756 53. R. Laso-Jadart, K. Sugier, E. Petit, K. Labadie, P. Peterlongo, C. Ambroise, P. Wincker, J. L. Jamet,
757 M. A. Madoui, Investigating population-scale allelic differential expression in wild populations of
758 *Oithona similis* (Cyclopoida, Claus, 1866). *Ecol Evol* (2020), doi:10.1002/ece3.6588.
- 759 54. L. Schlüter, K. T. Lohbeck, J. P. Gröger, U. Riebesell, T. B. H. Reusch, Long-term dynamics of
760 adaptive evolution in a globally important phytoplankton species to ocean acidification. *Sci Adv.* **2**
761 (2016), doi:10.1126/sciadv.1501660.
- 762 55. P. Jin, S. Agustí, Fast adaptation of tropical diatoms to increased warming with trade-offs. *Sci Rep.*
763 **8** (2018), doi:10.1038/s41598-018-36091-y.
- 764 56. A. J. Irwin, Z. v. Finkel, F. E. Müller-Karger, L. Troccoli Ghinaglia, Phytoplankton adapt to changing
765 ocean environments. *Proceedings of the National Academy of Sciences* (2015),
766 doi:10.1073/pnas.1414752112.
- 767 57. D. J. R. Thornalley, D. W. Oppo, P. Ortega, J. I. Robson, C. M. Brierley, R. Davis, I. R. Hall, P. Moffa-
768 Sanchez, N. L. Rose, P. T. Spooner, I. Yashayaev, L. D. Keigwin, Anomalously weak Labrador Sea
769 convection and Atlantic overturning during the past 150 years. *Nature.* **556** (2018),
770 doi:10.1038/s41586-018-0007-4.
- 771 58. L. Johannes, D. P. D., Risk of tipping the overturning circulation due to increasing rates of ice
772 melt. *Proceedings of the National Academy of Sciences.* **118**, e2017989118 (2021).
- 773 59. K. H. Kilbourne, A. D. Wanamaker, P. Moffa-Sanchez, D. J. Reynolds, D. E. Amrhein, P. G. Butler,
774 G. Gebbie, M. Goes, M. F. Jansen, C. M. Little, M. Mette, E. Moreno-Chamarro, P. Ortega, B. L.
775 Otto-Bliesner, T. Rossby, J. Scourse, N. M. Whitney, Atlantic circulation change still uncertain. *Nat*
776 *Geosci.* **15**, 165–167 (2022).
- 777 60. S. Rahmstorf, J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, E. J. Schaffernicht,
778 Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat Clim*
779 *Chang* (2015), doi:10.1038/nclimate2554.
- 780 61. A. Vorobev, M. Dupouy, Q. Carradec, T. O. Delmont, A. Annamalé, P. Wincker, E. Pelletier,
781 Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities
782 via high-throughput metagenomics and metatranscriptomics. *Genome Res.* **30** (2020),
783 doi:10.1101/gr.253070.119.
- 784 62. H. B. Nielsen, M. Almeida, A. S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, D. R. Plichta, L. Gautier,
785 A. G. Pedersen, E. le Chatelier, E. Pelletier, I. Bonde, T. Nielsen, C. Manichanh, M. Arumugam, J.
786 M. Batto, M. B. Quintanilha Dos Santos, N. Blom, N. Borruel, K. S. Burgdorf, F. Boumezeur, F.
787 Casellas, J. Doré, P. Dworzynski, F. Guarner, T. Hansen, F. Hildebrand, R. S. Kaas, S. Kennedy, K.
788 Kristiansen, J. R. Kultima, P. Leonard, F. Levenez, O. Lund, B. Moumen, D. le Paslier, N. Pons, O.

- 789 Pedersen, E. Prifti, J. Qin, J. Raes, S. Sørensen, J. Tap, S. Tims, D. W. Ussery, T. Yamada, P. Renault,
790 T. Sicheritz-Ponten, P. Bork, J. Wang, S. Brunak, S. D. Ehrlich, A. Jamet, A. Mérieux, A. Cultrone, A.
791 Torrejon, B. Quinquis, C. Brechot, C. Delorme, C. M'rini, W. M. de Vos, E. Maguin, E. Varela, E.
792 Guedon, F. Gwen, F. Haimet, F. Artiguenave, G. Vandemeulebrouck, G. Denariaz, G. Khaci, H.
793 Blottière, J. Knol, J. Weissenbach, J. E. T. van Hylckama Vlieg, J. Torben, J. Parkhill, K. Turner, M.
794 van de Guchte, M. Antolin, M. Rescigno, M. Kleerebezem, M. Derrien, N. Galleron, N. Sanchez, N.
795 Grarup, P. Veiga, R. Oozeer, R. Dervyn, S. Layec, T. Bruls, Y. Winogradski, Z. Erwin, Identification
796 and assembly of genomes and genetic elements in complex metagenomic samples without using
797 reference genomes. *Nat Biotechnol.* **32** (2014), doi:10.1038/nbt.2939.
- 798 63. L. Guidi, M. Picheral, S. Pesant, C. Tara Oceans Consortium, P. Tara Oceans Expedition,
799 "Environmental context of all samples from the Tara Oceans Expedition (2009-2013), about
800 sensor data in the targeted environmental feature" in (PANGAEA, 2017;
801 <https://doi.org/10.1594/PANGAEA.875576>).
- 802 64. L. Guidi, P. Morin, L. Coppola, J.-É. Tremblay, S. Pesant, C. Tara Oceans Consortium, P. Tara
803 Oceans Expedition, "Environmental context of all samples from the Tara Oceans Expedition
804 (2009-2013), about nutrients in the targeted environmental feature" in (PANGAEA, 2017;
805 <https://doi.org/10.1594/PANGAEA.875575>).
- 806 65. H. Garcia, K. W. Weathers, C. R. Paver, I. Smolyar, T. P. Boyer, R. A. Locarnini, M. M. Zweng, A. V.
807 Mishonov, O. K. Baranova, D. Seidov, J. R. Reagan, World Ocean Atlas 2018. Volume 4: Dissolved
808 Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). *NOAA Atlas NESDIS 84.* **84**
809 (2018).
- 810 66. P. Brun, M. Vogt, M. R. Payne, N. Gruber, C. J. O'Brien, E. T. Buitenhuis, C. le Quéré, K. Leblanc, Y.
811 W. Luo, Ecological niches of open ocean phytoplankton taxa. *Limnol Oceanogr.* **60** (2015),
812 doi:10.1002/lno.10074.
- 813 67. N. Gruber, J. L. Sarmiento, Global patterns of marine nitrogen fixation and denitrification. *Global*
814 *Biogeochem Cycles.* **11** (1997), doi:10.1029/97GB00077.
- 815 68. C. Deutsch, J. L. Sarmiento, D. M. Sigman, N. Gruber, J. P. Dunne, Spatial coupling of nitrogen
816 inputs and losses in the ocean. *Nature.* **445** (2007), doi:10.1038/nature05392.
- 817 69. O. Aumont, C. Ethé, A. Tagliabue, L. Bopp, M. Gehlen, PISCES-v2: An ocean biogeochemical model
818 for carbon and ecosystem studies. *Geosci Model Dev* (2015), doi:10.5194/gmd-8-2465-2015.
- 819 70. A. Tagliabue, O. Aumont, R. Death, J. P. Dunne, S. Dutkiewicz, E. Galbraith, K. Misumi, J. K. Moore,
820 A. Ridgwell, E. Sherman, C. Stock, M. Vichi, C. Völker, A. Yool, How well do global ocean
821 biogeochemistry models simulate dissolved iron distributions? *Global Biogeochem Cycles.* **30**
822 (2016), doi:10.1002/2015GB005289.
- 823 71. P. C. Legendre, L. Legendre, Numerical Ecology, Volume 24. (*Developments in Environmental*
824 *Modelling*) (1988), doi:10.1017/CBO9781107415324.004.
- 825 72. S. Nielsen, Numerical Ecology. Legendre P. and Legendre L., second ed., Elsevier, Amsterdam, p.
826 853, 1998. *Ecological Modelling - ECOL MODEL.* **132**, 303–304 (2000).

- 827 73. G. Benoit, M. Mariadassou, S. Robin, S. Schbath, P. Peterlongo, C. Lemaitre, SimkaMin: Fast and
828 resource frugal de novo comparative metagenomics. *Bioinformatics*. **36** (2020),
829 doi:10.1093/bioinformatics/btz685.
- 830 74. G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A.
831 Magnusson, S. Moeller, M. Schwartz, B. Venables, Package “gplots”: Various R programming
832 tools for plotting data. *R package version 2.17.0*. (2016).
- 833 75. R. J. Simes, An improved bonferroni procedure for multiple tests of significance. *Biometrika*. **73**
834 (1986), doi:10.1093/biomet/73.3.751.
- 835 76. R. D. Finn, A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K.
836 Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, M. Punta, Pfam: The protein
837 families database. *Nucleic Acids Res*. **42** (2014), , doi:10.1093/nar/gkt1223.
- 838 77. S. Carbon, E. Douglass, B. M. Good, D. R. Unni, N. L. Harris, C. J. Mungall, S. Basu, R. L. Chisholm,
839 R. J. Dodson, E. Hartline, P. Fey, P. D. Thomas, L. P. Albou, D. Ebert, M. J. Kesling, H. Mi, A.
840 Muruganujan, X. Huang, T. Mushayahama, S. A. LaBonte, D. A. Siegele, G. Antonazzo, H. Attrill, N.
841 H. Brown, P. Garapati, S. J. Marygold, V. Trovisco, G. dos Santos, K. Falls, C. Tabone, P. Zhou, J. L.
842 Goodman, V. B. Strelets, J. Thurmond, P. Garmiri, R. Ishtiaq, M. Rodríguez-López, M. L. Acencio,
843 M. Kuiper, A. Lægreid, C. Logie, R. C. Lovering, B. Kramarz, S. C. C. Saverimuttu, S. M. Pinheiro, H.
844 Gunn, R. Su, K. E. Thurlow, M. Chibucos, M. Giglio, S. Nadendla, J. Munro, R. Jackson, M. J.
845 Duesbury, N. Del-Toro, B. H. M. Meldal, K. Paneerselvam, L. Perfetto, P. Porras, S. Orchard, A.
846 Shrivastava, H. Y. Chang, R. D. Finn, A. L. Mitchell, N. D. Rawlings, L. Richardson, A. Sangrador-
847 Vegas, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. M. Sitnikov, M. A.
848 Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bähler, E. R. Bolton, J. L. de Pons, M. R.
849 Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, C. Plasterer, M. A. Tutaj,
850 M. Vedi, S. J. Wang, P. D’Eustachio, L. Matthews, J. P. Balhoff, S. A. Aleksander, M. J. Alexander, J.
851 M. Cherry, S. R. Engel, F. Gondwe, K. Karra, S. R. Miyasato, R. S. Nash, M. Simison, M. S. Skrzypek,
852 S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, A. Morgat, E. Bakker, T. Z. Berardini, L. Reiser, S.
853 Subramaniam, E. Huala, C. N. Arighi, A. Auchincloss, K. Axelsen, G. Argoud-Puy, A. Bateman, M. C.
854 Blatter, E. Boutet, E. Bowler, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C. C. Casas, E. Coudert,
855 P. Denny, A. Es-Treicher, M. L. Famiglietti, G. Georghiou, A. N. Gos, N. Gruaz-Gumowski, E.
856 Hatton-Ellis, C. Hulo, A. Ignatchenko, F. Jungo, K. Laiho, P. le Mercier, D. Lieberherr, A. Lock, Y.
857 Lussi, A. MacDougall, M. Ma-Grane, M. J. Martin, P. Masson, D. A. Natale, N. Hyka-Nouspikel, S.
858 Orchard, I. Pedruzzi, L. Pourcel, S. Poux, S. Pundir, C. Rivoire, E. Speretta, S. Sundaram, N. Tyagi,
859 K. Warner, R. Zaru, C. H. Wu, A. D. Diehl, J. N. Chan, C. Grove, R. Y. N. Lee, H. M. Muller, D. Raciti,
860 K. van Auken, P. W. Sternberg, M. Berriman, M. Paulini, K. Howe, S. Gao, A. Wright, L. Stein, D. G.
861 Howe, S. Toro, M. Westerfield, P. Jaiswal, L. Cooper, J. Elser, The Gene Ontology resource:
862 Enriching a GOld mine. *Nucleic Acids Res*. **49** (2021), doi:10.1093/nar/gkaa1113.
- 863 78. Student, The Probable Error of a Mean. *Biometrika*. **6**, 1–25 (1908).
- 864 79. E. T. Jaynes, *Probability Theory* (Cambridge University Press, 2003;
865 <https://www.cambridge.org/core/product/identifier/9780511790423/type/book>).

- 866 80. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful
867 Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*.
868 **57** (1995), doi:10.1111/j.2517-6161.1995.tb02031.x.
- 869 81. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets are
870 compositional: And this is not optional. *Front Microbiol.* **8** (2017), ,
871 doi:10.3389/fmicb.2017.02224.
- 872 82. I. Erb, C. Notredame, How should we measure proportionality on relative gene expression data?
873 *Theory in Biosciences.* **135** (2016), doi:10.1007/s12064-015-0220-8.
- 874 83. B. Efron, R. Tibshirani, J. D. Storey, V. Tusher, Empirical bayes analysis of a microarray
875 experiment. *J Am Stat Assoc.* **96** (2001), doi:10.1198/016214501753382129.
- 876 84. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing
877 radiation response. *Proc Natl Acad Sci U S A.* **98** (2001), doi:10.1073/pnas.091062498.
- 878 85. I. Rivals, L. Personnaz, L. Taing, M. C. Potier, Enrichment or depletion of a GO category within a
879 class of genes: Which test? *Bioinformatics.* **23** (2007), , doi:10.1093/bioinformatics/btl633.
- 880 86. A. Brionne, A. Juanchich, C. Hennequet-Antier, ViSEAGO: A Bioconductor package for clustering
881 biological functions using Gene Ontology and semantic similarity. *BioData Min.* **12** (2019),
882 doi:10.1186/s13040-019-0204-1.
- 883 87. T. Slimani, Description and Evaluation of Semantic Similarity Measures Approaches. *Int J Comput*
884 *Appl.* **80** (2013), doi:10.5120/13897-1851.
- 885 88. G. Wang, D. H. Oh, M. Dassanayake, GOMCL: A toolkit to cluster, evaluate, and extract non-
886 redundant associations of Gene Ontology-based functions. *BMC Bioinformatics.* **21** (2020),
887 doi:10.1186/s12859-020-3447-4.
- 888 89. V. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet*
889 *Physics Doklady.* **10** (1966).
- 890 90. R. R. Sokal, A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull.* **38**
891 (1958).
- 892 91. [https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/ocean-](https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/ocean-circulation/arctic-ocean-circulation/)
893 [circulation/arctic-ocean-circulation/](https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/ocean-circulation/arctic-ocean-circulation/).

894
895



896

897 **Figure 1 | Lagrangian connectivity of the North Atlantic Ocean and Arctic Ocean in the *Tara***

898 **Oceans and *Tara* Oceans Polar Circle expeditions and polar front at the time of sampling.**

899 **(a) *Tara* Oceans and *Tara* Oceans Polar Circle stations considered in this study and shortest**

900 **Lagrangian path (minimum travel time T_{min}) between stations (starting at Station_143). (b) Mean**

901 **Sea Surface Temperature in June 2013 around the time of sampling of stations 155 to 168**

902 **(between May and July). The orange arrow indicates approximately the polar front between**

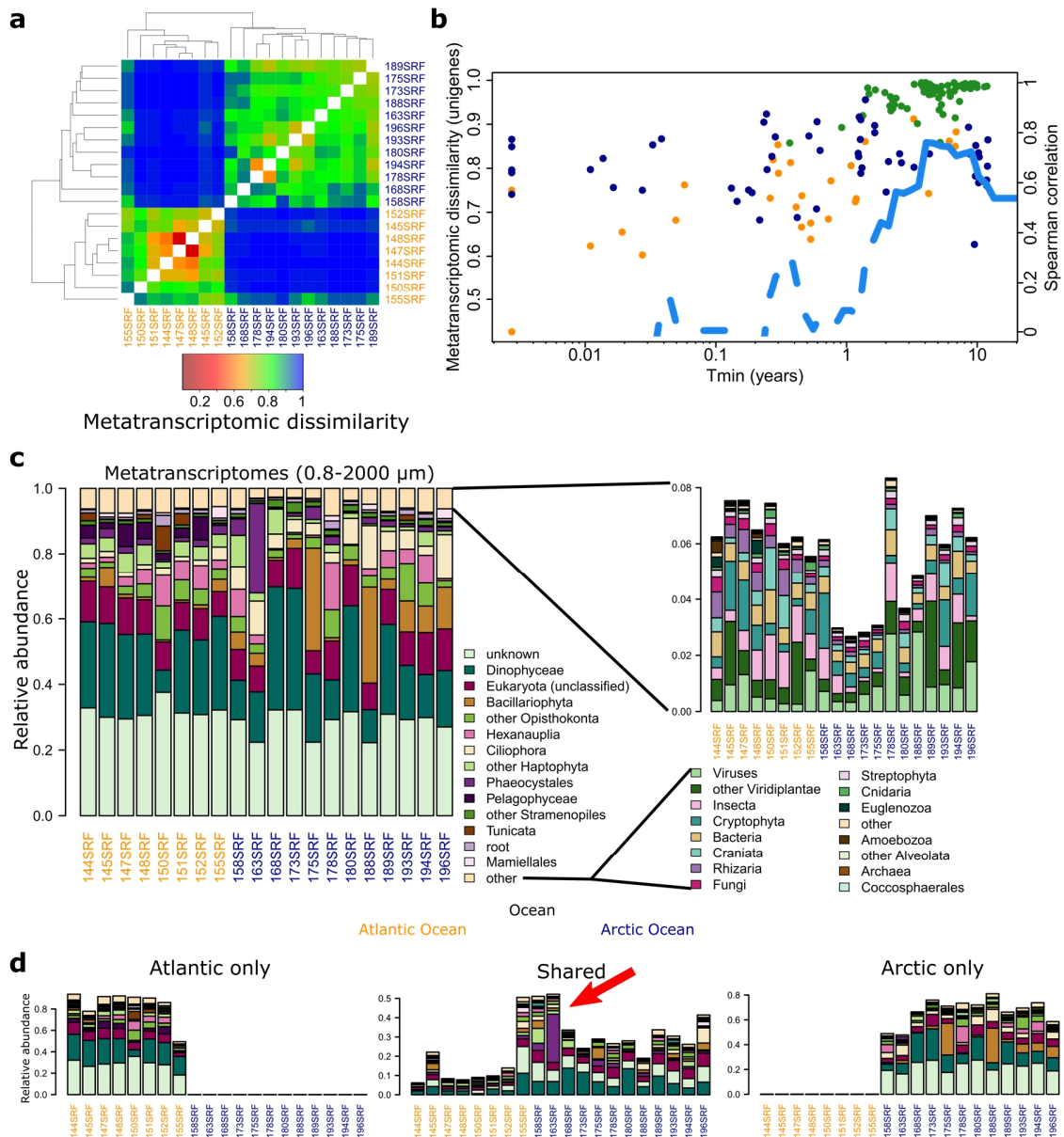
903 **Stations 158 and 163 where a steep temperature gradient was observed (black lines). Main inflow**

904 **and outflow currents are represented, respectively, by red arrows (warm Atlantic inflows) and blue**

905 **arrows (cold Arctic outflows). They are adapted from the Woods Hole Oceanographic Institution**

906 **website (91). The orange arrow indicates the polar front.**

907

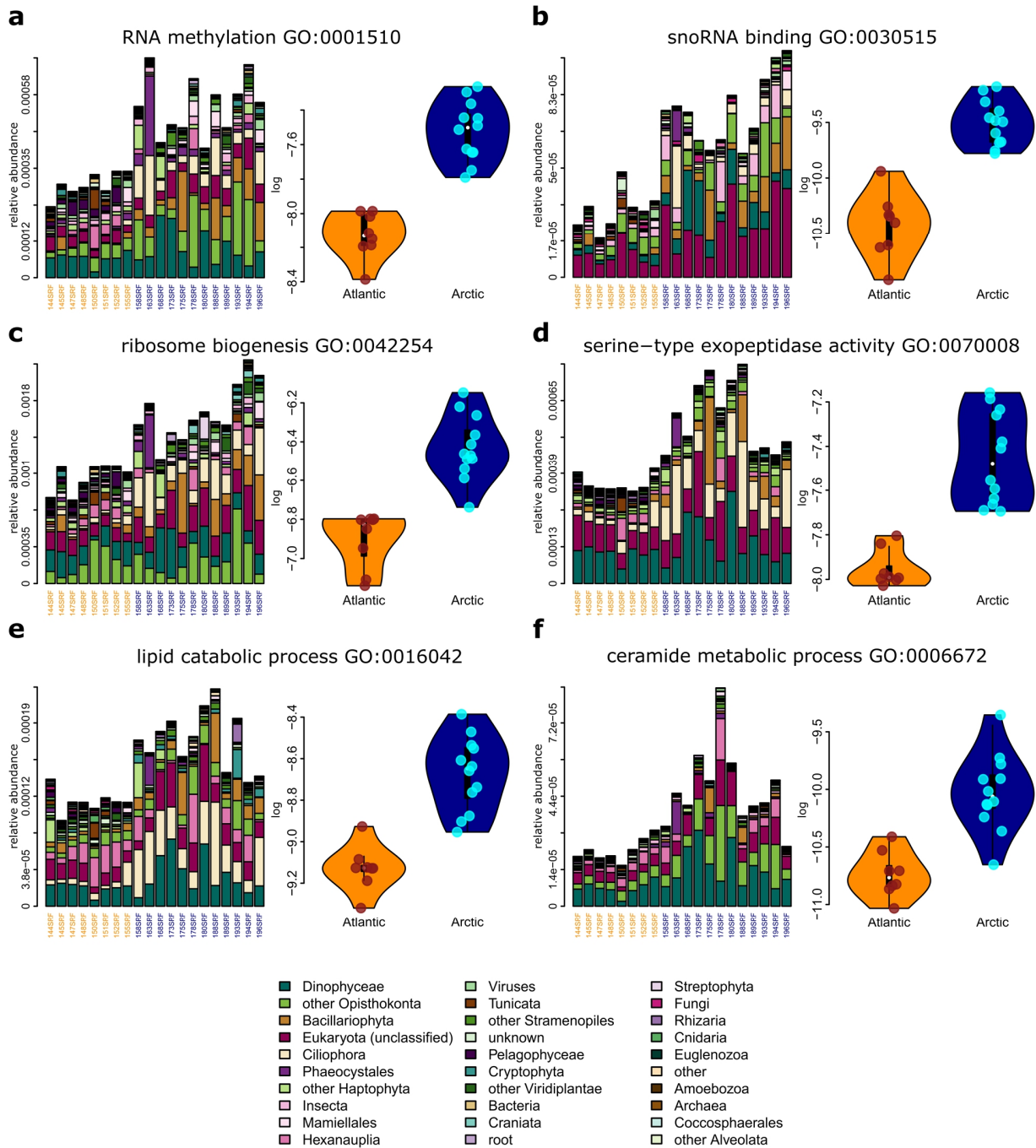


908

909 **Figure 2 | Metatranscriptomic and taxonomical composition connectivity with respect to**
 910 **ocean currents of surface ocean plankton (size fraction 0.8-2000 μm).** (a) Matrix of
 911 metatranscriptomic pairwise β diversity estimates based on unigene relative abundances (Bray-
 912 Curtis dissimilarity) clustered using the UPGMA (90) algorithm. (b) Metatranscriptomic β
 913 diversity as a function of Lagrangian connectivity between pairs of *Tara* Oceans and *Tara* Oceans
 914 Polar Circle stations (minimum transport time) and cumulative Spearman coefficient
 915 between the two metrics (blue line). Orange, dark blue and green points respectively indicate
 916 Atlantic, Arctic and Atlantic-Arctic pairs of stations. (c) Left: global metatranscriptomic taxonomical
 917 composition based on unigene relative abundances of major plankton taxa (taxonomic level can

28

918 vary) in the North Atlantic and Arctic Oceans. Right: Zoom on the 16 least abundant taxa (other).
919 **(d)** Decomposition between basin-specific unigenes and shared unigenes between the two basins
920 (classified based on presence or absence of expression). A clear transition enriched in shared
921 genes appears in Stations 153, 158 and 163.
922



923

924

925

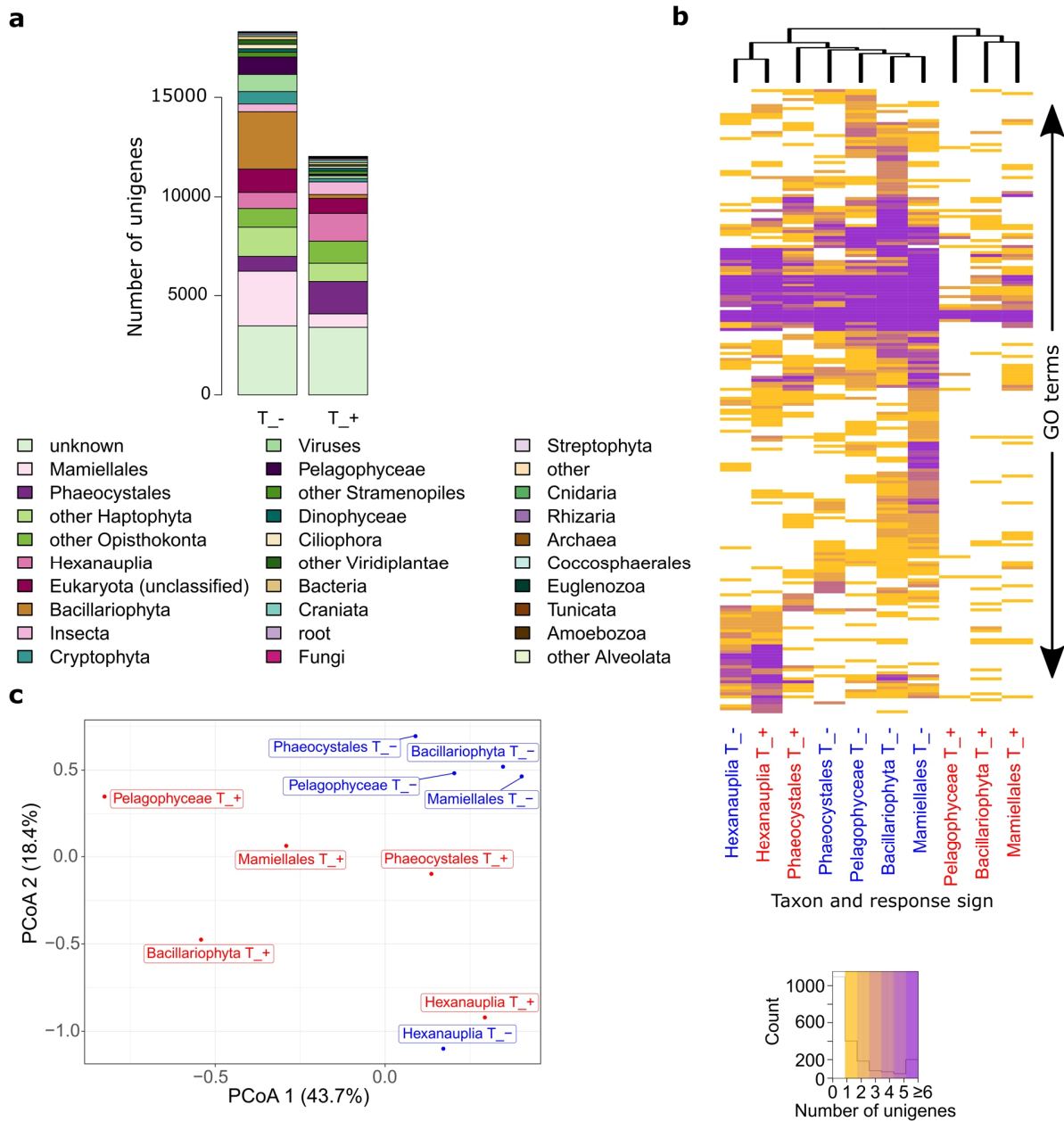
926

927

928

929

Figure 3 | Examples of Gene Ontology (GO) terms that are more abundant in Arctic metatranscriptomes. Each panel (a-f) represents (left) the sum of the metatranscriptomic relative abundances of unigenes associated with the abovementioned GO terms across considered taxa and (right) the distribution of this relative abundance in the Atlantic (dark blue) versus Arctic basin (purple). Each of these GO terms is significantly more abundant in the Arctic basin.



930

931

932

933

934

935

936

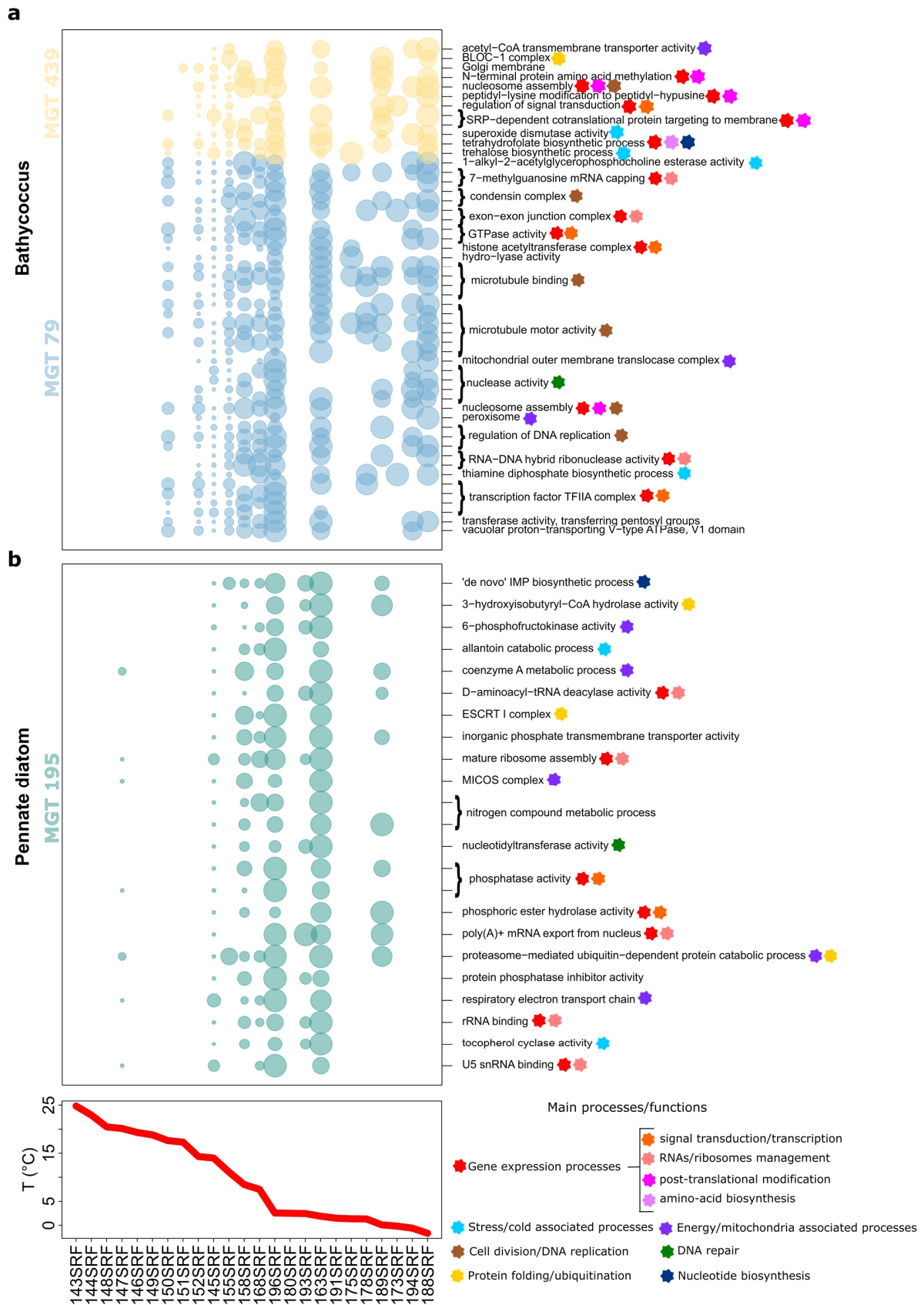
937

938

Figure 4 | Functional response to temperature at the unigene expression level for different taxa at the NAO/AO transition. (a) Taxonomic decomposition of unigenes significantly correlated with temperature, either negatively (left bar) or positively (right bar). (b) Heatmap of the number of unigenes and associated GO terms that correlate with temperature across main taxa. Each column represents the number of unigenes associated to a GO term (in line) for a given taxa and a correlation sign (blue: negative, red: positive). Columns are clustered based on Bray-Curtis dissimilarity index. (c) Bray-Curtis-based Principal Coordinate Analysis (PCoA) of the (b) panel. Noticeably the decrease in the functional response of the different phytoplankton groups

939 to temperature is similar while Hexanauplia (copepods) have a different functional landscape of
940 response to temperature changes.

941



942

33

187

943 **Figure 5 | Expression profiles and Gene Ontology annotation of enriched unigenes**
944 **negatively correlated with temperature for three algal Metagenomics-based**
945 **Transcriptomes.** Expression profiles (left) of unigenes whose Gene Ontology terms (right) are
946 enriched within unigenes significantly negatively correlated to temperature are displayed for two
947 algal taxa from three Metagenomics-based Transcriptomes (MGTs): **(a)** *Bathycoccus* (MGT 79
948 and 439) **(b)** a pennate diatom (MGT 195, annotated as a Cymatosiraceae). For this MGT
949 unigenes, unassigned functions (unknown) are enriched; however they are too numerous to be
950 represented here.

951

952

953

954

Supplement of

Changes in functional composition and gene expression in eukaryotic plankton at the Atlantic-Arctic Polar front

Frémont et al.

Correspondence to: Paul Frémont (pfremont@genoscope.cns.fr), Olivier Jaillon (ojillon@genoscope.cns.fr)

Supplementary Discussion 1 | Environmental differences between the North Atlantic Ocean (NAO) and the Arctic Ocean (AO) and link with current connectivity.

Using *in situ* environmental parameters from the *Tara* schooner, the North Atlantic Ocean and the Arctic Ocean are clearly distinguishable (Supplementary Figure 1a). The North Atlantic is warmer, saltier and has a stronger temperature seasonal variability while the Arctic Ocean is richer in nutrients, has longer daylight duration (in the considered seasons) and a higher nutrient seasonality (Supplementary Figure 1a). Stations 191 and 193 in the Arctic appear as outliers. These stations are more coastal and we note that stations 191 to 196 are influenced by Pacific waters (1). Correlating environmental distance and minimum travel time between stations, a maximum of correlation is found between 4 and 8 years of transport (Supplementary Figure 1b) and although the transition seem continuous environmentally (blue points to red points), high variability is also found within Arctic samples (green points and high values red points).

Supplementary Discussion 2 | Difference in taxonomic composition of the NAO/AO metatranscriptomes and metagenomes

We explore the differences in taxonomic composition of North Atlantic and Arctic Oceans metatranscriptomes and metagenomes. Composition of the metatranscriptomes (Figure 2 b-d) differs significantly from the metagenomes composition (Supplementary Figure 3-4). Strikingly, while Dinophyceae is the 1st most abundant taxa (after unknowns) in the metatranscriptomes, it represents a much lesser fraction of the metagenomes (10th most abundant taxa, 6th in terms of number of unigenes). In the metatranscriptomes, Ciliophora and Other Stramenopiles are significantly more relatively abundant in the Arctic than in the Atlantic while Pelagophyceae, Rhizaria, Amoebozoa, Other Alveolata, Archaea and Cocosphaerales are significantly more abundant in the Atlantic (Wilcoxon test, Hommel correction $p < 0.05$, Supplementary Figure 5). In metagenomes, Ciliophora is significantly relatively more abundant in the Arctic while Pelagophyceae, Amoebozoa and 'other' are significantly relatively more abundant in the Atlantic (Wilcoxon test, Hommel correction $p < 0.05$, Supplementary Figure 5). Importantly, a large part of metagenomes unigenes are not necessarily expressed in the metatranscriptomes representing a continuously decreasing part from western to eastern parts (from ~30% of the relative abundance in station 144 to ~10% in station 155) of the North Atlantic and then diminishing in the Arctic (Supplementary Figure 4). This might reflect the presence of dead cells in the Atlantic winter or at least less active cells. Indeed this North Atlantic temperate transect was sampled from January to March so one might expect an increasing activity of phytoplankton population as the spring bloom is approaching.

Supplementary Discussion 3 | Other functions more abundant in the Arctic metatranscriptomes.

Other functions more abundant in the Arctic metatranscriptomes are directly linked to lipid metabolism ('lipid catabolic process' (Figure 3e), 'fatty-acyl-CoA binding'), to organization of organelles membranes ('intrinsic component of organelle membrane' and 'integral component of organelle membrane') and membrane fluidity ('Ceramide metabolic process' and 'Sphingolipid metabolic process', also more abundant in metagenomes). The important abundance of functions linked to lipid catabolic process most likely reflect the need for membrane remodeling and free fatty acid removal in cold environments (2, 3). Some enzymes involved in lipid catabolism have been reported to be involved in freezing tolerance in *A. thaliana* (4, 5) and in response to cold (6) and N stress in algae (3). The regulation of membrane fluidity is directly linked to cold environment acclimation as for example shown in the chloroplasts of the green algae *Nannochloropsis* (7) and *Chlamydomonas reinhardtii* (8) and more generally in all plants especially with an important role of sphingolipids (9) and particularly ceramides (10). We also find functions linked to protein trafficking such as 'protein transmembrane transport' and 'mitochondrial inner membrane presequence translocase complex' more abundant in the Arctic. In contrast, functions linked to protein retention in their cellular compartments like 'maintenance of location', 'maintenance of protein localization in organelle' and 'protein retention in ER lumen' are found more abundant in North Atlantic metagenomes suggesting a less rapid turnover of proteins in winter Atlantic waters (Supplementary Table 1, Supplementary Discussion 2). Finally, functions linked to telomere maintenance such as 'telomere maintenance' and 'telomeric DNA binding' are also found more abundant in the Arctic metatranscriptomes which might reflect the need to maintain the integrity of chromosomes in extreme environments (11, 12).

Supplementary Discussion 4 | PFAMs more abundant in the Arctic Ocean in link with cold stress acclimation

We investigate the different functioning of the spring/summer Arctic Ocean and the winter Atlantic Ocean by analyzing the differential abundances of a set of 22 PFAMs reported to be important in cold/hot stress response. We use the ALDEx2 R package (13) to characterize more abundant and overexpressed PFAMs between the two basins. The relevance of each PFAM to cold/hot stress acclimation is summarized in Supplementary Table 2. First, we visualize the inter-basin dispersion in function of the intra-basin dispersion of metagenomic and metatranscriptomic abundances and expression values fitted by the ALDEx2 package (Supplementary Figure 8). On a total of 11950 PFAMs, we find 1214 (respectively 758) and 4461 (respectively 3657) PFAMs significantly more abundant (Benjamini-Hochberg corrected Welch test, $p < 0.05$) respectively in

the North Atlantic and the Arctic Ocean metagenomes (respectively metatranscriptomes). Most more abundant PFAMs in the North Atlantic have a high maximum within basin dispersion while PFAMs more abundant in the Arctic have lower maximum within basin dispersion. This observation might reflect the higher variability of the Arctic basin (Figure 2 and Supplementary Figure 3). Zooming on specific taxa (Supplementary Figure 8), we often find a bimodal pattern of within condition dispersion both in more abundant PFAMs in the Atlantic and Arctic. We formulate three hypotheses to explain these patterns. First, some specific lineages/species whose genomes are enriched in some functions probably differentiate the two basins. Second, some specific lineages enriched in specific functions might be sampled only in few stations (e.g. Diatoms which are particularly abundant in few arctic stations). Third, some functions might be highly overexpressed. On the other hand, expression dispersion is comparatively lower compared to abundances (Supplementary Figure 8) and only a few PFAMs are found to be highly overexpressed. We find relatively a lot of PFAMs highly overexpressed especially in the Arctic mainly in Dinoflagellates (Supplementary Figure 8f) suggesting an important role of expression regulation in these organisms. These results needs however to be taken with caution as here we normalize the metatranscriptomic abundances by metagenomic abundances that are not necessarily from the same unigenes between metagenomes and metatranscriptomes especially in Dinoflagellates (high genomic complexity of this group).

Regarding cold acclimation functions, we find that nearly all of them are significantly more abundant in the Arctic Ocean in both metagenomes and metatranscriptomes but not overexpressed (Supplementary Figure 9a). Strikingly this signal is the strongest when considering the overall abundance of these PFAMs (*i.e.* within the whole community) while specific taxons display various patterns (Supplementary Figure 9a-g). Particularly, these PFAMs tend to be more abundant in Mamiellales, Bacillariophyta, Pelagophyceae, Dinophyceae and Unknown while not necessarily in Phaeocystales and Ciliophora (which are more abundant in the Arctic). Noticeably, some of these PFAMs are overexpressed in Dinophyceae (Supplementary Figure 9f). Among the PFAMs we consider, the ice binding domain (PF11999) is largely the most excessively more abundant PFAM both at the community level and for specific taxons ($\sim 2^{15}$ higher abundance in the Arctic) while other PFAMs are more abundant at a lesser extent ($\sim 2^3$ to 5 higher abundance in the Arctic) (Supplementary Figure 9). Overall, these results might reflect the different strategies of the different plankton organisms to adapt and acclimate to an extreme environment such as the Arctic Ocean.

Supplementary Discussion 5 | Functions more abundant in Atlantic metagenomes and metatranscriptomes

We find 85 and 11 more abundant Gene Ontology (GO) terms (3 of which are common) respectively in Atlantic metagenomes and metatranscriptomes in comparison to the Arctic in size fraction 0.8-2000 μm (Supplementary Figure 10 and exhaustive lists in Supplementary Table 1). The first functions that jump to the eye when looking at more abundant functions in Atlantic metagenomes are central carbon metabolism functions. Indeed a few enzymes and reactions from the glycolysis are found more abundant in the Atlantic such as: "glycerol-3-phosphate (G3P) dehydrogenase [NAD⁺] activity" and "phosphoglycerate kinase activity". In addition "pentose-phosphate shunt, non-oxidative branch" (Supplementary Figure 10b) which is an alternative to glycolysis is also a more abundant GO terms in Atlantic metagenomes. The pentose phosphate shunt oxidative branch is known to reduce oxidative stress and support lipobiogenesis (14) while the non-oxidative branch ends to the beginning and middle of the glycolysis (generating a G3P and a fructose 6 phosphate molecule) and generates 5-carbon sugars used for the synthesis of nucleotides, nucleic acids and amino acids (15, 16). The term "Fatty acid biosynthetic process" and other functions linked to fatty acid biosynthesis like "acetyl-CoA carboxylase complex" (which is biotin dependent for which 2 GO terms are found more abundant) are also found more abundant in Atlantic metagenomes although these terms are dominated by bacterial genes (Supplementary Figure 16c). Overall, the high abundance of central carbon metabolism function in the Atlantic might reflect the overall prevalence of heterotrophy in this ocean as expected (17) especially as it was mostly sampled in winter. In addition, "nitrogenous compound fermentation" (Supplementary Figure 16d) is also found more abundant in the Atlantic. Interestingly, their abundance is dominated by bacteria followed by unclassified eukaryotes (Supplementary Figure 10d) which might reflect the presence of unknown eukaryotes capable of fermentation. The high abundance of these terms as well as "anaerobic amino acid catabolic process" might reflect a higher abundance of organisms capable of living in anaerobic condition in the Atlantic (Figure 1b). Although Tara samples are from the surface, the presence of such organisms might be due to the winter sampling and might come from the deep mixing of oceanic waters in winter. More intriguingly, five GO terms linked to photosynthesis are found to be more abundant in Atlantic metagenomes and three of which are also more abundant in the metatranscriptomes: "photosynthetic electron transport chain", "photosynthetic electron transport in photosystem II", "electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity" (Supplementary Figure 10e), "photosynthesis light reaction" and "light-harvesting complex" (the two latter more abundant only in metagenomes). The high abundance of these functions in the Atlantic in winter is interesting as it suggests that algae are adapting to low light levels and saving energy. Indeed, the high abundance of photoreceptors (light reaction)

suggests the need to harvest light in a more efficient way. Second the cyclic electron transport path of photosynthesis is known to be used to balance the energy (18) and protect photosystems from photoinhibition (19). The cyclic electron flow (CEF) is also known to be involved in various acclimation responses in plants (20). We verify that these functions correspond to only a few PFAMs that are mostly more abundant (or not) in the Atlantic metagenomes and metatranscriptomes using ALDEx2 package (13, 21) (Supplementary Figure 18): PF00124 (Photosynthetic reaction centre Photo_RC), PF00421 (PSII, photosystem II), PF00556 (Light Harvesting Complex LHC), PF03967 (Photosynthetic reaction centre, H-chain N-terminal region PRCH), and PF11623 (NADH dehydrogenase subunit S NdhS). In fact most other PFAMs associated to photosynthesis appear overall more abundant in Arctic metagenomes which is expected although it depends on taxons (Supplementary Figure 18).

Among other functions that we find more abundant in Atlantic metagenomes or metatranscriptomes, we can cite functions linked to mobility of genomic elements such as “transposase activity” and “transposition, DNA-mediated” (more abundant in metatranscriptomes, dominated by unknown followed by bacteria) or bacteria specific functions (Supplementary Table 1).

Supplementary Discussion 6 | Details of enriched function in persisting species.

In MGT 79 (*Bathycoccus*, Figure 5a) a unigene involved in thiamine (Vitamin B1) diphosphate biosynthetic process is negatively correlated to temperature, known to be important in early stage response to cold and osmotic stress in *Arabidopsis* (22) and in the poplar (23). In MGT 195, a pennate diatom (Figure 5b), the expression profiles of a gene involved in tocopherol cyclase activity (also in MGT 42 *Aureococcus*, Supplementary Figure 16b) and one in allantoin catabolic process are significantly negatively correlated to temperature. Tocopherol cyclase activity is known to be important in cold (24) but also salt (25) and high light acclimation (26) in plants. Allantoin on the other hand is a metabolic intermediate of purine catabolism that often accumulates in stressed plants (27) and was reported to enhance cold stress tolerance in rice grains (28). In MGT 4 (Supplementary Figure 16c), superoxide dismutase (SOD) (also in MGT 439, Figure 5a), heat shock protein binding and glutamyl-tRNA reductase activity are reported to be enriched GO terms in negatively correlated unigenes. SOD is known to increase stress tolerance in potato through mitochondrial activity (29) but also in wheat (30) and barley (31). Some heat shock proteins are known to respond to cold stress in *Synechococcus* (32), some plants (33) but also fishes (34). Finally, Glutamyl-tRNA reductase, the key enzyme for 5-aminolevulinic acid (ALA) synthesis, has been shown to be highly expressed in cold stress response and to remain at relatively high levels after acclimation although it varies across plants (35). In MGT 439 (Figure

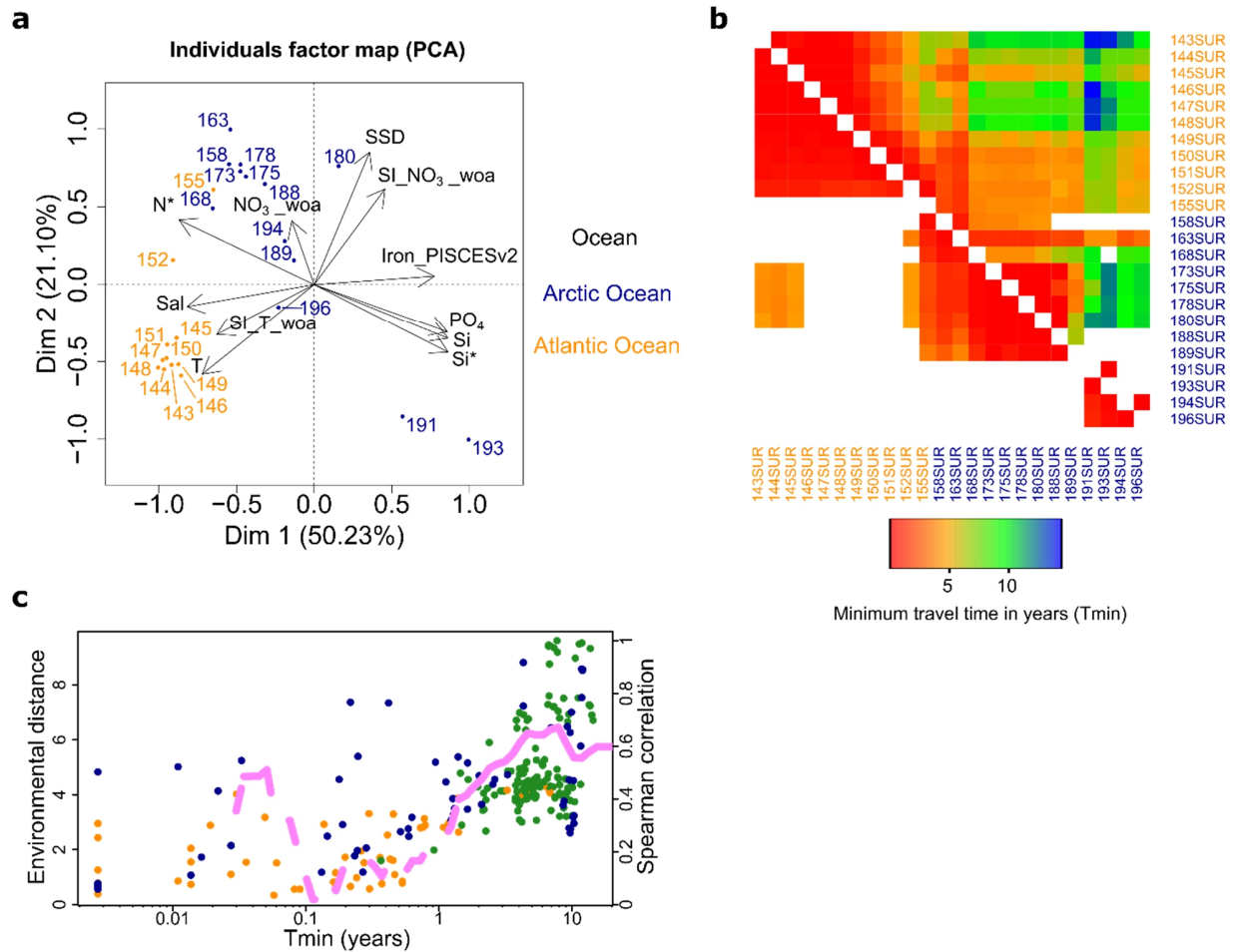
5a), trehalose biosynthetic process is also an enriched GO term in negatively correlated unigenes linked to cold acclimation (36). Finally, phospholipase A2 known to have a role in cold acclimation in wheat (37) is found to be enriched in MGT 76 (Supplementary Figure 16a). Zooming out at the higher taxonomy levels from Figure 2, other or same functions associated to cold stress are also found enriched such as tocopherol cyclase activity (Pelagophyceae) (24–26), polyketide metabolic process (Dinophyceae) (38, 39) or pantothenate biosynthesis (Pelagophyceae) (40) (Supplementary Figure 17). Overall, we find a wide variety of functions whose expressions are negatively correlated to temperature and known to be important in cold acclimation both in terrestrial plants and aquatic organisms highlighting the quasi universality and diversity of these functions for cold acclimation in the *Plantae* kingdom.

In addition to these functions directly linked to cold acclimation, many enriched functions are linked to gene expression processes and regulation, energy consumption and mitochondria related processes, division/DNA replication, motility, nucleotide biosynthesis, protein folding and DNA repair (Figure 5). In particular, GO terms linked to the triggering and control of gene expression such as ‘transcription factor TFIIA complex’ (MGT 79), ‘transcription by RNA polymerase II’ (MGT 76), ‘transcription corepressor activity’ (MGT 76) but also to post transcriptional modification and protein biosynthesis such as ‘histone acetyltransferase’ (MGT 79 and 76), ‘translation initiation factor activity’ (MGT 4), ‘poly(A)+ mRNA export from nucleus’ (MGT 195), ‘cellular amino acid biosynthetic process’ (MGT 4), are found to be enriched. This suggests that the entrance in the Arctic Ocean triggers a tightly genetically controlled transcriptional burst that allows the persistence of some organisms (mainly algae here) to the new physico-chemical conditions they encounter. Indeed, enriched functions include all levels of genetic expression control and processes including signal transduction (phosphorylation, GTPase, phosphatidylinositol phosphorylation...), transcription regulation, RNAs/ribosomes management, amino-acid biosynthesis and post-translation modifications (Figure 5, Supplementary Figure 16-17).

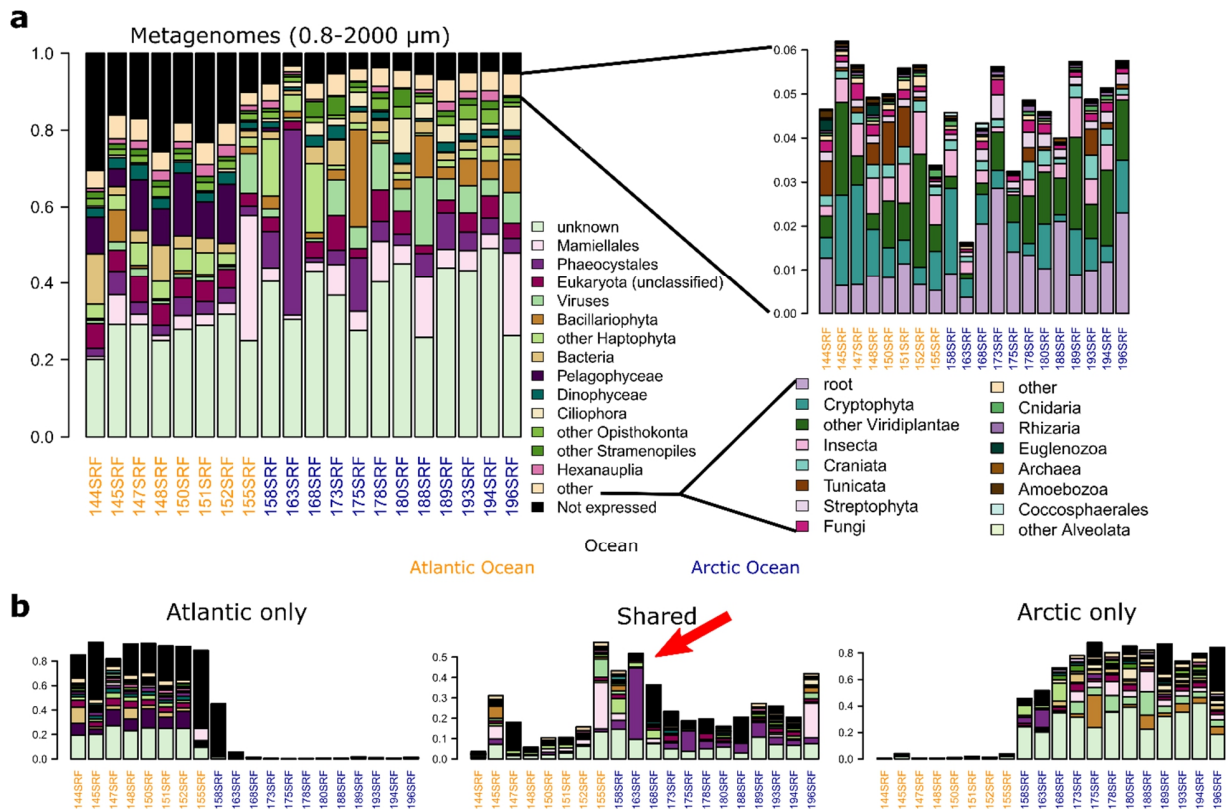
Overall many functions linked to energy and respiratory processes are found enriched in the 6 MGTs such as ‘respiratory electron transport chain complex’ and ‘proton transporting ATP synthase activity, rotational mechanism’. This effect might be due to an increase in metabolic activities due to the spring and summer season (2). These seasons are known to lead to loss of synchrony of cell division of microalgae (41, 42) in agreement with enriched functions in link with cell division and DNA replication.

Critically, many enriched GO terms are associated to DNA replication and consequently cell division such as ‘regulation of DNA replication’ (MGT 79 and Phaeocystales) and ‘DNA replication

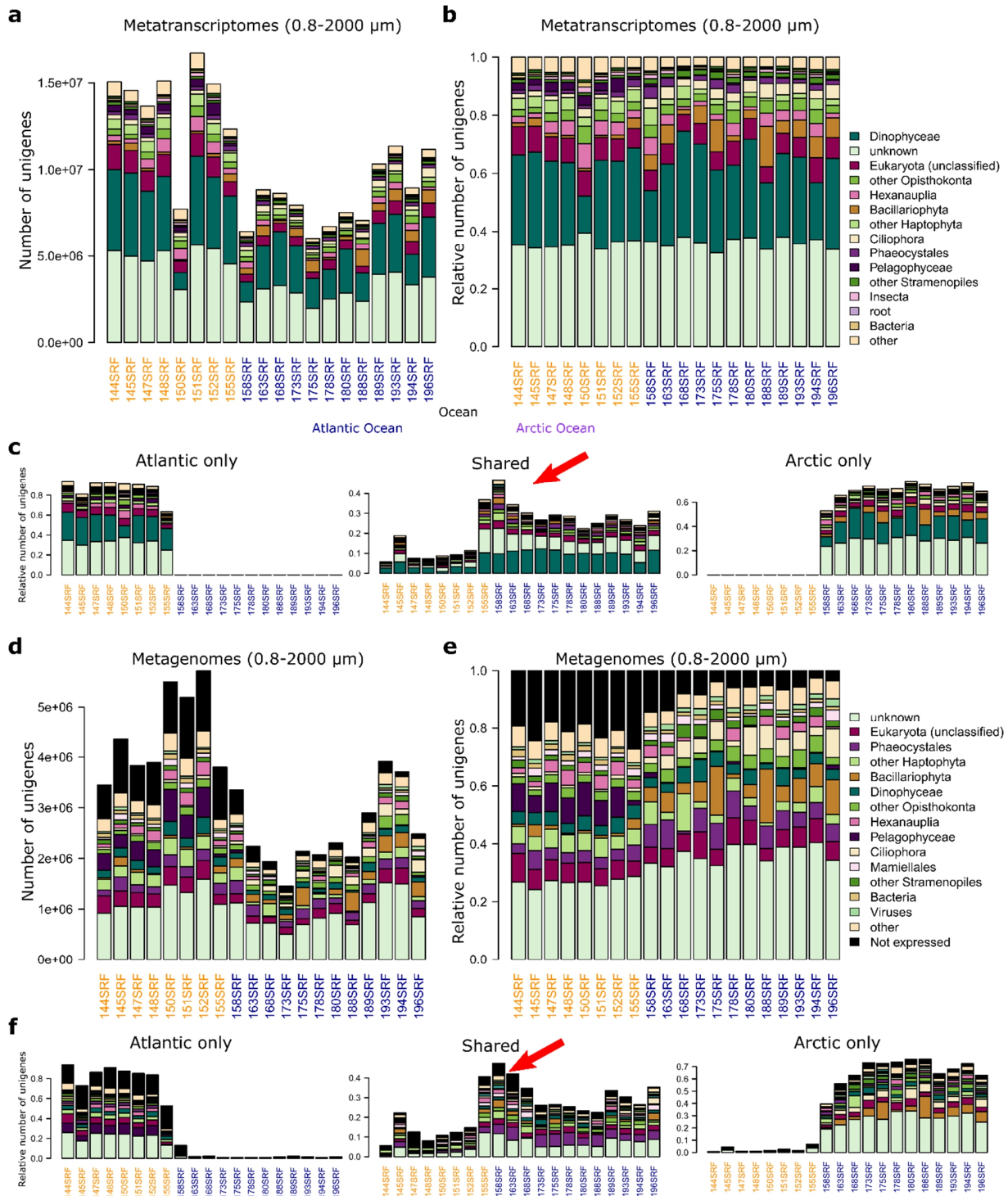
initiation' (MGT 76 and Mamiellales) or Ndc80 complex (Bacillariophyta). This might reflect the diminution in synchrony of cell division in link with the circadian clock (41, 42) as days are getting longer as cells enter the Arctic. In addition, it might show the capacity of persisting organism to keep on dividing as they enter in the polar waters.



Supplementary Figure 1 | Environmental differentiation of the NAO and AO basins in link with their Lagrangian connectivity by currents. (a) Principal Component Analysis (PCA) of the environmental parameters characterizing the stations. (b) Lagrangian minimum travel time connectivity matrix. The upper matrix represents the minimum travel time by currents from row stations to column stations and vice versa. (c) Environmental distance in function of Lagrangian minimum travel time (Tmin) and cumulative Spearman correlation (brown line).

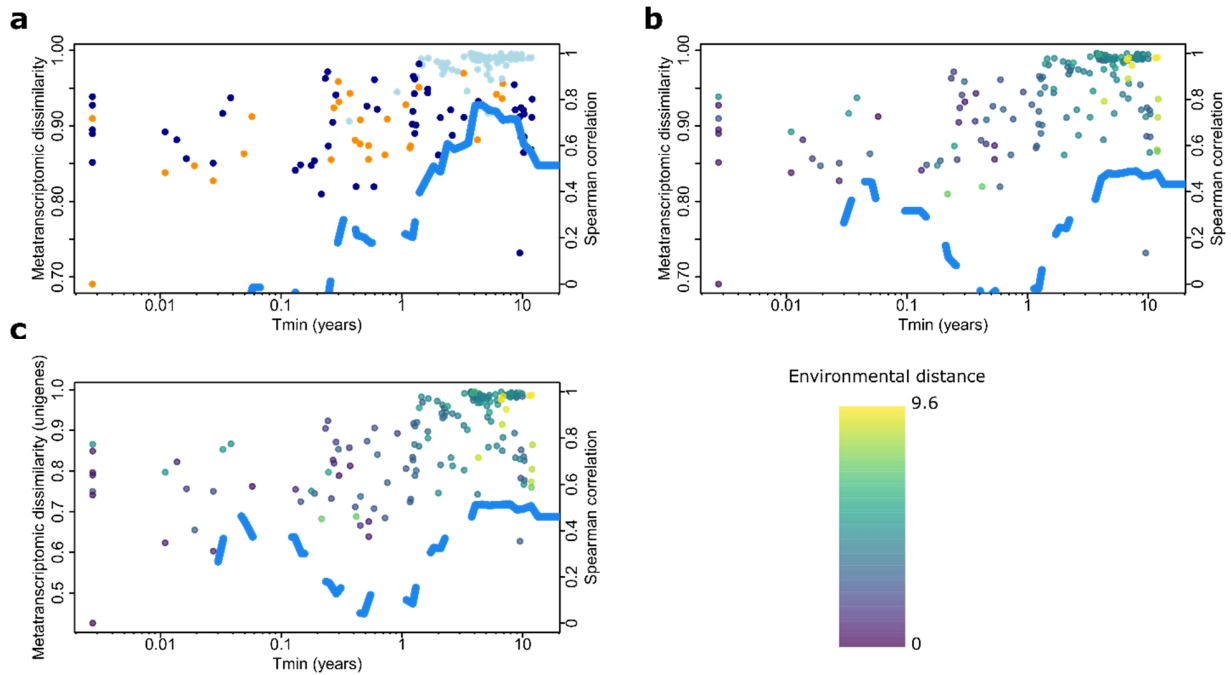


Supplementary Figure 2 | Taxonomic composition of surface ocean plankton metagenomes of size fraction 0.8-2000 μm . (a) Left: global metagenomic taxonomic composition based on expressed unigenes relative abundances of major plankton taxa (taxonomic level can vary) in the North Atlantic and Arctic Ocean. Right: Zoom on the 16 least abundant taxa (other). (b) Decomposition between basins specific unigenes and shared unigenes between the two basins (classified based on presence or absence of expression). A clear transition enriched in shared genes appears in station 153, 158 and 163.

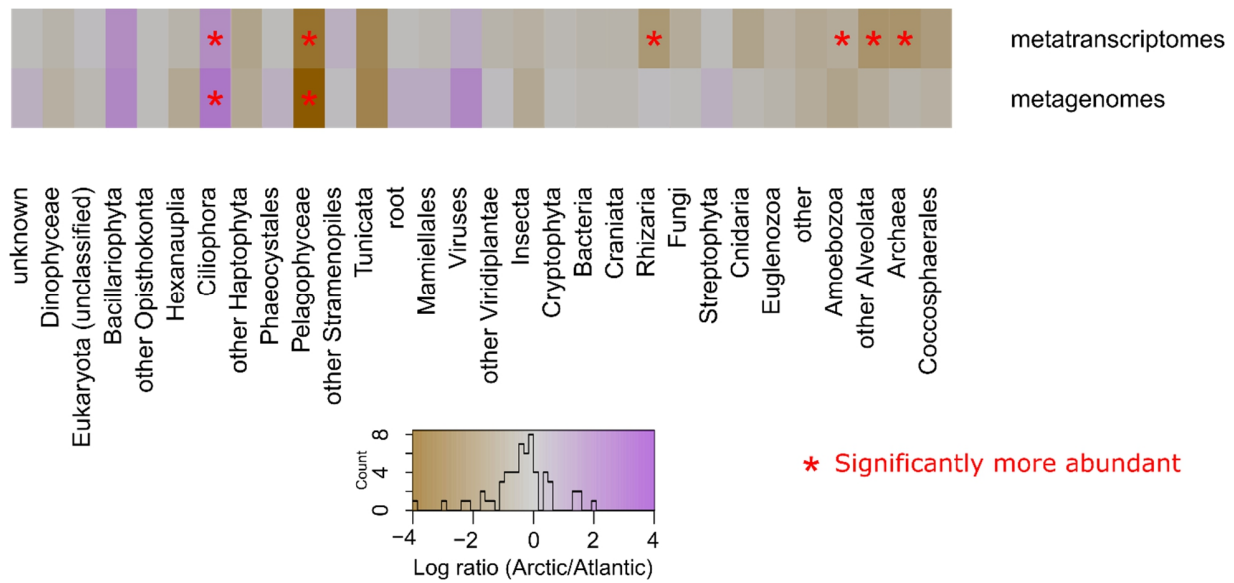


Supplementary Figure 3 | Taxonomic composition in unigenes of surface ocean plankton metatranscriptomes and metagenomes of size fraction 0.8-2000 μm . (a,d) Metatranscriptomic/metagenomic taxonomic composition in number of unigenes of North Atlantic and Arctic Ocean sampling stations. (b, e) Metatranscriptomic/metagenomic taxonomic

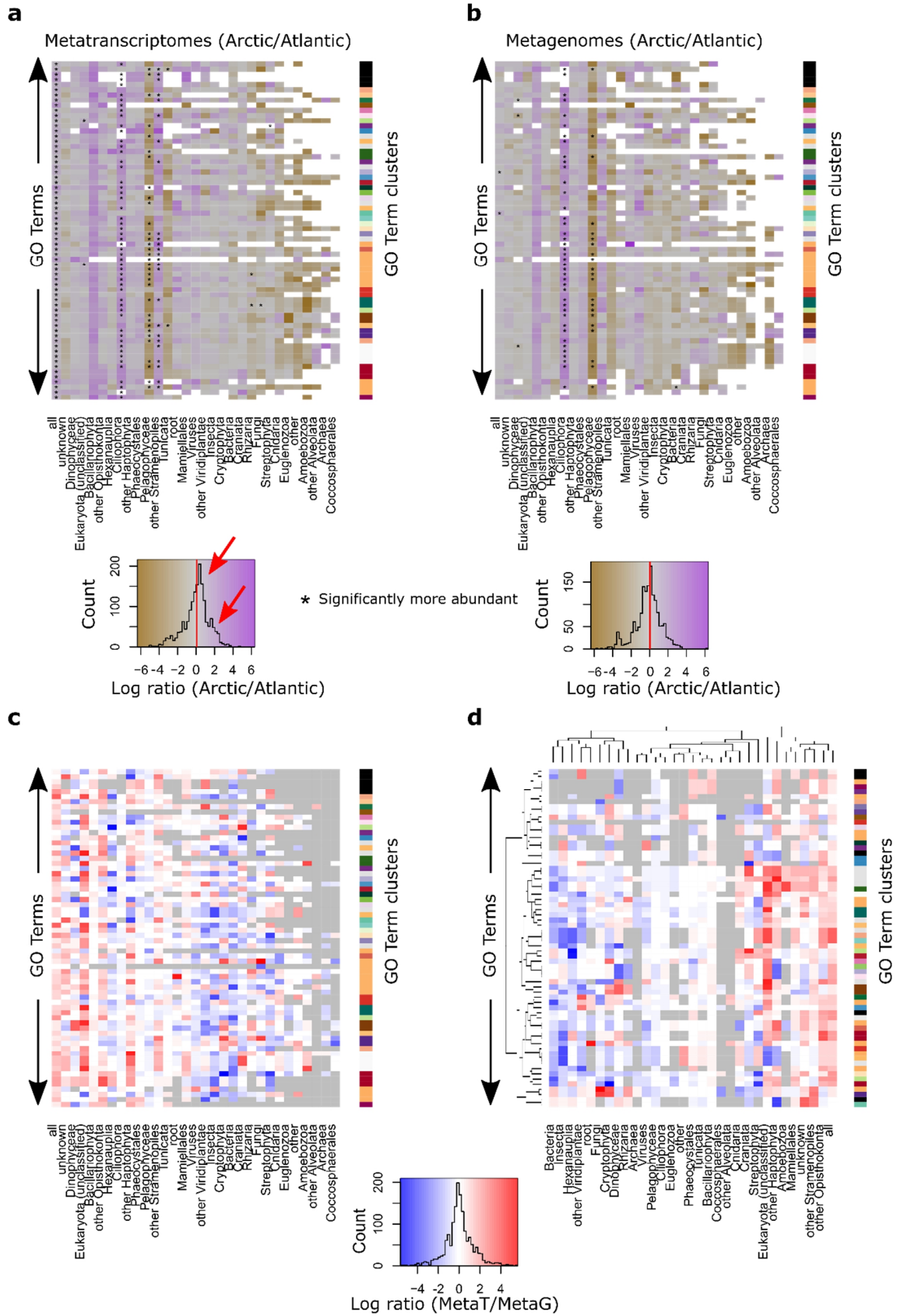
composition in relative number of unigenes of North Atlantic and Arctic Ocean sampling stations. **(c, f)** Decomposition of metatranscriptomes/metagenomes in relative number of unigenes between basins specific unigenes and shared unigenes between the two basins (classified based on presence or absence of expression).



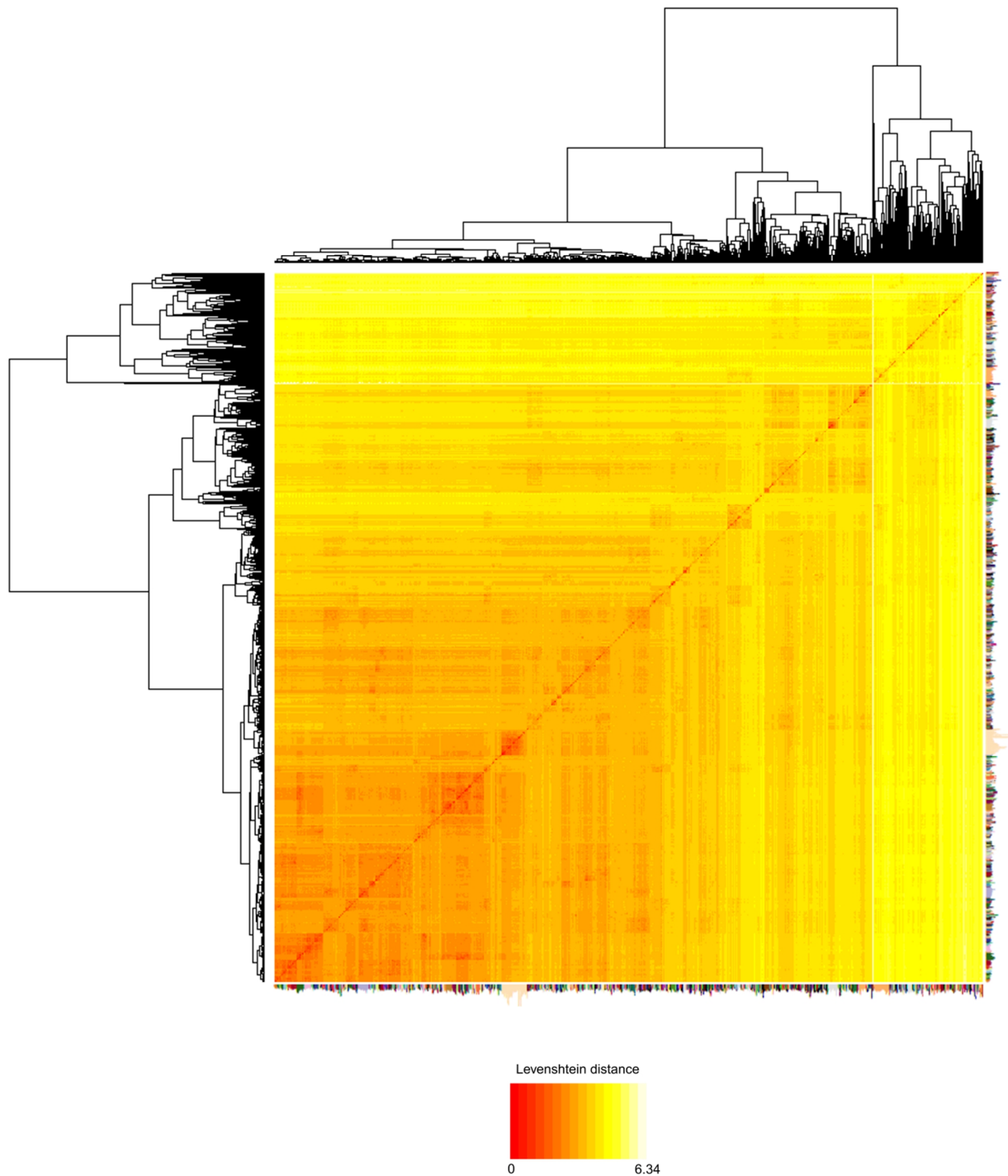
Supplementary Figure 4 | Metatranscriptomic β diversity in link with the Lagrangian and environmental connectivity of the NAO and AO basins. (a) Metatranscriptomic β diversity (SimkaMin (43)) in function of Lagrangian connectivity between pairs of Tara station (minimum transport time) and cumulative spearman correlation coefficient between the two metrics (blue line). Metatranscriptomic β diversity based on (b) SimkaMin (43) and (c) unigenes metatranscriptomic abundance-based Bray-Curtis (44) in function of Lagrangian connectivity between pairs of Tara station and cumulative spearman correlation coefficient between the environmental distance (points color) and β diversity (blue line).



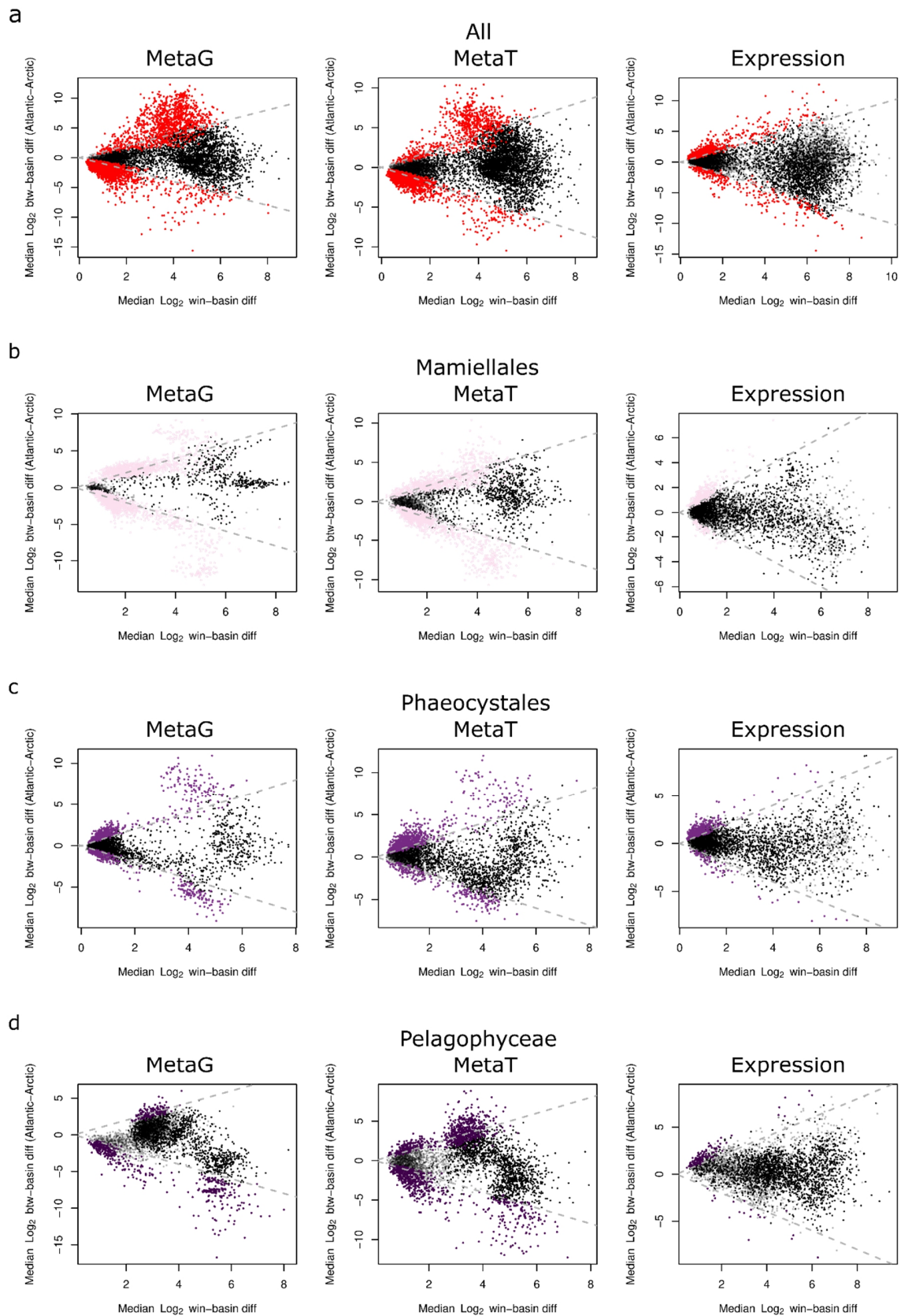
Supplementary Figure 5 | Median \log_e -fold change between Arctic Ocean and North Atlantic Ocean metatranscriptomes and metagenomes relative abundances of main plankton groups. The heatmap represents for each taxon the difference between their median logarithms of relative abundance between both basins (Arctic-Atlantic). Significantly more abundant taxa (Wilcoxon test, Hommel correction $p < 0.05$) either in metatranscriptomes (top row) or metagenomes (bottom row) are indicated with a red star. Here we use the logarithms of abundances and not the centered log ratio as the former might bias some abundance values as it is centered on zero in the case of skewed distributions (which is the case here).

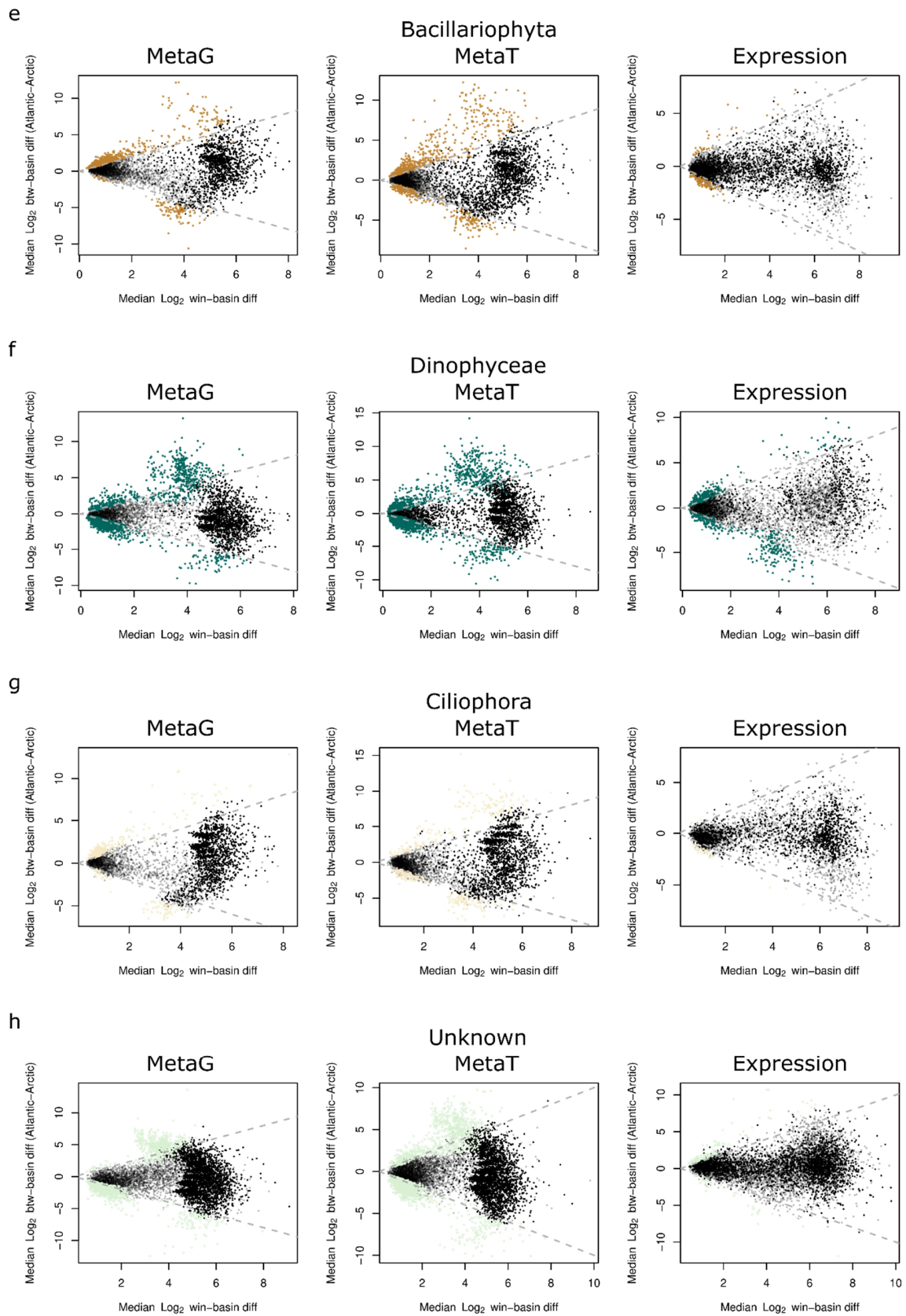


Supplementary Figure 6 | Median \log_e -fold change in metatranscriptomic and metagenomic relative abundances between Arctic Ocean and North Atlantic Ocean of significantly more abundant Gene Ontology (GO) terms in Arctic metatranscriptomes for all taxa and individual taxa and their ratios. (a,b) The heatmaps represent for all taxons (left columns of **a-c**) and each taxon in decreasing order of abundance (other columns) the difference between the median logarithms of relative abundance between both basins (Arctic-Atlantic) for Gene Ontology terms significantly more abundant in the Arctic metatranscriptomes. This ratio is plotted either for metatranscriptomes (**a**) or metagenomes (**b**). Significantly more abundant GO terms (Wilcoxon test, Hommel correction $p < 0.05$) are indicated with a black star. GO terms (rows) are colored and ordered according to the GO cluster they belong to, reflecting close functional annotation (see Supplementary Figure 7). The first 6 GO terms (6 first rows) don't belong to any cluster. (c) \log_e ratio of metatranscriptomes (**a**) and metagenomes (**b**) relative abundance ratio matrices and (d) hierarchical clustering of this matrix.



Supplementary Figure 7 | Clustering of the Gene Ontology (GO) terms based on the Levenshtein distance. Each GO term present in our unigene database was included in the clustering. The hierarchy of each GO term is converted in a string and pairwise Levenshtein distance⁽⁴⁵⁾ are computed. The resulting matrix is clustered using UPGMA⁽⁴⁶⁾ clustering (*Methods*). GO terms (lines and columns) are colored according to the cluster they belong to.

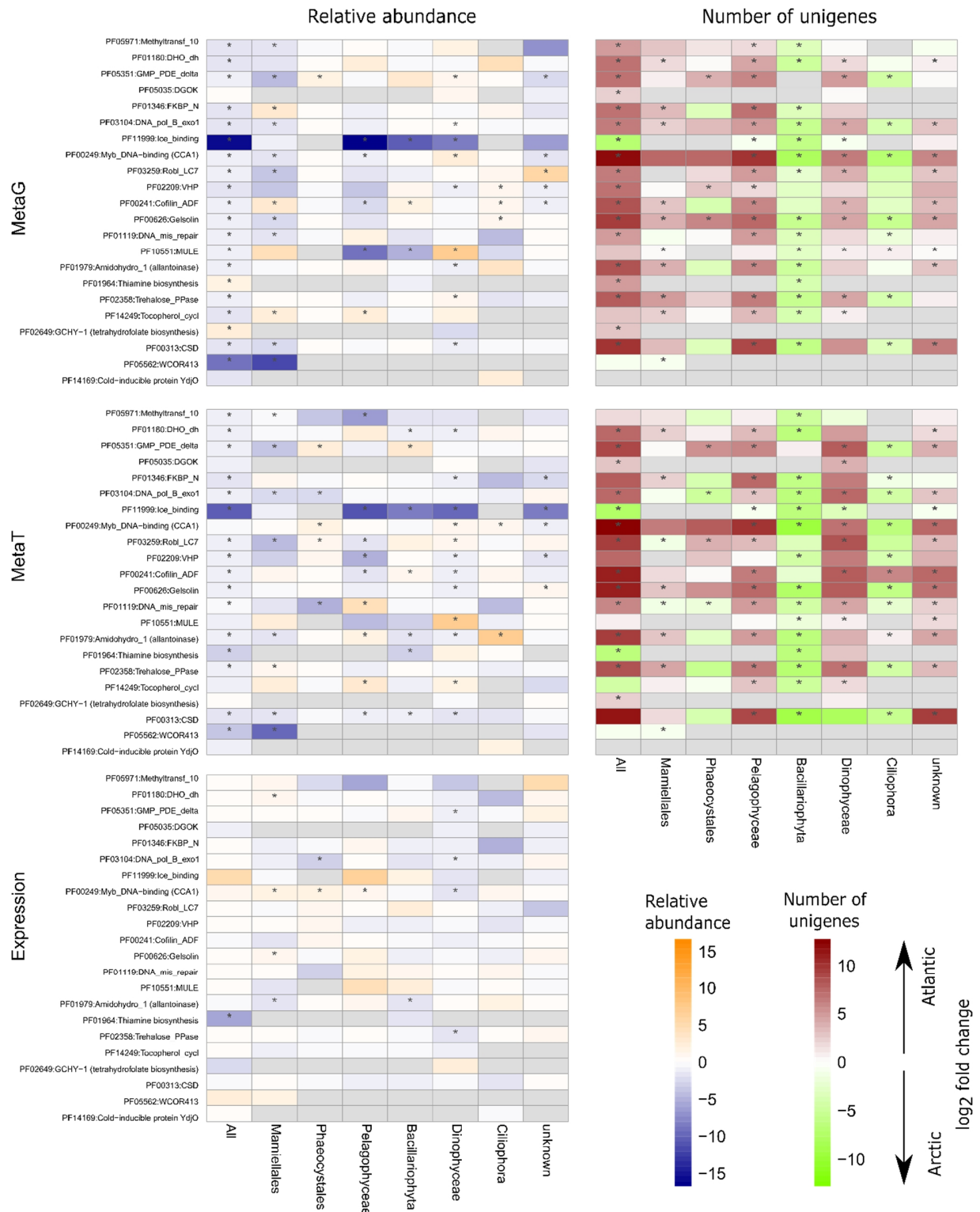




Supplementary Fig. 8 | Differential metagenomic, metatranscriptomic abundances and

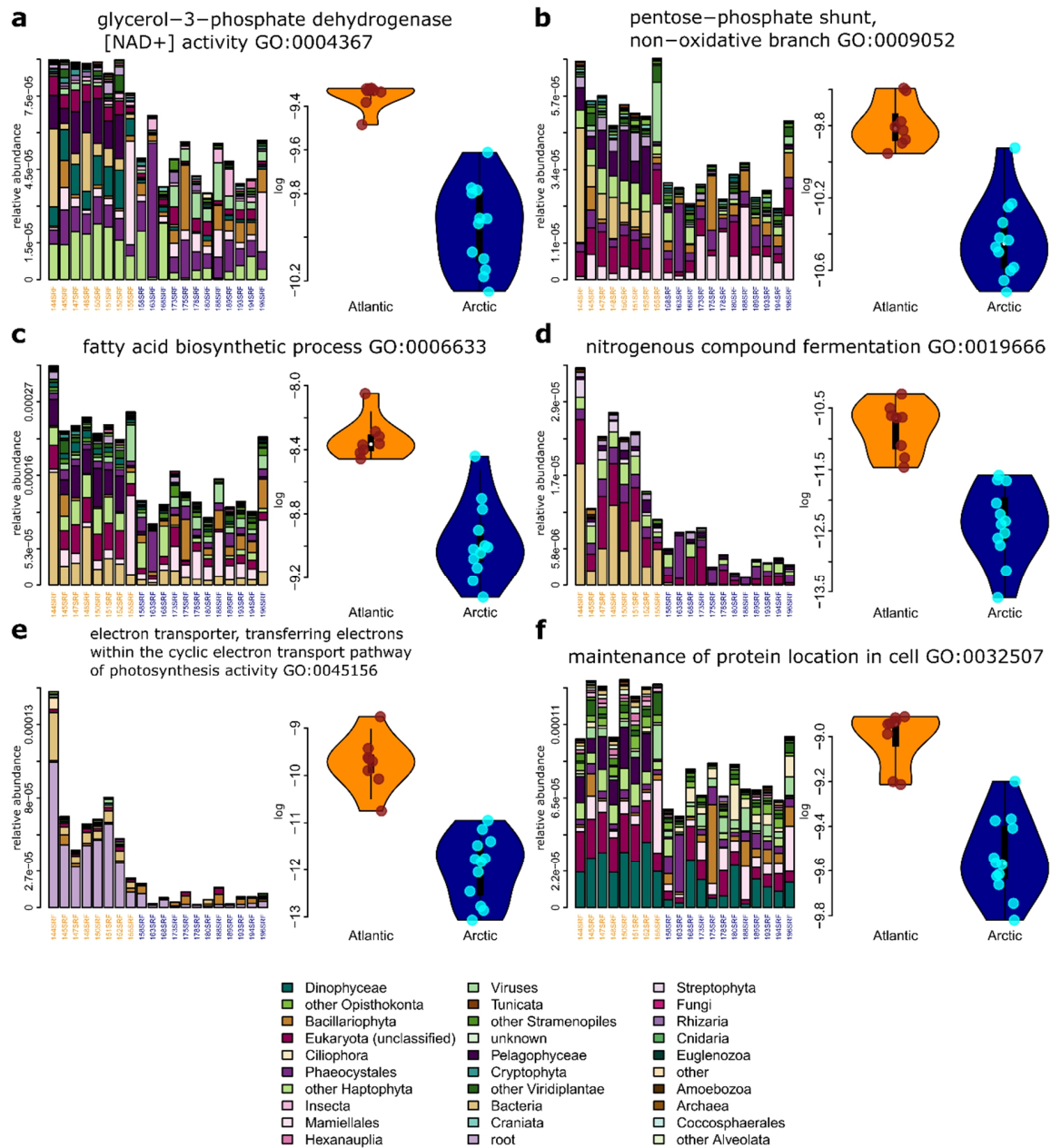
19

expression of PFAM domains between the North Atlantic basin and Arctic basin considering all taxons and specific taxons. We use the ALDEx2 package (13, 21) to compute differentially abundant functions (PFAMs) in either metagenomes (left columns), metatranscriptomes (middle column) and expression (metatranscriptomic abundance normalized by metagenomic abundance, right column) between the North Atlantic Ocean and the Arctic Ocean. For each PFAM the total. For each panel, the median \log_2 difference between basins (equation 7) is plotted in function of the median \log_2 dispersion (for each PFAM) (equation 8) and significantly more abundant PFAMs are colored (bottom part of the plot in the Arctic and top part of the plot in the Atlantic). The dotted lines indicate the limits at which inter-basin dispersion is superior to the intra-basin dispersion.

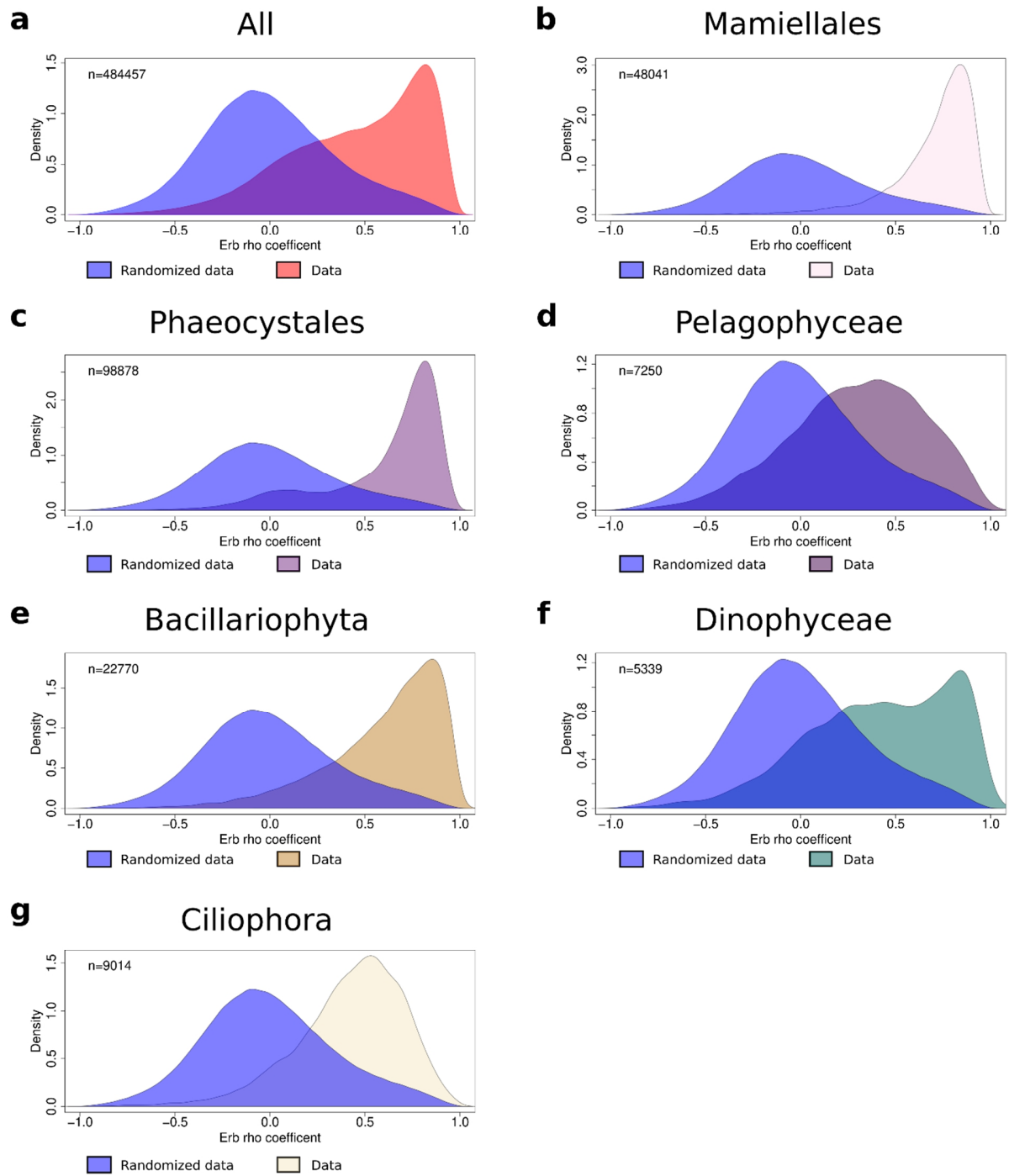


Supplementary Figure 9 | Differential abundance in metatranscriptomes and metagenomes and differential expression of PFAMs associated to temperature response. Right column: Median log₂-fold change between Atlantic and Arctic metagenomes, metatranscriptomes

(equation 6) and expression (metatranscriptomes normalized by metagenomes) of PFAMs associated to temperature response based on estimated centered log ratios from ALDEx2 package for all taxons, Mamiellales, Phaeocystales, Pelagophyceae, Bacillariophyta, Dinophyceae, Ciliophora and Unknown. Significant differences are highlighted by stars. Left column: Median log₂-fold change of the number of unigenes between Atlantic and Arctic metagenomes and metatranscriptomes for aforementioned taxa.

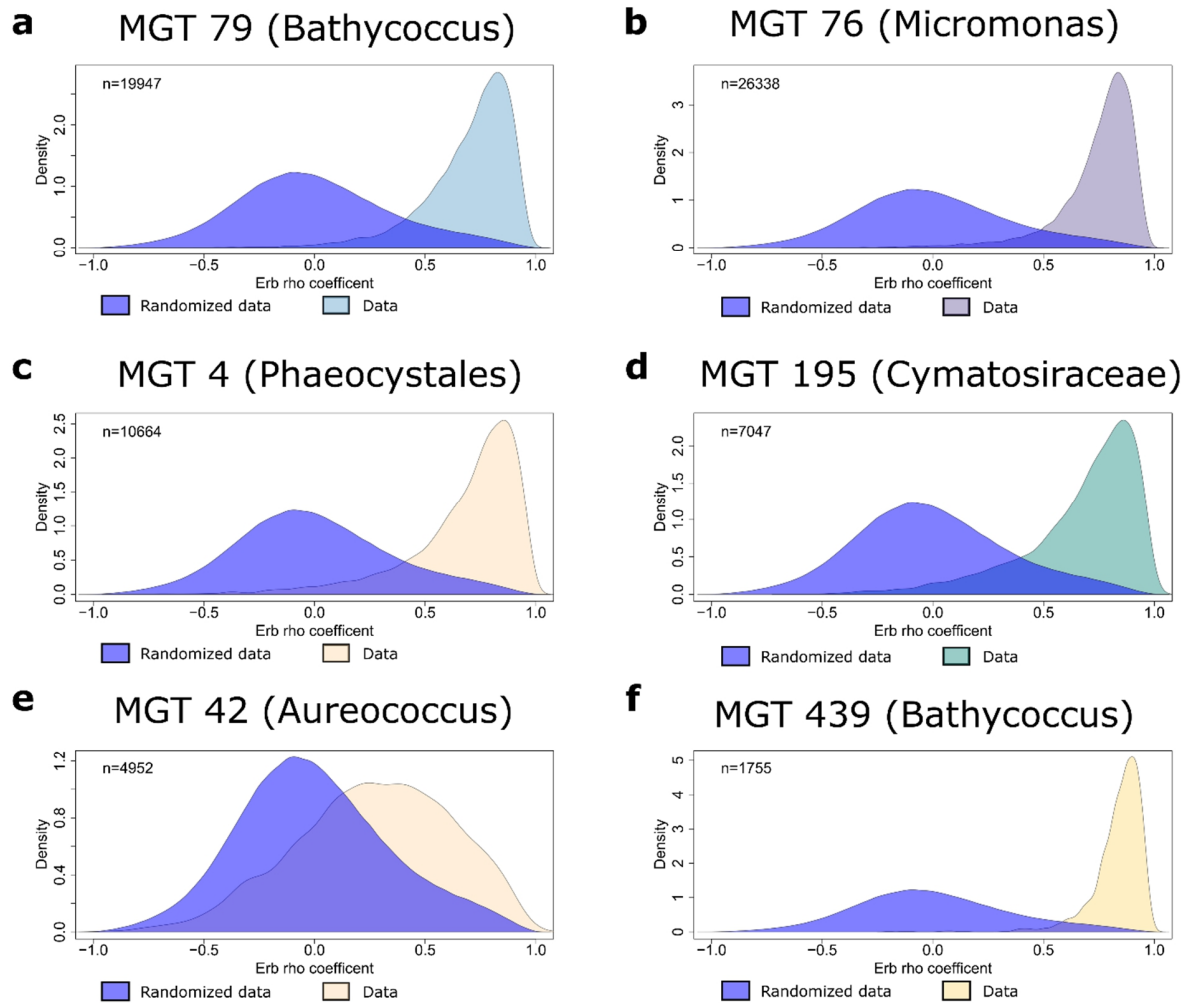


Supplementary Figure 10 | Examples of Gene Ontology (GO) terms more abundant in the Arctic metagenomes. Each panel (a-f) represents (left) the sum of the metagenomic relative abundance of unigenes associated to the abovementioned GO term across considered taxa and (right) the distribution of this relative abundance in the Atlantic (dark blue) versus Arctic basin (purple). Each of these GO terms is significantly more abundant in the Arctic basin.

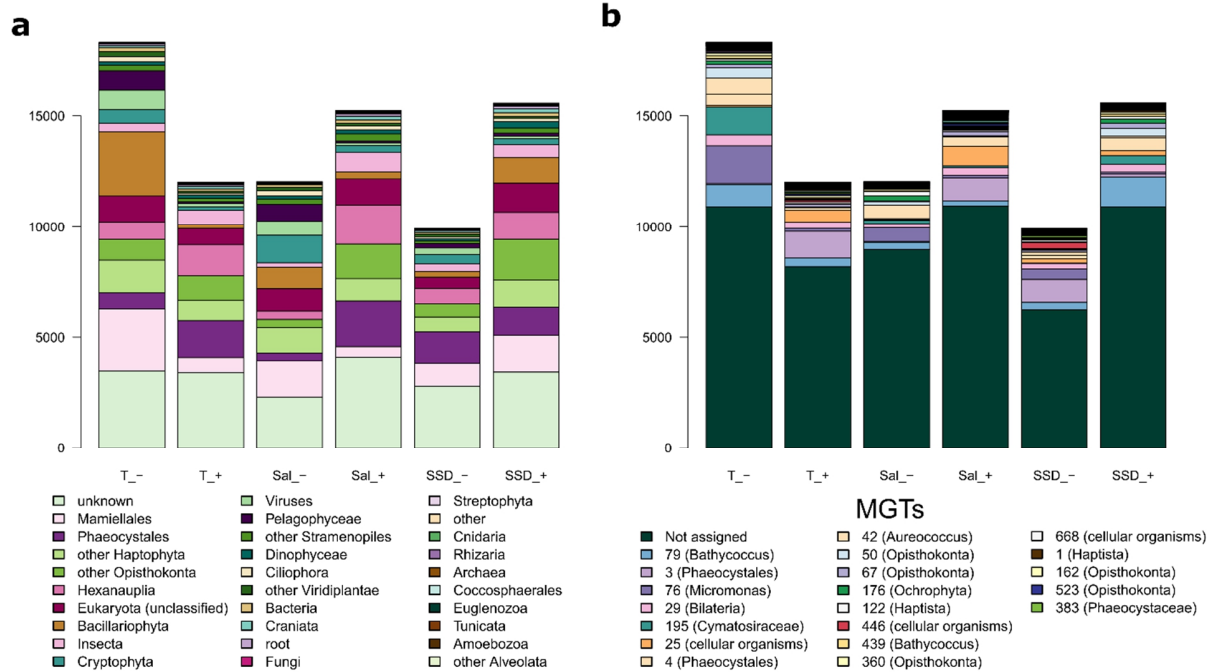


Supplementary Figure 11 | Distribution of Erb rho(47) correlation index between centered log ratios of randomized or real metatranscriptomic and metagenomic relative abundance of unigenes present in at least 5 stations. The correlation index distribution is represented for all taxons (a) and individual taxon (b-g). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. Importantly, metatranscriptomic and

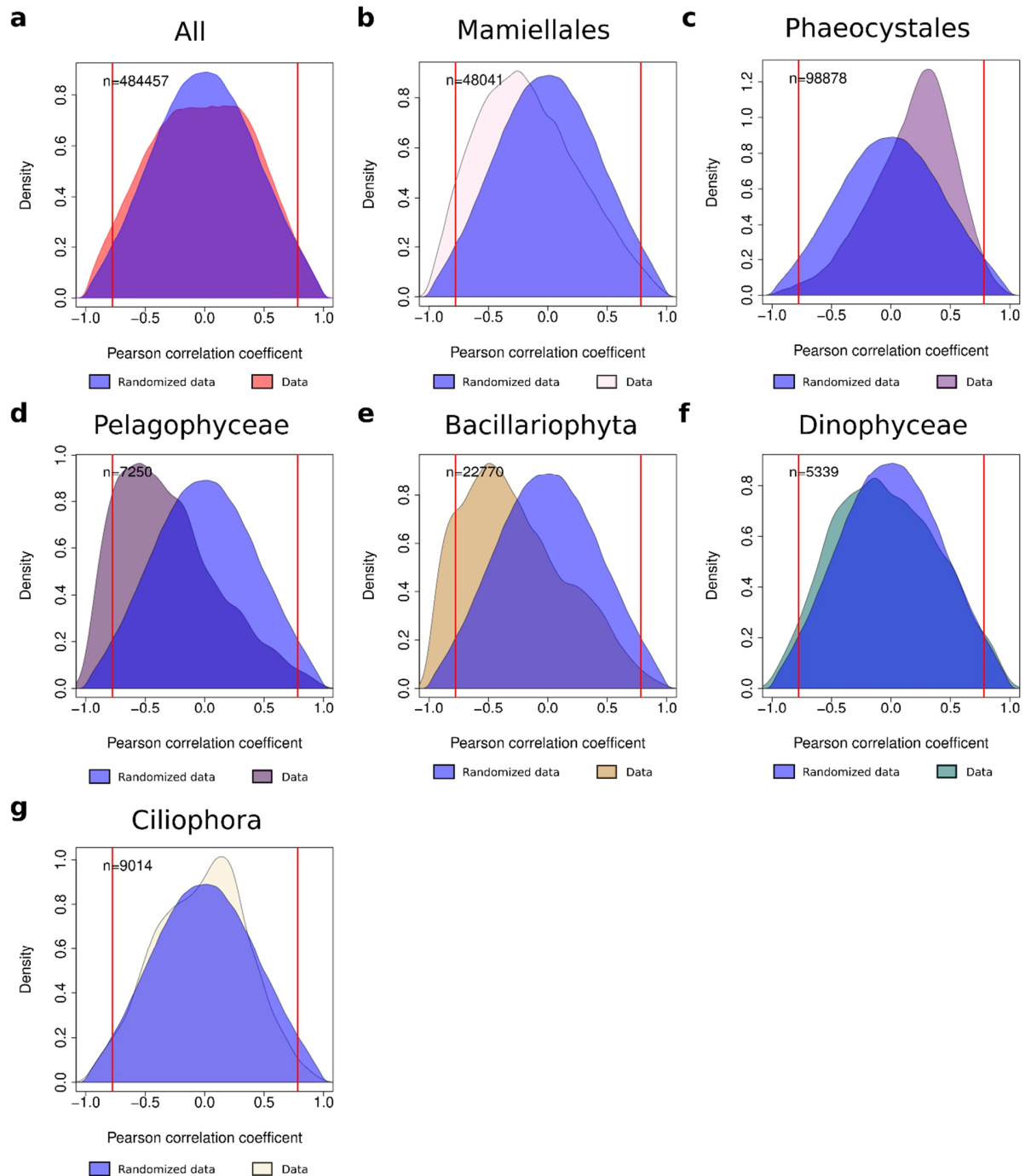
metagenomic centered log ratios (clr) are relatively well correlated although more or less strongly depending on the taxon. These good correlations justify the use of the metagenomic clr to normalize the metatranscriptomic clr to define unigene expression (Figure 4, 5, Supplementary Figure 12-13-14).



Supplementary Figure 12 | Distribution of Erb rho (47) correlation index between centered log ratios of randomized or real metatranscriptomic and metagenomic relative abundance of unigenes present in at least 5 stations of 6 algal Metagenomics-based Transcriptomes (MGTs). The correlation index distribution is plotted for 6 MGTs (a-f). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. Importantly, metatranscriptomic and metagenomic centered log ratios (clr) are relatively well correlated although more or less strongly depending on the taxon. These good correlations justify the use of the metagenomic clr to normalize the metatranscriptomic clr to define unigene expression (Figure 4, 5, Supplementary Figure 12-13-14).

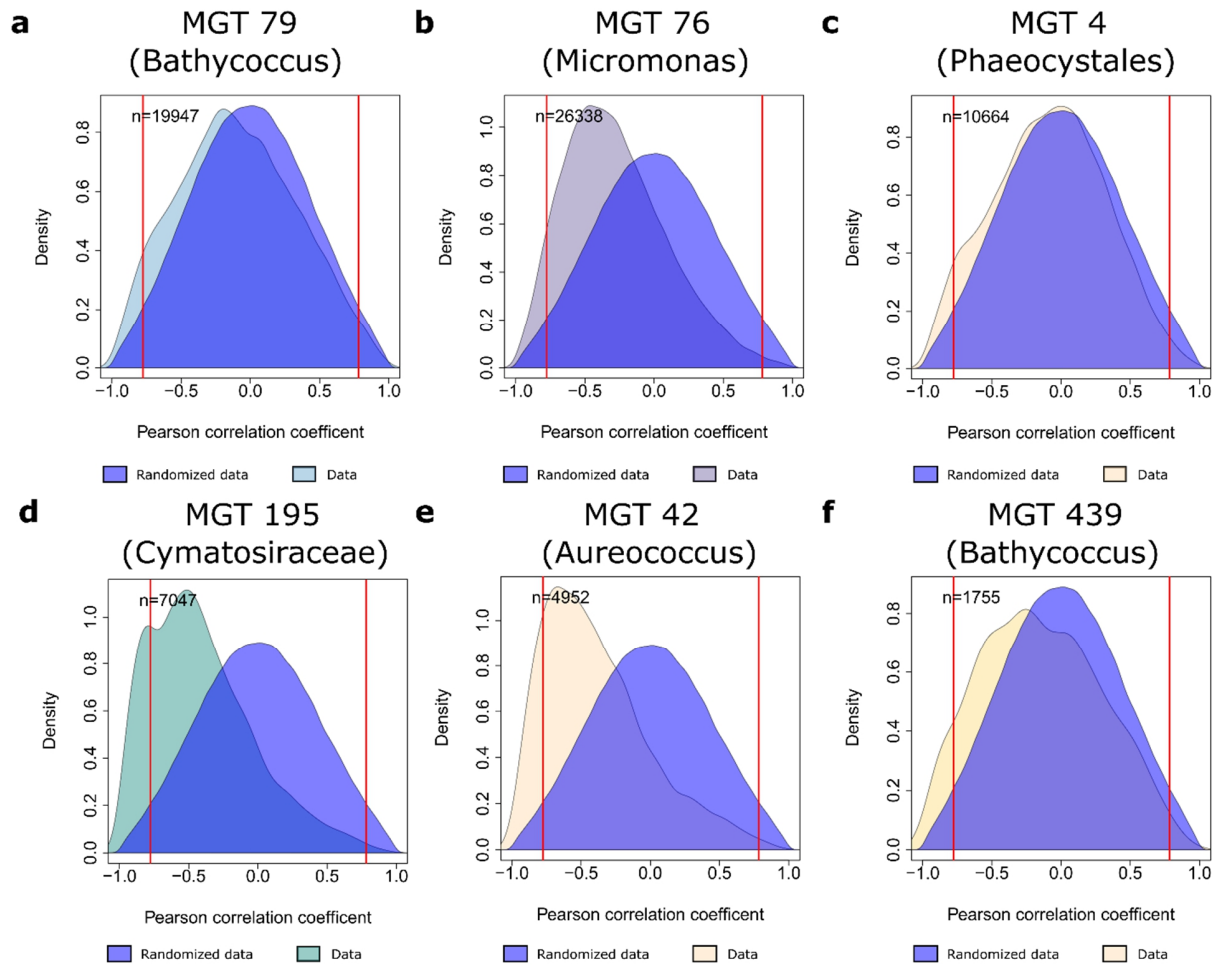


Supplementary Figure 13 | Taxonomic decomposition of unigenes significantly correlated with physical environmental parameters, either negatively or positively. (a) Taxonomic and (b) Metagenomics-based Transcriptomes (MGTs) annotation of unigenes. Temperature is the parameter for which the highest number of unigenes expression profile is correlated to temperature. Salinity (Sal) and day length (SSD) are highly correlated to temperature so we only analyze unigenes functions which are correlated to temperature (Figure 4-5).

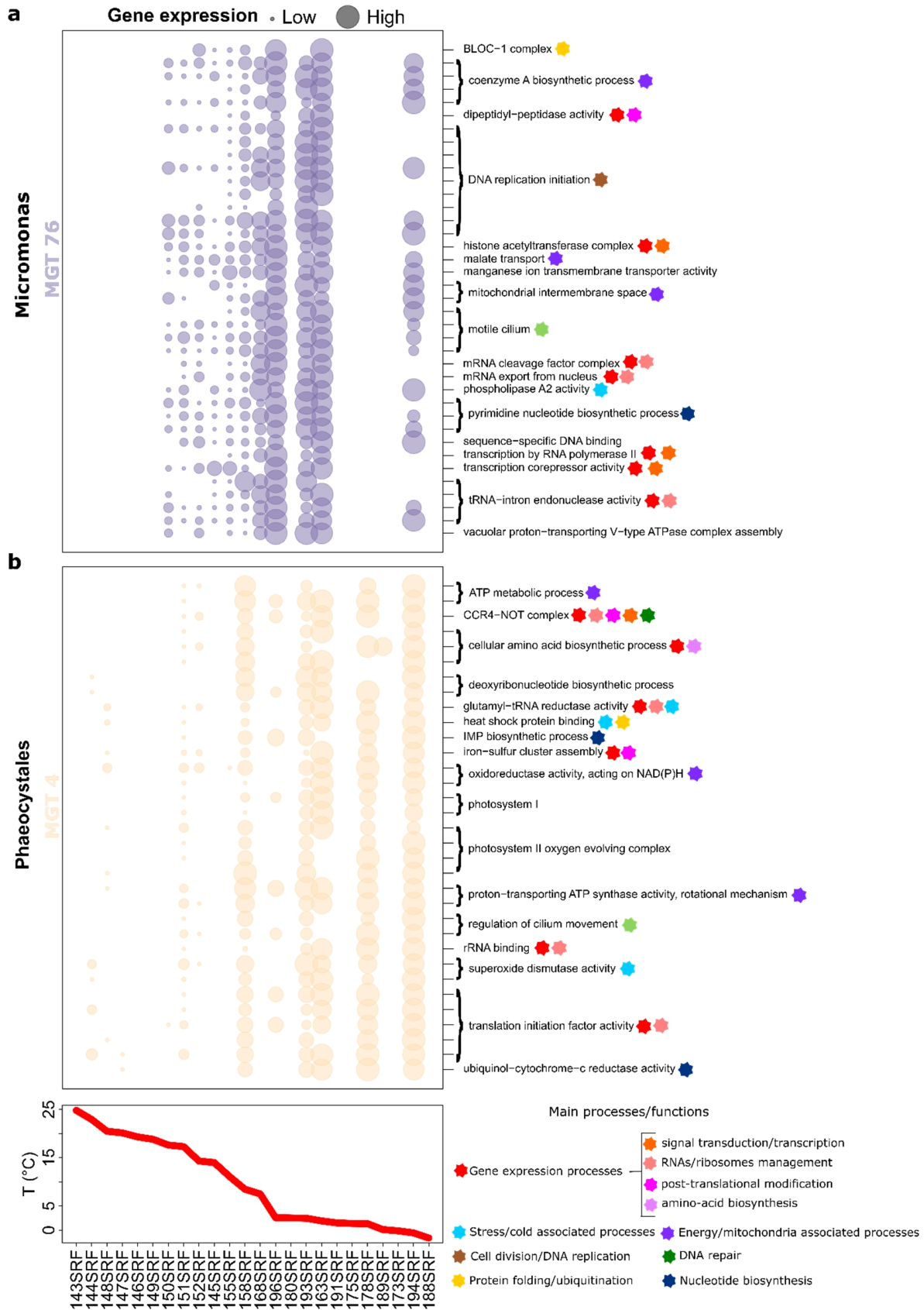


Supplementary Figure 14 | Distribution of Pearson correlation coefficient between temperature and randomized or real expression values of unigenes present in at least 5 stations. The Pearson correlation distribution is represented for all taxons (a) and individual taxon (b-g). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. The expression value of a unigene in a given station is calculated as the

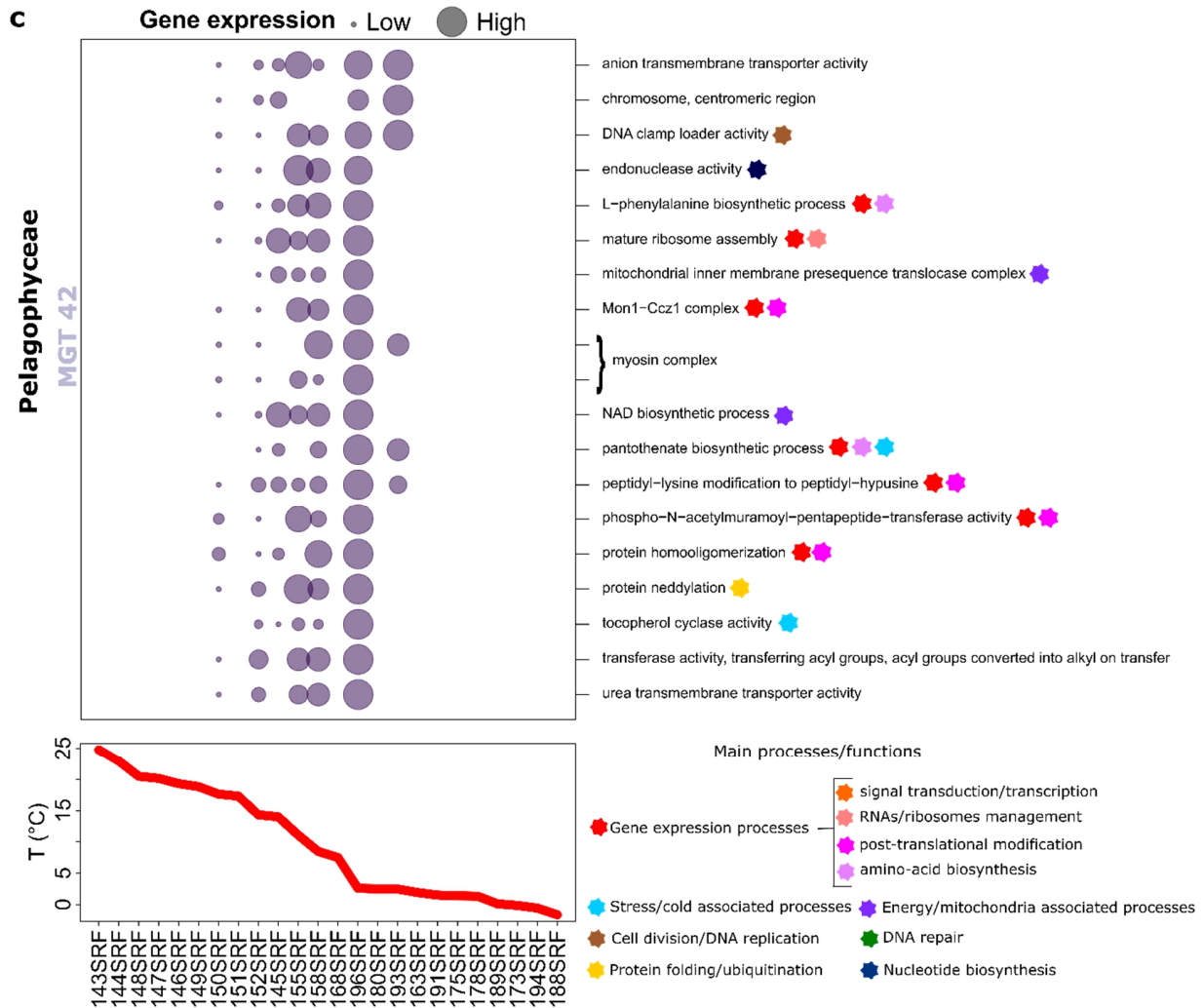
difference between the centered log ratios of its metatranscriptomic relative abundance and metagenomic abundance. For some taxon correlation distribution are clearly distinguishable from the randomized data distribution indicating a clear response to temperature. It is particularly the case for Mamiellales, Pelagophyceae and Bacillariophyta that display skewed distribution enriched in unigenes whose expression profile is negatively correlated with temperature.



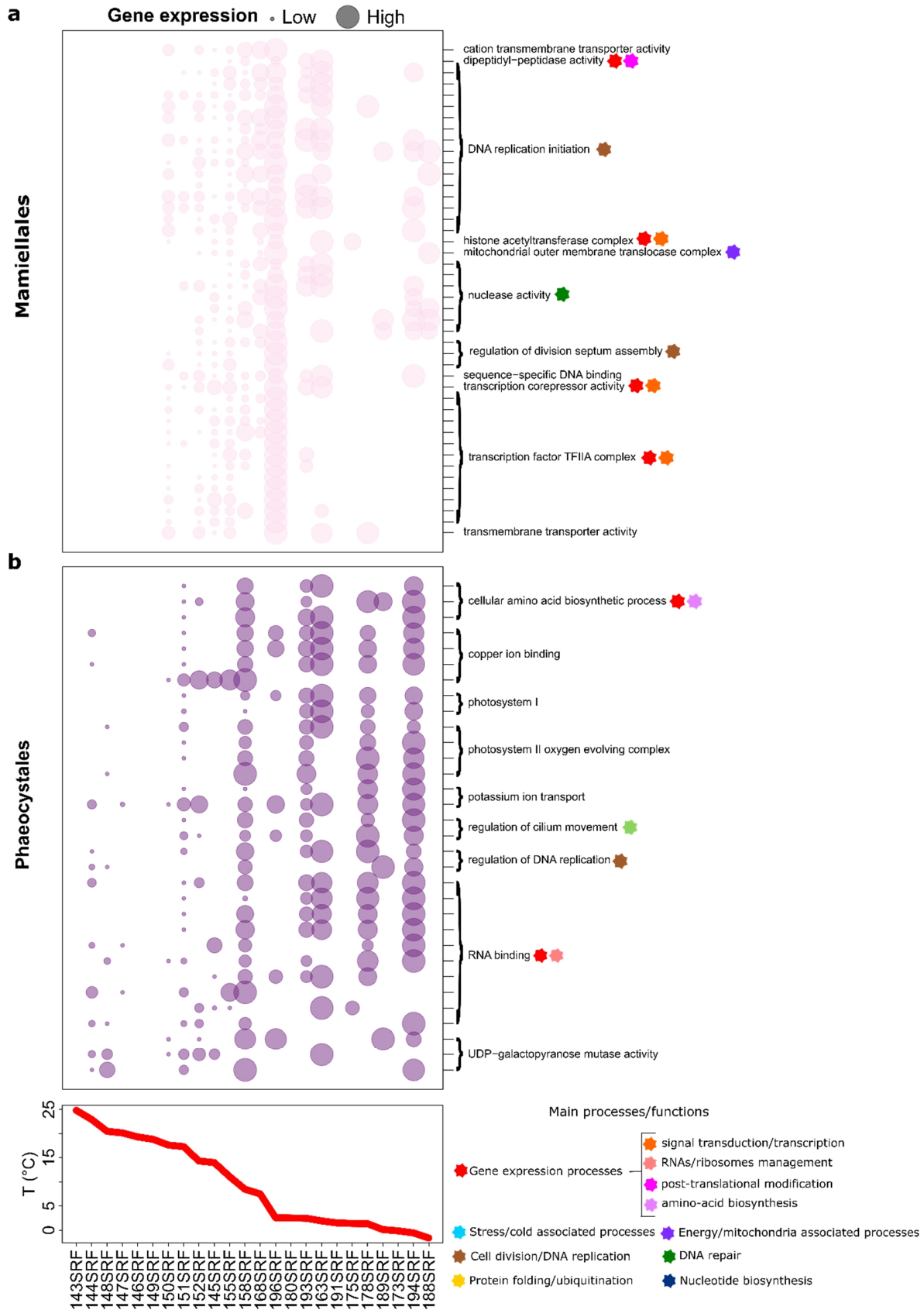
Supplementary Figure 15 | Distribution of Pearson correlation coefficient between temperature and randomized or real expression values of unigenes present in at least 5 stations for 6 algal Metagenomics-based Transcriptomes. The Pearson correlation distribution is represented for 6 algal MGTs (a-f). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. The expression value of a unigene in a given station is calculated as the difference between the centered log ratios of its metatranscriptomic relative abundance and metagenomic abundance. For some MGT correlation distribution are clearly distinguishable from the randomized data distribution indicating a clear response to temperature. It is particularly the case for MGT 76, MGT 195, MGT 42 and MGT 439 that display skewed distribution enriched in unigenes whose expression profile is negatively correlated with temperature.

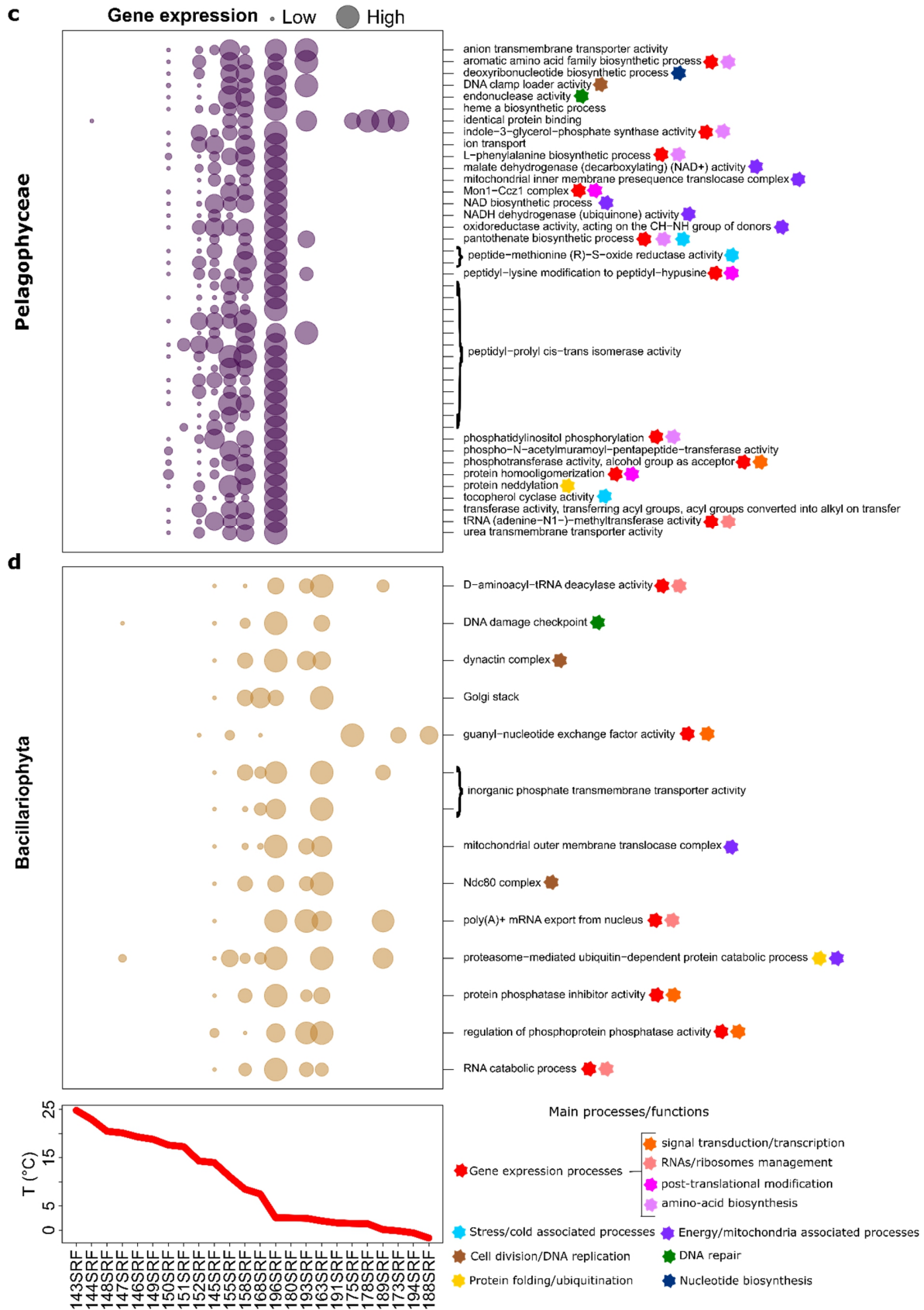


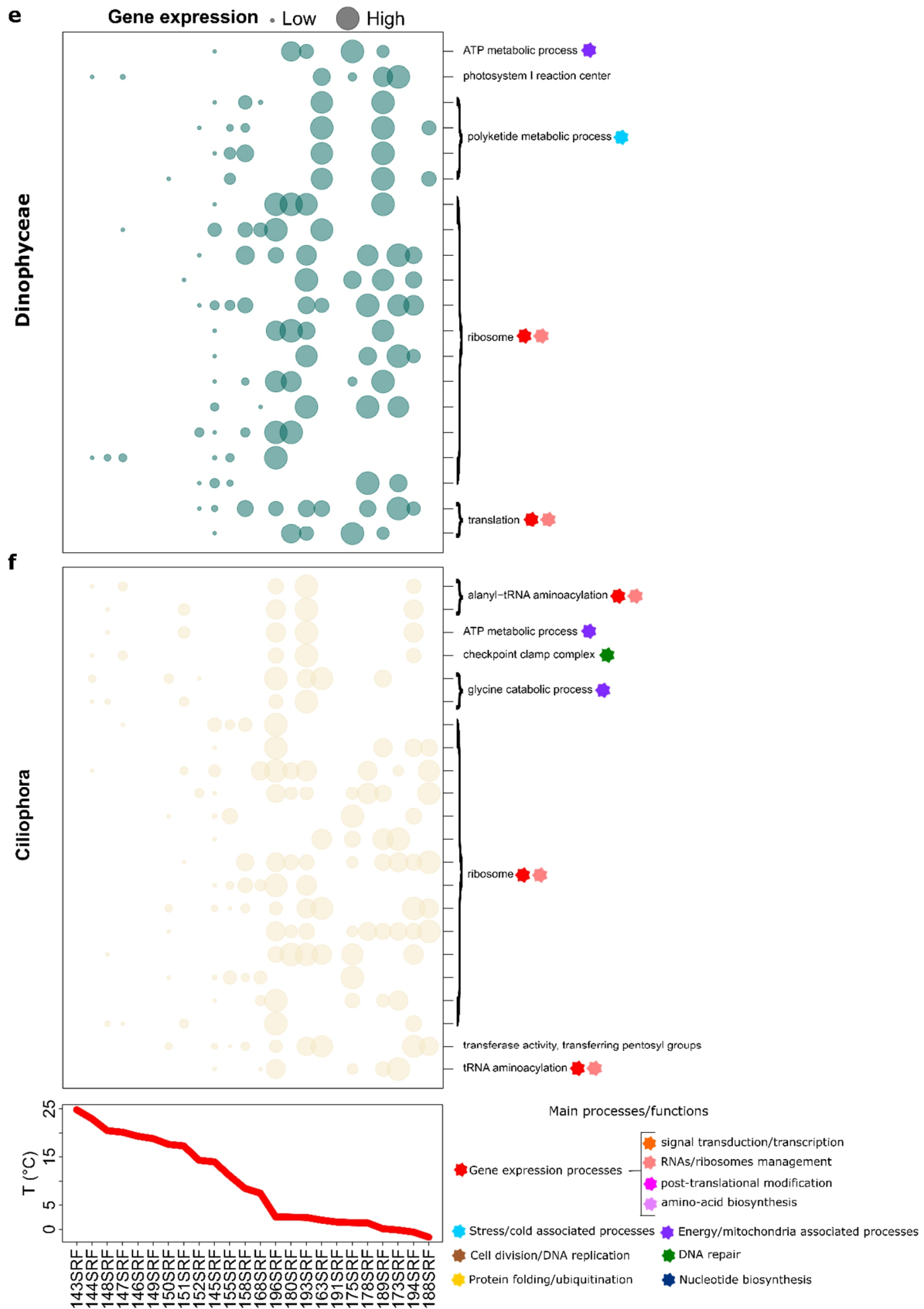
31



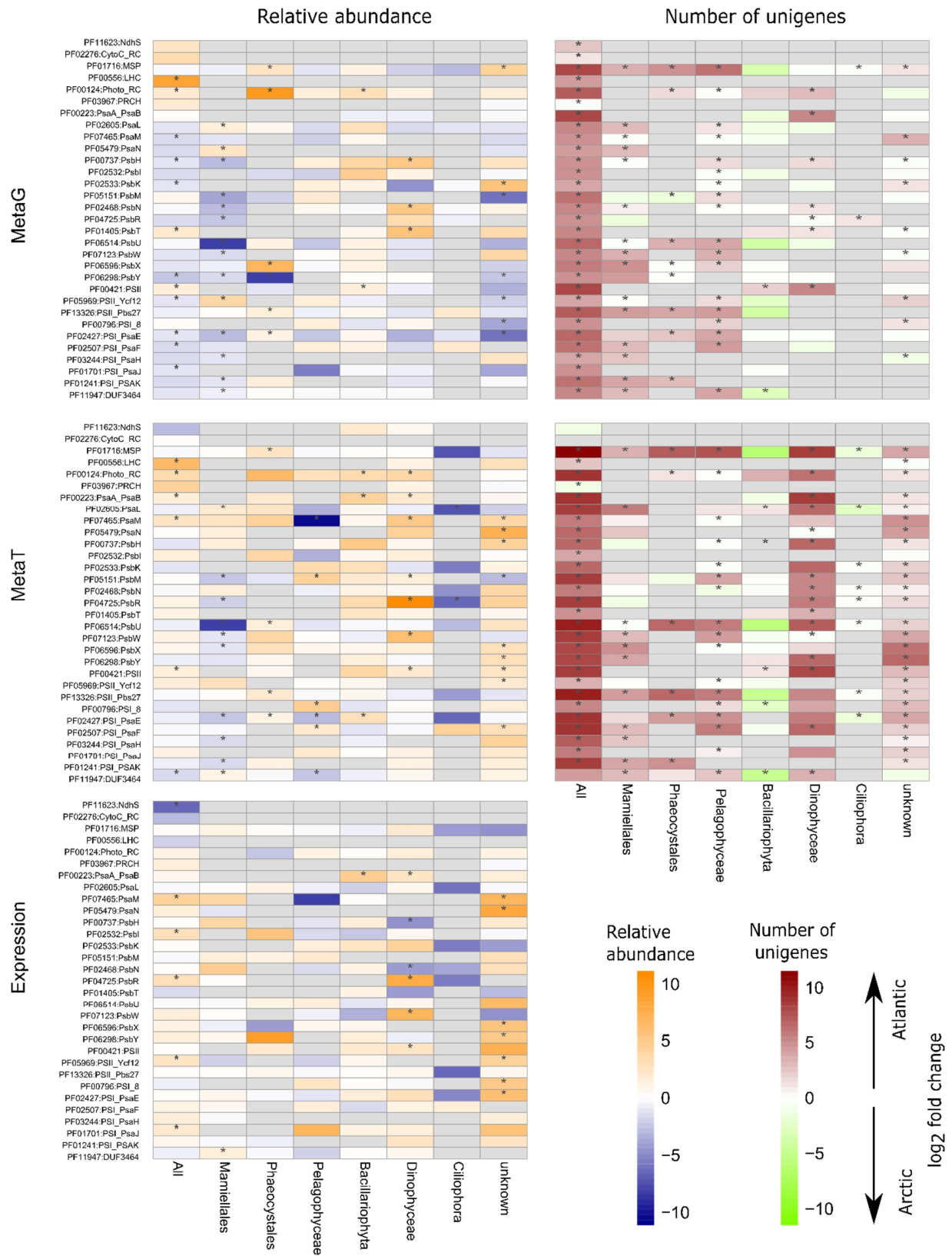
Supplementary Figure 16 | Expression profiles and Gene Ontology annotation of enriched unigenes negatively correlated with temperature for three algal Metagenomics-based Transcriptomes. Expression profiles of unigenes whose Gene Ontology terms (right) are enriched within unigenes significantly negatively correlated to temperature are displayed for three algal taxa from three Metagenomics-based Transcriptomes (MGTs): (a) *Micromonas* (MGT 76) (b) *Phaeocystales* (MGT 4). For MGT 4, unigenes associated to the 'ribosome' GO term were also enriched but removed as they were too numerous (24). (c) *Pelagophyceae* (MGT 42). Note that stations are ordered according to the temperature gradient.





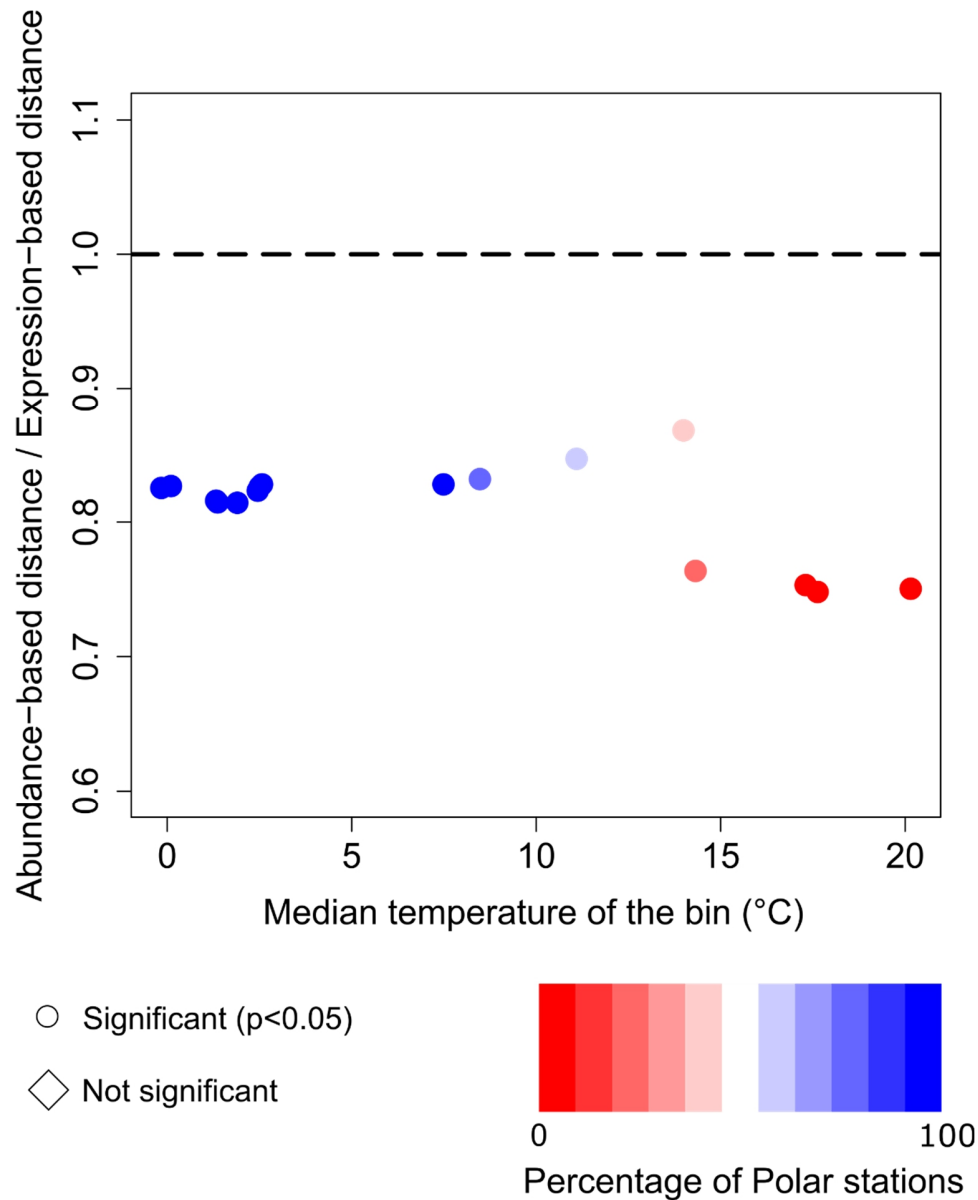


Supplementary Figure 17 | Expression profiles and Gene Ontology annotation of enriched unigenes negatively correlated with temperature for six algal taxa. Expression profiles of unigenes whose Gene Ontology terms (right) are enriched within unigenes significantly negatively correlated to temperature are displayed for six algal taxa: **(a)** Mamiellales (unigenes associated to the ‘nucleus’ GO term were also enriched but removed as they were too numerous (27) as well as the unknown), **(b)** Phaeocystales (unigenes associated to the ‘ribosome’ GO term were also enriched but removed as they were too numerous (40)) **(c)** Pelagophyceae **(d)** Bacillariophyta **(e)** Dinophyceae and **(f)** Ciliophora. Note that stations are ordered according to the temperature gradient.



Supplementary Figure 18 | Differential abundance in metatranscriptomes and

metagenomes and differential expression of PFAMs associated to photosynthesis across all and different taxons using ALDEx2 package (13, 21). Right column: Median log₂-fold change between Atlantic and Arctic metagenomes, metatranscriptomes (equation 6) and expression (metatranscriptomes normalized by metagenomes) of PFAMs associated to photosynthesis based on estimated centered log ratios from ALDEx2 package for all taxons, Mamiellales, Phaeocystales, Pelagophyceae, Bacillariophyta, Dinophyceae, Ciliophora and Unknown. Significant differences are highlighted by stars. Left column: Median log₂-fold change of the number of unigenes between Atlantic and Arctic metagenomes and metatranscriptomes for aforementioned taxa.



Supplementary Figure 19 | Relative contribution of changes in community composition and gene expression changes to changes in metatranscriptomic composition along the temperature gradient of the NAO-AO transition. The median ratio between an abundance-based distance and an expression based distance (see Methods) is plotted for temperature bins each containing 5 stations. A Wilcoxon test and holm correction ($p < 0.05$) is performed for each bin to test whether the median ratio is significantly different than 1.

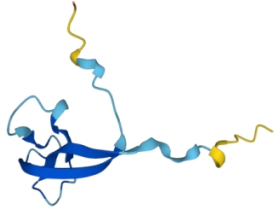
	Significantly more abundant functions	
Location	Metagenomes	Metatranscriptomes
Arctic	<p>"cellulose binding"</p> <p>"sphingolipid metabolic process"</p> <p>"tetraacyldisaccharide 4'-kinase activity"</p> <p>"ceramide metabolic process"</p> <p>"polynucleotide dephosphorylation"</p> <p>"polynucleotide 5' dephosphorylation"</p> <p>"Mon1-Ccz1 complex"</p> <p>"polynucleotide phosphatase activity"</p> <p>"polynucleotide 5'-phosphatase activity"</p> <p>"transcription, RNA-templated"</p> <p>"RNA-directed 5'-3' RNA polymerase activity"</p>	<p>"hydrolase activity, acting on acid carbon-carbon bonds"</p> <p>"cis-trans isomerase activity"</p> <p>"ribonucleoprotein complex biogenesis"</p> <p>"pseudouridine synthase activity"</p> <p>"hydrolase activity, acting on acid carbon-carbon bonds, in ketonic substances"</p> <p>"preribosome, small subunit precursor"</p> <p>"ribosome biogenesis"</p> <p>"anatomical structure homeostasis"</p> <p>"fumarylacetoacetase activity"</p> <p>"snoRNA binding"</p> <p>"telomere organization"</p> <p>"ribonuclease H2 complex"</p> <p>"lipid catabolic process"</p> <p>"telomeric DNA binding"</p> <p>"S-adenosylmethionine-dependent methyltransferase activity"</p> <p>"intrinsic component of organelle membrane"</p> <p>"microtubule anchoring"</p> <p>"methyltransferase complex"</p> <p>"macromolecule methylation"</p> <p>"integral component of organelle membrane"</p> <p>"vacuole"</p> <p>"sphingolipid metabolic process"</p> <p>"urea metabolic process"</p> <p>"microbody"</p> <p>"tRNA methyltransferase complex"</p> <p>"peroxisome"</p> <p>"tRNA (m1A) methyltransferase complex"</p> <p>"protein transmembrane transport"</p> <p>"chromosome, telomeric region"</p> <p>"ceramide metabolic process"</p> <p>"vacuolar part"</p> <p>"nucleic acid phosphodiester bond hydrolysis"</p> <p>"mitochondrial RNA metabolic process"</p> <p>"peptidyl-amino acid modification"</p> <p>"RNA processing"</p> <p>"RNA modification"</p> <p>"rRNA metabolic process"</p> <p>"peptidyl-proline modification"</p> <p>"protein peptidyl-prolyl isomerization"</p> <p>"pseudouridine synthesis"</p> <p>"ncRNA processing"</p> <p>"serine-type exopeptidase activity"</p> <p>"nuclease activity"</p> <p>"nucleolus"</p>

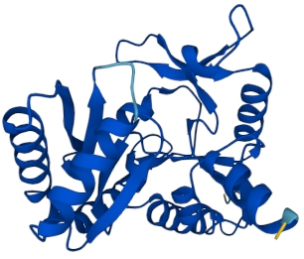
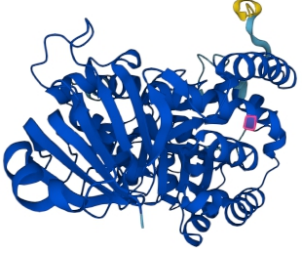
		<p>"vacuolar membrane"</p> <p>"urea catabolic process"</p> <p>"RNA methylation"</p> <p>"serine-type carboxypeptidase activity"</p> <p>"endonuclease activity"</p> <p>"tRNA processing"</p> <p>"intracellular protein transmembrane transport"</p> <p>"peptidyl-prolyl cis-trans isomerase activity"</p> <p>"tRNA modification"</p> <p>"tRNA dihydrouridine synthesis"</p> <p>"rRNA processing"</p> <p>"telomere maintenance"</p> <p>"nuclear chromosome, telomeric region"</p> <p>"vacuolar proton-transporting V-type ATPase complex"</p> <p>"RNA methyltransferase activity"</p> <p>"tRNA methylation"</p> <p>"mitochondrial inner membrane presequence translocase complex"</p> <p>"transcription by RNA polymerase I"</p> <p>"tRNA dihydrouridine synthase activity"</p> <p>"tRNA methyltransferase activity"</p> <p>"tRNA (adenine) methyltransferase activity"</p> <p>"tRNA (adenine-N1-)-methyltransferase activity"</p>
Atlantic	<p>"glycosaminoglycan binding"</p> <p>"transferase activity, transferring aldehyde or ketonic groups"</p> <p>"ligase activity, forming carbon-carbon bonds"</p> <p>"argininosuccinate synthase activity"</p> <p>"hydroxymethylbilane synthase activity"</p> <p>"1-deoxy-D-xylulose-5-phosphate synthase activity"</p> <p>"CoA carboxylase activity"</p> <p>"phosphotransferase activity, carboxyl group as acceptor"</p> <p>"maintenance of location"</p> <p>"acetyl-CoA carboxylase activity"</p> <p>"chorismate synthase activity"</p> <p>"fatty acid synthase activity"</p> <p>"ribose-5-phosphate isomerase activity"</p> <p>"mannosamine metabolic process"</p> <p>"cellulase activity"</p> <p>"ER retention sequence binding"</p> <p>"N-acylglucosamine-6-phosphate 2-epimerase activity"</p> <p>"uridylyltransferase activity"</p> <p>"3-oxoacyl-[acyl-carrier-protein] synthase activity"</p> <p>"sulfate adenylyltransferase activity"</p> <p>"N-acetylmannosamine metabolic process"</p> <p>"photosynthesis, light reaction"</p> <p>"light-harvesting complex"</p> <p>"cytokinetic process"</p> <p>"regulation of cytoskeleton organization"</p>	<p>"transposition"</p> <p>"transposase activity"</p> <p>"chlorophyll binding"</p> <p>"photosynthetic electron transport chain"</p> <p>"ion channel complex"</p> <p>"photosynthetic electron transport in photosystem II"</p> <p>"electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity"</p> <p>"protein splicing"</p> <p>"transposition, DNA-mediated"</p> <p>"single strand break repair"</p> <p>"deoxyribonucleotide catabolic process"</p>

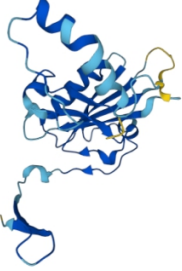
	<p>"sulfate adenyltransferase (ATP) activity"</p> <p>"fermentation"</p> <p>"DNA helicase complex"</p> <p>"glycopeptide alpha-N-acetylgalactosaminidase activity"</p> <p>"maintenance of protein location"</p> <p>"maintenance of location in cell"</p> <p>"oxidoreductase activity, acting on CH or CH2 groups, with an iron-sulfur protein as acceptor"</p> <p>"oxidoreductase activity, acting on diphenols and related substances as donors, with copper protein as acceptor"</p> <p>"regulation of microtubule-based movement"</p> <p>"regulation of cilium movement"</p> <p>"glycerol-3-phosphate dehydrogenase [NAD+] activity"</p> <p>"methylenetetrahydrofolate reductase (NAD(P)H) activity"</p> <p>"regulation of Toll signaling pathway"</p> <p>"plastoquinol--plastocyanin reductase activity"</p> <p>"5-methyltetrahydrofolate-dependent methyltransferase activity"</p> <p>"4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity"</p> <p>"terpenoid metabolic process"</p> <p>"acetyl-CoA carboxylase complex"</p> <p>"photosynthetic electron transport chain"</p> <p>"nitrogenous compound fermentation"</p> <p>"alditol phosphate metabolic process"</p> <p>"glycerol-3-phosphate metabolic process"</p> <p>"glutamyl-tRNA reductase activity"</p> <p>"peptidoglycan turnover"</p> <p>"photosynthetic electron transport in photosystem II"</p> <p>"proteasome accessory complex"</p> <p>"chorismate metabolic process"</p> <p>"electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity"</p> <p>"cell septum assembly"</p> <p>"phosphoglycerate kinase activity"</p> <p>"fatty acid metabolic process"</p> <p>"isoprenoid biosynthetic process"</p> <p>"maintenance of protein location in cell"</p> <p>"phycobilisome"</p> <p>"terpenoid biosynthetic process"</p> <p>"glycerol-3-phosphate catabolic process"</p> <p>"maintenance of protein localization in organelle"</p> <p>"methionine metabolic process"</p> <p>"dipeptidyl-peptidase activity"</p> <p>"maintenance of protein localization in endoplasmic reticulum"</p> <p>"L-lysine metabolic process"</p>	
--	---	--

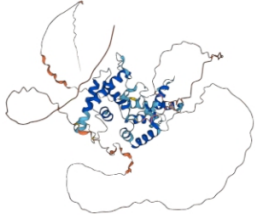

	<p>"protein polyubiquitination"</p> <p>"protein retention in ER lumen"</p> <p>"biotin transport"</p> <p>"exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters"</p> <p>"fatty acid biosynthetic process"</p> <p>"aspartate family amino acid catabolic process"</p> <p>"exodeoxyribonuclease activity"</p> <p>"monocarboxylic acid transmembrane transporter activity"</p> <p>"lysine catabolic process"</p> <p>"anaerobic amino acid catabolic process"</p> <p>"exodeoxyribonuclease activity, producing 5'-phosphomonoesters"</p> <p>"chlorophyll biosynthetic process"</p> <p>"methionine biosynthetic process"</p> <p>"L-lysine catabolic process"</p> <p>"pentose-phosphate shunt, non-oxidative branch"</p> <p>"queuosine salvage"</p> <p>"methionine synthase activity"</p> <p>"L-lysine catabolic process to acetate"</p> <p>"biotin transmembrane transporter activity"</p>	
--	--	--

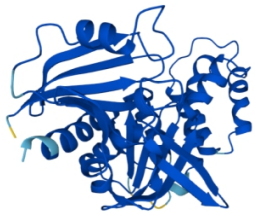

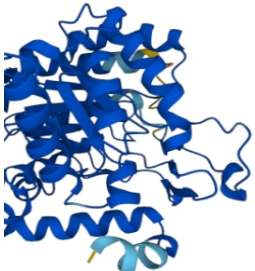
Supplementary Table 1 | List of more abundant GO terms in both North Atlantic and Arctic Ocean metagenomes and metatranscriptomes

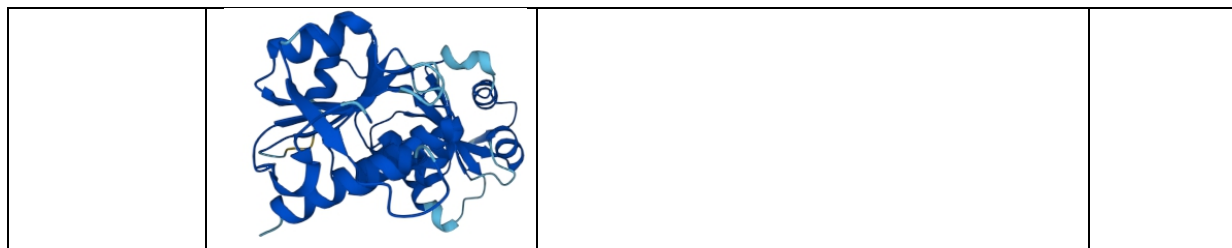
PFAM	Molecular function and structure	Relevance to cold/hot stress acclimation	References
PF14169: Cold-inducible protein YdjO	Molecular: Unknown Structure (example of YdjO) (48): 	Strongly cold-shock-induced in a transcriptome analysis of <i>B. subtilis</i>	(49)
PF05562: WCOR413	Molecular: putative transport of substances or ions, mediation of protein/protein interaction Structure: Highly hydrophobic protein with six transmembrane regions	A low-temperature-regulated protein from wheat resembling (25% to 56% identity) a stress-responsive gene from the resurrection plant <i>Xerophyta viscosa</i> Baker (XVSAP1). Hypothesis: "It is possible that XVSAP1 may be involved in the transport of substances or ions across the plasma membrane as a region stretching from amino-terminal glycine residue of a growing polypeptide. Proteins so modified have diverse functions and the myristate appears to be critical for mediating protein/protein and/or protein/membrane interactions"(50)	(50, 51)
PF00313: CSD	Molecular: Part of this domain is highly similar to the RNP-1 RNA-binding motif(52). Structure: Protein domain of about 70 amino acids which has been found in prokaryotic and eukaryotic DNA-binding proteins(53, 54).	"may have a function in activating transcription or unwinding or masking RNA molecules."(54)	(52–54)
PF02649: GCHY-1	Molecular: tetrahydrofolate biosynthesis. type I GTP cyclohydrolase Structure: highly conserved glutamate residue at position 216 in folE2: likely to be the substrate ligand. The metal ligand is likely to be the cysteine at position 147. Zn ²⁺ dependent(55).	Cold-upregulated in a thermophilic alga: "Glycine decarboxylase, which feeds methylene-tetrahydrofolate into the folate cycle was 1.7-fold upregulated at 28°C. In contrast, transcript frequency for formate-tetrahydrofolate ligase, which would drain the folate cycle, was reduced to 0.3-fold at 28°C"(56) "Both in vitro and in vivo experiments indicate that OtNOS, unlike mammalian NOS, efficiently uses tetrahydrofolate as a cofactor in <i>Arabidopsis</i> plants. The modulation of NO production to alleviate abiotic stress disturbances in higher plants highlights the potential of genetic manipulation to influence NO metabolism as a tool to improve plant fitness under adverse growth conditions."(57)	(55–57)
PF14249: Tocopherol_cycl	Molecular:tocopherol cyclase activity. Secondary metabolism Structure: 3 β-feuillet and 1 α-helix, strong confidence except the tip of the β and α(48) (N and C-terminal domains)	Abstract: "As expected, the synthesis of tocopherols (vitamin E) from 2,3-dimethyl-5-phytyl-1,4-benzoquinol by the recombinant protein demonstrated that the EgSXD1 gene product is a TC. Northern hybridizations with RNA isolated from cell suspension cultures clearly showed a fast EgSXD1 downregulation by cold."(24)	(24)
PF01964: Thiamine biosynthesis	Molecular: Thiamine biosynthesis, "Escherichia coli K-12 synthesizes thiamine pyrophosphate (vitamin B1)	"The expression of <i>Arabidopsis</i> genes involved in the thiamine diphosphate biosynthesis pathway, including that of TH11, TH1C, TH1 and TPK, was analyzed for 48 h in seedlings	(22, 58)

	<p><i>de novo</i>. Two precursors [4-methyl-5-(beta-hydroxyethyl)thiazole monophosphate and 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate] are coupled to form thiamine monophosphate, which is then phosphorylated to make thiamine pyrophosphate.”(22)</p> <p>Structure: A β-feuillet surrounded by α-helices and α-β-α on the side (C-terminal)(58)</p>	<p>subjected to NaCl or sorbitol treatment. These genes were found to be predominantly up-regulated in the early phase (2-6 h) of the stress response.”(22)</p>	
<p>PF02358: Trehalose_PPase</p>	<p>Molecular: Trehalose phosphatase Structure (example of Trehalose phosphatase)(48):</p> 	<p>Abstract: “A mutant <i>Escherichia coli</i> strain unable to produce trehalose died much faster than the wild type at 4°C. Transformation of the mutant with the <i>otsA/otsB</i> genes, responsible for trehalose synthesis, restored trehalose content and cell viability at 4°C. After temperature downshift from 37°C to 16°C (“cold shock”), trehalose levels in wild-type cells increased up to 8-fold.”(59)</p>	(59)
<p>PF01979: Amidohydro_1</p>	<p>Molecular: Allantoinase, Catalyzes the conversion of allantoin (5-ureidohydantoin) to allantoinic acid by hydrolytic cleavage of the five-member hydantoin ring(60). Structure (example of allantoinase)(48):</p> 	<p>“Allantoin is a metabolic intermediate of purine catabolism that often accumulates in stressed plants(27).”</p>	(27, 60, 61)
<p>PF10551: MULE</p>	<p>Molecular: Mutator-like element (MULE) transposases. Zn dependent(62) Structure: This domain was identified by Babu and colleagues(62). “Typical structural features of MULEs include long TIRs (>100 bp) and 8–10 bp flanking TSDs(63)”</p>	<p>NONE (control)</p>	(62, 63)

<p>PF01119: DNA_mis_repair</p>	<p>Function : DNA repair(64). Structure: "This family represents the C-terminal domain of the <i>mutL/hexB/PMS1</i> family. This domain has a ribosomal S5 domain 2-like fold."(64)</p>	<p>Title: DNA Damage Inducible Protein 1 is Involved in Cold Adaption of Harvested Cucumber Fruit(65)</p>	<p>(64, 65)</p>
<p>PF00626: Gelsolin</p>	<p>Function : Gelsolin is a cytoplasmic, calcium-regulated, actin-modulating protein that binds to the barbed ends of actin filaments, preventing monomer exchange (end-blocking or capping). It can promote nucleation (the assembly of monomers into filaments), as well as sever existing filaments(66). Structure (example of Gelsolin)(48) : uncertain</p> 	<p>"Cold-acclimation had a negligible effect on steady state protein expression in the heart, with just 7 differentially abundant (DA) proteins (0.5% of the quantified cardiac proteome). Furthermore, only gelsolin isoform X1 (log2FC = -0.289) was not influenced by a confounding effect of available development stage"</p>	<p>(66)</p>
<p>PF00241: Cofilin_ADF</p>	<p>Function: "They bind both actin-monomers and filaments and promote rapid filament turnover in cells by depolymerising/fragmenting actin filaments. ADF/cofilins bind ADP-actin with higher affinity than ATP-actin and inhibit the spontaneous nucleotide exchange on actin monomers"(67) Structure: "In molecular biology, ADF-H domain (actin-depolymerising factor homology domain) is an approximately 150 amino acid motif that is present in three phylogenetically distinct classes of eukaryotic actin-binding proteins(68)."</p>	<p>"Based on the above information and on the identification of a wheat cold-regulated ADF homolog, we designed the present experiments to determine the function and regulation of TaADF."(69)</p>	<p>(67–69)</p>
<p>PF02209: VHP</p>	<p>Molecular: Bundles, nucleates, caps, and severs actin in a Ca²⁺-dependent</p>	<p>"Compared with Table I, only eight spots, which represents seven proteins (38.9%) including a villin (<i>Os03g24220</i>), an actin 3 (<i>Os03g61970</i>), an actin 7 (<i>Os05g01600</i>), an aminotransferase (<i>Os08g41990</i>), an ATP-dependent Clp protease (<i>Os04g32560</i>),</p>	<p>(70–72)</p>

	<p>manner(70).</p> <p>Structure: Villin headpiece domain. <i>"The core of the GH domain consists of a five-stranded β-sheet sandwiched between one long and one short helix"</i>(71)</p>	<p><i>a putative subtilisin (Os02g53860), and a sucrose synthase (Os03g28330), showed similar up- or down-regulated patterns.</i>(72)</p>	
<p>PF03259: Robl_LC7</p>	<p>Molecular: associated to flagellar outer arm of dynein. Structure: See Koonin et al.(73)</p>	<p>This family includes proteins that may modulate specific dynein functions but it may also play a structural or regulatory role. Dynein were shown to be regulated in cold conditions and act as a <i>"thermally tunable motor"</i>(74)</p>	(73, 74)
<p>PF00249: Myb_DNA-binding</p>	<p>Molecular: Myb-like DNA-binding domain (transcription factor) Structure(48):</p> 	<p>Putative role in oxidative stress response in <i>Arabidopsis Thaliana</i>: regulation of ROS in a circadian manner (probable regulation by CCA1)(75)</p>	(75)
<p>PF11999: Ice_binding</p>	<p>Molecular: Ice binding, antifreeze glycoproteins, they bind to crystals of ice Structure (example of Ice binding protein 1)(48):</p> 	<p><i>"Sub-zero temperatures put plants at risk of damage associated with the formation of ice crystals in the apoplast. Some freeze-tolerant plants mitigate this risk by expressing ice-binding proteins (IBPs), that adsorb to ice crystals and modify their growth."</i>(76)</p>	(48, 76)
<p>PF03104: DNA_pol_B_exo1</p>	<p>Molecular: 3' to 5' exonuclease activity Structure: ribonuclease H type fold</p>	<p><i>"Previous studies [15] showed changes in RNase activity in wheat leaves and roots following exposure to low temperatures (2 8C) for long periods (60 days)."</i>(77)</p>	(77)
<p>PF01346: FKBP_N</p>	<p>Molecular: peptidyl-prolyl isomerase (foldase/chaperone), post translation modification Structure(78): Domain amino terminal to FKBP-type</p>	<p>Some peptidyl prolyl isomerase have been shown to be important in protein folding under cold stress condition by increasing the rate of cis-trans isomerization(79)</p>	(78, 79)
<p>PF05035: DGOK</p>	<p>Molecular: 2-dehydro-3-deoxygalactonokinase activity Structure (example of</p>	<p>NONE (control)</p>	(48)

	<p>deoxygalactonokinase)(48):</p> 		
<p>PF05351: GMP_PDE_delta</p>	<p>Molecular: "PDE delta subunit is thought to be a specific soluble transport factor for certain prenylated proteins and Arl2-GTP a regulator of PDE-mediated transport."(80)</p> <p>Structure (example of GMP_PDE_delta containing protein) (48):</p> 	<p>"The enzyme is stable only under specific conditions, and the activation property of the enzyme is lost relatively easy. Irreversible modifications occur at temperatures below 0° and above 30°C, and at pH below 6.0. Several other conditions such as high ion concentrations, temperatures just above 0°C and pH above 8.0 lead to reversibel modifications of enzyme activity."(81)</p>	<p>(80, 81)</p>
<p>PF01180: DHO_dh</p>	<p>Molecular: cytosolic dihydroorotate dehydrogenase Structure (example of dihydroorotate dehydrogenase) (48):</p> 	<p>"In the present paper, we identified and cloned OsDHODH1 encoding a putative cytosolic dihydroorotate dehydrogenase (DHODH) in rice. Expression analysis indicated that OsDHODH1 is upregulated by salt, drought and exogenous abscisic acid (ABA), but not by cold"(82)</p> <p>"Even though there is no data on the effect of temperature on DhoDH, we can expect that its thermal sensitivity will affect more mitochondrial functions than only the reduction of the Q pool."(83)</p>	<p>(82, 83)</p>
<p>PF05971: Methyltransf_10</p>	<p>Molecular: RNA methyltransferase Structure (example of Methyltransferase)(48):</p>	<p>RNA methylation is known to be important in stress response especially cold acclimation (stability of RNA)(84)</p>	<p>(84, 85)</p>



Supplementary table 2 | Summary table of function and structure of a set of PFAMs known to be involved in cold/hot stress acclimation.

References

1. P. Wassmann, K. N. Kosobokova, D. Slagstad, K. F. Drinkwater, R. R. Hopcroft, S. E. Moore, I. Ellingsen, R. J. Nelson, E. Carmack, E. Popova, J. Berge, The contiguous domains of Arctic Ocean advection: Trails of life and death. *Prog Oceanogr.* **139** (2015), doi:10.1016/j.pocean.2015.06.011.
2. A. C. Palmisano, C. W. Sullivan, PHYSIOLOGY OF SEA ICE DIATOMS. I. RESPONSE OF THREE POLAR DIATOMS TO A SIMULATED SUMMER-WINTER TRANSITION. *J Phycol.* **18** (1982), doi:10.1111/j.1529-8817.1982.tb03215.x.
3. F. Kong, I. T. Romero, J. Warakanont, Y. Li-Beisson, Lipid catabolism in microalgae. *New Phytologist.* **218** (2018), doi:10.1111/nph.15047.
4. W. Li, R. Wang, M. Li, L. Li, C. Wang, R. Welti, X. Wang, Differential degradation of extraplastidic and plastidic lipids during freezing and post-freezing recovery in *Arabidopsis thaliana*. *J Biol Chem.* **283**, 461–468 (2008).
5. B. Usadel, O. E. Bläsing, Y. Gibon, F. Poree, M. Höhne, M. Günter, R. Trethewey, B. Kamlage, H. Poorter, M. Stitt, Multilevel genomic analysis of the response of transcripts, enzyme activities and metabolites in *Arabidopsis* rosettes to a progressive decrease of temperature in the non-freezing range. *Plant Cell Environ.* **31**, 518–547 (2008).
6. I. A. Guschina, J. L. Harwood, Lipids and lipid metabolism in eukaryotic algae. *Prog Lipid Res.* **45**, 160–186 (2006).
7. S. S. Gill, S. Willette, B. Dungan, J. M. Jarvis, T. Schaub, D. M. van Leeuwen, R. St Hilaire, F. Omar Holguin, Suboptimal temperature acclimation affects Kennedy pathway gene expression, lipidome and metabolite profile of *nannochloropsis salina* during PUFA enriched TAG synthesis. *Mar Drugs.* **16** (2018), doi:10.3390/md16110425.
8. L. Valledor, T. Furuhashi, A. M. Hanak, W. Weckwerth, Systemic cold stress adaptation of *Chlamydomonas reinhardtii*. *Molecular and Cellular Proteomics.* **12** (2013), doi:10.1074/mcp.M112.026765.
9. E. Huby, J. A. Napier, F. Baillieux, L. v. Michaelson, S. Dhondt-Cordelier, Sphingolipids: towards an integrated view of metabolism during the plant stress response. *New Phytologist.* **225** (2020), , doi:10.1111/nph.15997.
10. J. H. T. M. Fabri, N. P. de Sá, I. Malavazi, M. del Poeta, The dynamics and role of sphingolipids in eukaryotic organisms upon thermal adaptation. *Prog Lipid Res.* **80** (2020), , doi:10.1016/j.plipres.2020.101063.
11. D. Carrero, J. G. Pérez-Silva, V. Quesada, C. López-Otín, Differential mechanisms of tolerance to extreme environmental conditions in tardigrades. *Sci Rep.* **9** (2019), doi:10.1038/s41598-019-51471-8.
12. J. J. Luxton, M. J. McKenna, A. Lewis, L. E. Taylor, K. A. George, S. M. Dixit, M. Moniz, W. Benegas, M. J. Mackay, C. Mozsary, D. Butler, D. Bezdán, C. Meydan, B. E. Crucian, S. R. Zwart, S. M. Smith,

- C. E. Mason, S. M. Bailey, Telomere Length Dynamics and DNA Damage Responses Associated with Long-Duration Spaceflight. *Cell Rep.* **33** (2020), doi:10.1016/j.celrep.2020.108457.
13. A. D. Fernandes, J. N. S. Reid, J. M. Macklaim, T. A. McMurrugh, D. R. Edgell, G. B. Gloor, Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome.* **2** (2014), doi:10.1186/2049-2618-2-15.
 14. Y. P. Wang, L. S. Zhou, Y. Z. Zhao, S. W. Wang, L. L. Chen, L. X. Liu, Z. Q. Ling, F. J. Hu, Y. P. Sun, J. Y. Zhang, C. Yang, Y. Yang, Y. Xiong, K. L. Guan, D. Ye, Regulation of G6PD acetylation by SIRT2 and KAT9 modulates NADPH homeostasis and cell survival during oxidative stress. *EMBO Journal.* **33** (2014), doi:10.1002/emj.201387224.
 15. L. Jin, Y. Zhou, Crucial role of the pentose phosphate pathway in malignant tumors (review). *Oncol Lett.* **17** (2019), doi:10.3892/ol.2019.10112.
 16. H. Liu, D. Huang, D. L. McArthur, L. G. Boros, N. Nissen, A. P. Heaney, Fructose induces transketolase flux to promote pancreatic cancer growth. *Cancer Res.* **70** (2010), doi:10.1158/0008-5472.CAN-09-4615.
 17. J. M. Holding, C. M. Duarte, J. M. Arrieta, R. Vaquer-Suyner, A. Coello-Camba, P. Wassmann, S. Agustí, Experimentally determined temperature thresholds for Arctic plankton community metabolism. *Biogeosciences.* **10** (2013), doi:10.5194/bg-10-357-2013.
 18. N. T. Miller, M. D. Vaughn, R. L. Burnap, Electron flow through NDH-1 complexes is the major driver of cyclic electron flow-dependent proton pumping in cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* **1862**, 148354 (2021).
 19. W. Huang, Y.-J. Yang, H. Hu, S.-B. Zhang, Different roles of cyclic electron flow around photosystem I under sub-saturating and saturating light intensities in tobacco leaves. *Front Plant Sci.* **6** (2015), doi:10.3389/fpls.2015.00923.
 20. M. Suorsa, Cyclic electron flow provides acclimatory plasticity for the photosynthetic machinery under various environmental conditions and developmental stages. *Front Plant Sci.* **6** (2015), doi:10.3389/fpls.2015.00800.
 21. A. D. Fernandes, J. M. Macklaim, T. G. Linn, G. Reid, G. B. Gloor, ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One.* **8** (2013), doi:10.1371/journal.pone.0067019.
 22. M. Rapala-Kozik, N. Wolak, M. Kujda, A. K. Banas, The upregulation of thiamine (vitamin B 1) biosynthesis in *Arabidopsis thaliana* seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response. *BMC Plant Biol.* **12** (2012), doi:10.1186/1471-2229-12-2.
 23. S. Ferreira, K. Hjernø, M. Larsen, G. Wingsle, P. Larsen, S. Fey, P. Roepstorff, M. Salomé Pais, Proteome profiling of *Populus euphratica* Oliv. upon heat stress. *Ann Bot.* **98** (2006), doi:10.1093/aob/mcl106.

24. W. el Kayal, G. Keller, C. Debayles, R. Kumar, D. Weier, C. Teulieres, C. Marque, Regulation of tocopherol biosynthesis through transcriptional control of tocopherol cyclase during cold hardening in *Eucalyptus gunnii*. *Physiol Plant*. **126** (2006), doi:10.1111/j.1399-3054.2006.00614.x.
25. S. Q. Ouyang, S. J. He, P. Liu, W. K. Zhang, J. S. Zhang, S. Y. Chen, The role of tocopherol cyclase in salt stress tolerance of rice (*Oryza sativa*). *Sci China Life Sci*. **54** (2011), doi:10.1007/s11427-011-4138-1.
26. R. Szymańska, J. Kruk, Plastoquinol is the main prenyllipid synthesized during acclimation to high light conditions in arabidopsis and is converted to plastochromanol by tocopherol cyclase. *Plant Cell Physiol*. **51** (2010), doi:10.1093/pcp/pcq017.
27. H. Takagi, Y. Ishiga, S. Watanabe, T. Konishi, M. Egusa, N. Akiyoshi, T. Matsuura, I. C. Mori, T. Hirayama, H. Kaminaka, H. Shimada, A. Sakamoto, Allantoin, a stress-related purine metabolite, can activate jasmonate signaling in a MYC2-regulated and abscisic acid-dependent manner. *J Exp Bot*. **67** (2016), doi:10.1093/jxb/erw071.
28. P. Wang, C. H. Kong, B. Sun, X. H. Xu, Distribution and function of allantoin (5-ureidohydantoin) in rice grains. *J Agric Food Chem*. **60** (2012), doi:10.1021/jf2051043.
29. M. M. Seppänen, K. Fagerstedt, The role of superoxide dismutase activity in response to cold acclimation in potato. *Physiol Plant*. **108** (2000), doi:10.1034/j.1399-3054.2000.108003279.x.
30. K. H. Baek, D. Z. Skinner, Differential expression of manganese superoxide dismutase sequence variants in near isogenic lines of wheat during cold acclimation. *Plant Cell Rep*. **25** (2006), doi:10.1007/s00299-005-0073-6.
31. C. Kayihan, F. Eyidogan, N. Afsar, H. A. Oktem, M. Yucel, Cu/Zn superoxide dismutase activity and respective gene expression during cold acclimation and freezing stress in barley cultivars. *Biol Plant*. **56** (2012), doi:10.1007/s10535-012-0143-x.
32. J. Porankiewicz, A. K. Clarke, Induction of the heat shock protein ClpB affects cold acclimation in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *J Bacteriol*. **179** (1997), doi:10.1128/jb.179.16.5111-5117.1997.
33. N. Ukaji, C. Kuwabara, D. Takezawa, K. Arakawa, S. Yoshida, S. Fujikawa, Accumulation of small heat-shock protein homologs in the endoplasmic reticulum of cortical parenchyma cells in mulberry in association with seasonal cold acclimation. *Plant Physiol*. **120** (1999), doi:10.1104/pp.120.2.481.
34. L. E. Teigen, J. I. Orczewska, J. McLaughlin, K. M. O'Brien, Cold acclimation increases levels of some heat shock protein and sirtuin isoforms in threespine stickleback. *Comp Biochem Physiol A Mol Integr Physiol*. **188** (2015), doi:10.1016/j.cbpa.2015.06.028.
35. Y. Huang, D. Jin, C. Lu, X. Lan, P. Qiao, H. Li, Y. Chen, Proteomic responses associated with freezing tolerance in the callus of the Tibetan alpine plant *Saussurea laniceps* during cold acclimation. *Plant Cell Tissue Organ Cult*. **124** (2016), doi:10.1007/s11240-015-0876-2.

36. A. C. Seybold, D. A. Wharton, M. A. S. Thorne, C. J. Marshall, Investigating trehalose synthesis genes after cold acclimation in the Antarctic nematode *Panagrolaimus* sp. DAW1. *Biol Open*. **6**, 1953–1959 (2017).
37. D. Z. Skinner, Phospholipase A2 activity during cold acclimation of wheat. *Crop Sci*. **49** (2009), doi:10.2135/cropsci2008.06.0359.
38. X.-C. Wang, Q.-Y. Zhao, C.-L. Ma, Z.-H. Zhang, H.-L. Cao, Y.-M. Kong, C. Yue, X.-Y. Hao, L. Chen, J.-Q. Ma, J.-Q. Jin, X. Li, Y.-J. Yang, Global transcriptome profiles of *Camellia sinensis* during cold acclimation. *BMC Genomics*. **14**, 415 (2013).
39. T. Pang, C.-Y. Ye, X. Xia, W. Yin, De novo sequencing and transcriptome analysis of the desert shrub, *Ammopiptanthus mongolicus*, during cold acclimation using Illumina/Solexa. *BMC Genomics*. **14**, 488 (2013).
40. C. M. Paget, J.-M. Schwartz, D. Delneri, Environmental systems biology of cold-tolerant phenotype in *Saccharomyces* species adapted to grow at different temperatures. *Mol Ecol*. **23**, 5241–5257 (2014).
41. M. Moulager, A. Monnier, B. Jesson, R. Bouvet, J. Mosser, C. Schwartz, L. Garnier, F. Corellou, F. Y. Bouget, Light-dependent regulation of cell division in *Ostreococcus*: Evidence for a major transcriptional input. *Plant Physiol*. **144** (2007), doi:10.1104/pp.107.096149.
42. J. S. O'Neill, G. van Ooijen, L. E. Dixon, C. Troein, F. Corellou, F. Y. Bouget, A. B. Reddy, A. J. Millar, Circadian rhythms persist without transcription in a eukaryote. *Nature*. **469** (2011), doi:10.1038/nature09654.
43. G. Benoit, M. Mariadassou, S. Robin, S. Schbath, P. Peterlongo, C. Lemaitre, SimkaMin: Fast and resource frugal de novo comparative metagenomics. *Bioinformatics*. **36** (2020), doi:10.1093/bioinformatics/btz685.
44. P. J. Somerfield, Identification of the Bray-Curtis similarity index: Comment on Yoshioka (2008). *Mar Ecol Prog Ser* (2008), , doi:10.3354/meps07841.
45. V. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. **10** (1966).
46. R. R. Sokal, A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull*. **38** (1958).
47. I. Erb, C. Notredame, How should we measure proportionality on relative gene expression data? *Theory in Biosciences*. **135** (2016), doi:10.1007/s12064-015-0220-8.
48. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).

49. T. Kaan, G. Homuth, U. Mäder, J. Bandow, T. Schweder, Genome-wide transcriptional profiling of the *Bacillus subtilis* cold-shock response. *Microbiology (N Y)*. **148**, 3441–3455 (2002).
50. D. Garwe, J. A. Thomson, S. G. Mundree, Molecular characterization of XVSAP1, a stress-responsive gene from the resurrection plant *Xerophyta viscosa* Baker. *J Exp Bot*. **54**, 191–201 (2003).
51. X. Huang, A. Gaballa, M. Cao, J. D. Helmann, Identification of target promoters for the *Bacillus subtilis* extracytoplasmic function σ factor, σ^W . *Mol Microbiol*. **31**, 361–371 (1999).
52. D. Landsman, RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res*. **20**, 2861–2864 (1992).
53. J. Doniger, D. Landsman, M. A. Gonda, G. Wistow, The product of *unr*, the highly conserved gene upstream of *N-ras*, contains multiple repeats similar to the cold-shock domain (CSD), a putative DNA-binding motif. *New Biol*. **4**, 389–395 (1992).
54. P. G. Jones, M. Inouye, The cold-shock response--a hot topic. *Mol Microbiol*. **11**, 811–818 (1994).
55. B. Sankaran, S. A. Bonnett, K. Shah, S. Gabriel, R. Reddy, P. Schimmel, D. A. Rodionov, V. de Crécy-Lagard, J. D. Helmann, D. Iwata-Reuyl, M. A. Swairjo, Zinc-independent folate biosynthesis: genetic, biochemical, and structural investigations reveal new metal dependence for GTP cyclohydrolase IB. *J Bacteriol*. **191**, 6936–6949 (2009).
56. A. W. Rossoni, G. Schönknecht, H. J. Lee, R. L. Rupp, S. Flachbart, T. Mettler-Altmann, A. P. M. Weber, M. Eisenhut, Cold Acclimation of the Thermoacidophilic Red Alga *Galdieria sulphuraria*: Changes in Gene Expression and Involvement of Horizontally Acquired Genes. *Plant Cell Physiol*. **60**, 702–712 (2019).
57. N. Foresi, M. L. Mayta, A. F. Lodeyro, D. Scuffi, N. Correa-Aragunde, C. García-Mata, C. Casalongué, N. Carrillo, L. Lamattina, Expression of the tetrahydrofolate-dependent nitric oxide synthase from the green alga *Ostreococcus tauri* increases tolerance to abiotic stresses and influences stomatal development in *Arabidopsis*. *The Plant Journal*. **82**, 806–821 (2015).
58. S. Coquille, C. Roux, A. Mehta, T. P. Begley, T. B. Fitzpatrick, S. Thore, High-resolution crystal structure of the eukaryotic HMP-P synthase (THIC) from *Arabidopsis thaliana*. *J Struct Biol*. **184**, 438–444 (2013).
59. O. Kandror, A. DeLeon, A. L. Goldberg, Trehalose synthesis is induced upon exposure of *Escherichia coli* to cold and is essential for viability at low temperatures. *Proc Natl Acad Sci U S A*. **99**, 9727–9732 (2002).
60. <https://www.uniprot.org/uniprot/Q8XGS6>.
61. I. Ramazzina, L. Cendron, C. Folli, R. Berni, D. Monteverdi, G. Zanotti, R. Percudani, Logical identification of an allantoinase analog (*puuE*) recruited from polysaccharide deacetylases. *J Biol Chem*. **283**, 23295–23304 (2008).

62. M. M. Babu, L. M. Iyer, S. Balaji, L. Aravind, The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* **34**, 6505–6520 (2006).
63. K. Liu, S. R. Wessler, Transposition of Mutator-like transposable elements (MULEs) resembles hAT and Transib elements and V(D)J recombination. *Nucleic Acids Res.* **45**, 6644–6655 (2017).
64. S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).
65. B. Wang, G. Wang, S. Zhu, DNA Damage Inducible Protein 1 is Involved in Cold Adaption of Harvested Cucumber Fruit. *Front Plant Sci.* **10**, 1723 (2019).
66. A. G. WEEDS, J. GOOCH, B. POPE, H. E. HARRIS, Preparation and characterization of pig plasma and platelet gelsolins. *Eur J Biochem.* **161**, 69–76 (1986).
67. P. Lappalainen, M. M. Kessels, M. J. Cope, D. G. Drubin, The ADF homology (ADF-H) domain: a highly exploited actin-binding module. *Mol Biol Cell.* **9**, 1951–1959 (1998).
68. <https://pfam.xfam.org/family/PF00241>.
69. F. Ouellet, E. Carpentier, M. J. Cope, A. F. Monroy, F. Sarhan, Regulation of a wheat actin-depolymerizing factor during cold acclimation. *Plant Physiol.* **125**, 360–368 (2001).
70. E. Friederich, K. Vancompernelle, D. Louvard, J. Vandekerckhove, Villin function in the organization of the actin cytoskeleton. Correlation of in vivo effects to its biochemical activities in vitro. *J Biol Chem.* **274**, 26751–26760 (1999).
71. U. Ghoshdastider, D. Popp, L. D. Burtneck, R. C. Robinson, The expanding superfamily of gelsolin homology domain proteins. *Cytoskeleton (Hoboken).* **70**, 775–795 (2013).
72. C. Huo, B. Zhang, H. Wang, F. Wang, M. Liu, Y. Gao, W. Zhang, Z. Deng, D. Sun, W. Tang, Comparative Study of Early Cold-Regulated Proteins by Two-Dimensional Difference Gel Electrophoresis Reveals a Key Role for Phospholipase D α 1 in Mediating Cold Acclimation Signaling Pathway in Rice. *Mol Cell Proteomics.* **15**, 1397–1411 (2016).
73. E. V Koonin, L. Aravind, *Curr Biol*, in press, doi:10.1016/s0960-9822(00)00774-0.
74. W. Hong, A. Takshak, O. Osunbayo, A. Kunwar, M. Vershinin, The Effect of Temperature on Microtubule-Based Transport by Cytoplasmic Dynein and Kinesin-1 Motors. *Biophys J.* **111**, 1287–1294 (2016).
75. L. A. Grace, D. C. J., M.-R. Bernd, K. S. A., S. J. H. M., D. P. P., CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proceedings of the National Academy of Sciences.* **109**, 17129–17134 (2012).
76. M. Bredow, V. K. Walker, Ice-Binding Proteins in Plants. *Front Plant Sci.* **8** (2017), doi:10.3389/fpls.2017.02153.

77. C. Spanò, C. Crosatti, R. Pacchini, P. Meletti, I. Grilli, Ribonucleases during cold acclimation in winter and spring wheats. *Plant Science*. **162**, 809–815 (2002).
78. R. D. Finn, A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, M. Punta, Pfam: The protein families database. *Nucleic Acids Res*. **42** (2014), , doi:10.1093/nar/gkt1223.
79. C. Budiman, Y. Koga, K. Takano, S. Kanaya, FK506-Binding Protein 22 from a Psychrophilic Bacterium, a Cold Shock-Inducible Peptidyl Prolyl Isomerase with the Ability to Assist in Protein Folding. *Int J Mol Sci*. **12**, 5261–5284 (2011).
80. <https://www.ncbi.nlm.nih.gov/Structure/cdd/PF05351>.
81. R. Bulgakov, P. J. M. Van Haastert, Isolation and partial characterization of a cyclic GMP-dependent cyclic GMP-specific phosphodiesterase from Dictyostelium discoideum. *Biochimica et Biophysica Acta (BBA) - General Subjects*. **756**, 56–66 (1983).
82. W. Y. Liu, M. M. Wang, J. Huang, H. J. Tang, H. X. Lan, H. S. Zhang, The OsDHODH1 gene is involved in salt and drought tolerance in rice. *J Integr Plant Biol*. **51** (2009), doi:10.1111/j.1744-7909.2009.00853.x.
83. H. Lemieux, P. U. Blier, Exploring Thermal Sensitivities and Adaptations of Oxidative Phosphorylation Pathways. *Metabolites*. **12**, 360 (2022).
84. C. Zhang, G. Jia, Reversible RNA Modification N1-methyladenosine (m1A) in mRNA and tRNA. *Genomics Proteomics Bioinformatics*. **16** (2018), , doi:10.1016/j.gpb.2018.03.003.
85. <https://www.rcsb.org/structure/6b91>.

CHAPITRE 3 : Modélisation de la biogéographie du phytoplancton, un modèle basé individu dans un double gyre de l'Atlantique Nord idéalisé

I. Introduction : une compréhension mécanistique de la biogéographie et de la biodiversité?

Dans ce dernier chapitre, nous abordons la question de la biogéographie du plancton de manière plus théorique. En effet comme déjà énoncé dans l'introduction, l'observation de la nature permet d'ériger certaines lois mathématiques établissant des liens causaux entre les phénomènes naturels et expliquant notamment leur dynamique. Ce chapitre se distingue ainsi des deux autres qui étaient des études statistiques de données dont on a pu tirer certaines tendances voire lois. L'approche est ainsi inversée et la question n'est plus « Qu'est-ce que j'observe et qu'est ce que j'en déduis ? » mais plutôt « Qu'observe-t-on selon la simulation des lois générales observées ? Puis-je reconstruire les observations initiales? Si oui, puis-je faire des prédictions ? ». Cette approche dite mécanistique s'attachera ici en la modélisation de la dynamique des espèces phytoplanctoniques dans leur milieu à l'aide de lois connues théoriques ou empiriques décrivant le comportement de ces dernières. Plus précisément la question posée ici est : puis-je reconstruire une biogéographie selon les lois connues de croissance du phytoplancton ?

En particulier, pour approcher cette question de manière relativement simple j'ai développé un modèle de biogéographie du phytoplancton dans une circulation océanique idéalisée de l'Atlantique Nord. L'approche modèle basé individu a été choisie c'est-à-dire que le phytoplancton est modélisé par un grand nombre de 'particules' se déplaçant dans le champ de vitesse des courants océaniques. Le nombre de cellules de phytoplancton par litre d'eau de mer étant bien trop élevé (de l'ordre du million^{338,339}), chaque 'particule' du modèle représente un très grand nombre de cellules phytoplanctoniques (approche de la 'superparticule' ou de la 'supercellule'³⁴⁰). Cette approche est aussi dite « Lagrangienne » et se distingue de

l'approche « Euleurienne » qui modélise une concentration homogène en plancton (ou autre traceur) dans un point de grille donné. L'approche basée individu (ou Lagrangienne) bien que coûteuse en temps de calcul, permet notamment de suivre l'« histoire » de chaque particule (température, taux de croissance, division...) ou encore de modéliser des mutations dans un génome lors de divisions de ces dernières¹⁶⁰. Je décris maintenant brièvement le modèle de croissance utilisé.

Bien qu'ils existent des modèles plus avancées^{161-163,200,203}, nous utilisons ici l'équation de Monod²⁰² qui permet de décrire simplement la croissance d'une bactérie ou d'une algue en fonction de la concentration en nutritif et son affinité pour celui-ci. Il s'agit d'une équation différentielle du premier degré facilement intégrable décrivant la dynamique de croissance d'une population d'algues en fonction du temps. En plus de la dépendance à un élément nutritif, il a aussi été ajouté une dépendance de la croissance à la lumière, à la température ainsi qu'un terme de broutage par du zooplancton, approximé par la concentration totale en phytoplancton. Ainsi pendant leur advection par les courants, les 'supercellules' croissent selon ces différents paramètres (intégrés (pour la lumière) ou moyennés jusqu'à la couche de mélange (nutritif et température)) et se divisent après avoir atteint une masse critique. Le modèle inclut les plus grands types de phytoplancton connus par ordre croissant de taille : les cyanobactéries (*Prochlorococcus* et *Synechococcus*), divers eucaryotes et les diatomées. Les paramètres physiologiques de dépendance aux éléments nutritifs sont ainsi déterminés par des lois allométriques dépendantes de leur taille³⁴¹.

Je présente ici sous forme d'une ébauche d'article les résultats très préliminaires de cette étude. La dynamique globale et locale du modèle est présentée ainsi que ses grandes caractéristiques écologiques telles que la biomasse, la diversité et la biogéographie modélisée. Il est notamment souligné que le terme de broutage est nécessaire à l'obtention d'une biogéographie claire. Un second résultat est la mise en évidence que le nombre de particules est un facteur important déterminant la diversité totale en espèce modélisées. Enfin l'influence des courants sur les niches

des espèces modélisées est quantifiée.

Je le souligne encore, ce modèle et ces résultats restent encore à l'état d'ébauche. Néanmoins ce modèle est un outil intéressant et puissant pour étudier des questions théoriques tel que l'émergence de niche écologique du plancton dans les océans et l'influence des courants sur celle-ci. Il s'avérera ainsi nécessaire dans le futur de valider et optimiser ce modèle en comparaison de données réelles. Un développement intéressant serait l'ajout d'un module d'adaptation (par mutation) dans l'espace des traits physiologiques du phytoplancton (affinité aux nutritifs, niche de température...) ainsi qu'un cycle saisonnier non stationnaire et éventuellement suivant un scénario de changement climatique. L'idée serait ici de comprendre l'impact qu'aurait l'adaptation (absentes des modèles biogéochimiques de large échelle) sur la biogéographie du plancton selon un cycle saisonnier variable ou avec réchauffement climatique. Ainsi l'adaptation à des changements environnementaux rapides peut être rapide^{311,342,343} du fait de l'importante taille des populations de phytoplancton et leur cycle de vie rapide. De plus elle peut permettre le sauvetage partiel de populations adaptées localement³⁴⁴, mitigeant certains effets du changement climatique bien qu'avec certaines limites³⁴⁴⁻³⁴⁶.

II. Résultats préliminaires: Ecological characteristics of a size structured Individual Based Model of a phytoplankton community in an idealized North Atlantic double gyre flow

Cette étude a été dirigée par Daniele Iudicone de la Stazione Zoologica Anton Dohrn de Naples qui m'a accueilli 6 mois dans son laboratoire en 2019. Il a été conçu et réalisé en étroite collaboration avec mes collègues de la Stazione Zoologica Anton Dohrn, Daniele Iudicone, Maurizio Ribera D'alcala et Bruno Hay Mele. J'ai réalisé l'ensemble de la modélisation, des analyses et des figures des résultats préliminaires présentés sous forme d'ébauche d'article.

Ecological characteristics and limits of a size structured Individual Based Model of a phytoplankton community in an idealized North Atlantic double gyre flow

Abstract

Understanding the emergence and dynamics of plankton biogeography in the context of the oceanic circulation and geochemistry is an important and fundamental challenge. Here we develop and Individual Based Model of a phytoplankton community representing major phytoplankton types from cyanobacteria to diatoms. Particles of the model are advected in an idealized double gyre circulation mimicking the North Atlantic currents. We use the model to characterize the emerging biogeographical patterns from different transport and ecological configurations. North-South and East-West biogeographical structuring emerges when diffusivity and the number of modeled particles are low. However, in low diffusivity configuration and when the number of particles is increased we evidence a non-linear increase in species coexistence and a global homogenization of the plankton community at the scale of the whole double gyre domain. We suggest that the number of particles and by extension the number of plankton cells probably increase species coexistence and total richness. Finally we assess the role of currents on the niche temperature of the modeled species showing a relatively weak effect of currents in the case of a structured biogeography.

Introduction

Plankton communities are composed of complex assemblages of viruses, prokaryotes (bacteria and archaea), unicellular (protists) and multicellular eukaryotes (small metazoans) that are passively transported along currents and thrive under multiple types of climate. In particular, phytoplankton composed of cyanobacteria and microscopic algae are responsible for the primary production in the ocean via photosynthesis [1] (with a contribution similar to that of the terrestrial component) on which the rest of the food chain depends. In addition plankton also takes part in the biological carbon pump, the process by which organic matter from the surface sinks to the deep ocean, is stored in the sediments [2, 3], and from which oil is derived [4]. Critically, it has been shown that plankton is not distributed homogeneously in the oceans in terms of biomass [5, 6, 7, 8], diversity [5, 9, 10] and composition [11, 12, 13, 14, 15, 16]. Consequently, it is important to understand the mechanisms that shape the distribution of plankton communities including their thermal niche [17, 34], sizes [19, 20], acclimation [21] and adaptation abilities [21, 22, 23, 24]. These questions are particularly reinforced in the context of climate change caused by anthropogenic carbon dioxide emissions [25]. One of the key tools to explore these phenomena are biogeochemical models [26, 5, 27]. Generally in mechanistic biogeochemical models plankton are modelled using the Eulerian approach [26, 5, 27] so that each grid point of the model simulates a concentration of each plankton type that is non null. Though relatively computationally efficient, the Eulerian approach suffers several limitations. It considers that the tracer (any tracer) is homogeneous within each grid cell, doesn't allow to track water mass or particles trajectories and adaptation processes (*i.e* in case the tracer is plankton) can only be coded as diffusion processes and can not be tracked [28, 29].

In the past and present decade there is an increasing interest in using Individual Based Model (IBM) or particle tracking models (Lagrangian approach) to study phytoplankton ecology [28, 30, 31, 32, 33, 34, 35, 36]. Despite being computationally intensive, the IBM approach allows to track individual histories of particles (trajectories, temperature, growth rate, nutrient supply...), coding point mutations in genomes and model plankton as 'supercells' (*e.g.* each particle represents a large number of phytoplankton cells) that are not necessarily present at each grid point; the two latter assets being more satisfying from a conceptual point of view. Interestingly, the IBM approach has been used to show that neutral evolution is sufficient for the emergence of biogeographical patterns at the genomic level in the context of the ocean global circulation [33].

In a few studies, the IBM approach has been used to assess the role of temperature and currents in shaping plankton ecology [28, 34, 35]. For example, the role of currents in the temperature selection of plankton has been investigated [34]. Important differences between optimal temperature growth with or without advection

with differences of up to 4°C in regions of strong latitudinal flows were shown [34]. In a second example, by tracking the temperature lagrangian history of particles drifting in the ocean currents and in a circulation model it has been shown that currents play an important role in shaping plankton growth rates and in a very different manners than does the Eulerian approach [35]. In another study, numerical experiments of selection of temperature optima for growth on seasonal cycles and environmental change, using both the Lagrangian and Eulerian approach, result in different subpopulation of phytoplankton with different temperature optima [28]. Another major matter of interest in plankton ecological modelling studies is the role of selection and adaptation for nutrient allocation and cell size selection using either the IBM approach [31, 32] or the Eulerian approach [29]. The latter show that trait diffusion in a Nutrient Phytoplankton Zooplankton Detritus (NPZD) model is sufficient for the emergence of a size-structure food web with phytoplankton and zooplankton coevolving in a series of branching events (in the size trait space) [29]. This result in an increasingly discretized size spectrum of both phytoplankton and zooplankton as mutation rates decreases [29]. The previous two [31, 32] show complex relationships and trade-offs between nutrient affinity, resource allocation, cell size and light limitation [32]. Particularly they show that these trade-offs result in an inhomogeneous cell size distribution of picophytoplankton at global scale by combining the IBM approach with a biogeochemical and a global circulation model [32]. Interestingly in these models, some subcellular processes are resolved like the photosynthetic apparatus, the biosynthetic apparatus, and cell structure and the role of bottom up control by nutrient supply in oligotrophic regions on cell size is investigated [32].

To our knowledge and despite these various approaches, there exist no studies that directly investigates at which extent IBM models can represent the diversity of phytoplankton organisms in coupling with a geochemical forcing. In biogeochemical Eulerian models there has been important effort in modelling an increasing number of phytoplankton types [5]. Here we propose to investigate the properties and particularly the diversity emerging from a size structured phytoplankton IBM in coupling with a simple geochemical forcing and circulation model. First we present the main characteristics of our ecological and geochemical models. Main phytoplankton types and sizes are modeled and goes from small photosynthetic cyanobacteria *Prochlorococcus* and *Synechococcus* up to large diatom cells. Particles are advected in an 2 dimensional idealized double gyre flow respecting the North Atlantic dimensions and forced with constant seasonal Mixed Layer Depth (MLD), light, temperature and single nutrient (nitrate) fields. In this context we investigate four main questions: (1) What ecological dynamic emerges from the model? (2) What are the basic ecological features of the model: diversity, biogeography and biomass? (3) What is the impact of the number of particles modeled and of dispersion on the ecological features (particularly diversity) that emerges from the model? (4) What is the impact of currents and diffusivity on the temperature niches of the modeled species?

Methods

Ecological model

Particle growth model

For each particle of the Individual Based Model (IBM) we use a Monod equation of growth generalized to n species resembling the Generalized Lotka-Volterra (GLV) model [37, 38]. Particles are advected in a two dimensional double gyre flow described below. Zooplankton is approximated as the total phytoplankton biomass and all interaction terms are set to 1 in the GLV formulated in equation 2.1 of Pontarp 2017 [38]. The spatialized equation for one particle i of biomass B_i at position x, y and time t is the following:

$$\frac{dB_i(x, y, t)}{dt} = \mu_{max,i} \cdot \psi(T) \cdot \gamma(I) \cdot \beta(N) \cdot B_i(x, y, t) - d \cdot B_i(x, y, t) - \left(\frac{\sum_{Kj} B_{Kj}}{\tau(T, N, I) \cdot K} - g \cdot \eta(T) \cdot \sum_{Kj} B_{Kj} \right) \cdot B_i(x, y, t) \quad (1.1)$$

$B_i(x, y, t)$ is the biomass concentration of particle i expressed in $mmolN.m^{-2}$. d is the basal death rate of phytoplankton set to 0.1. g is the maximal grazing rate set to 0.6 or 0 (in simulations with no grazing) in $(mmolN.m^{-2})^{-1}.d^{-1}$. $\sum_{Kj} B_{Kj}$ in $mmolN.m^{-2}$ is the sum of the concentration of particles located in the control grid cell K . K is a constant in $mmolN.m^{-2}$ set to 1. It approximates the zooplankton concentration (grazing term) and is used in the bottom up control term to set the maximum phytoplankton concentration (see subsection 'Growth equilibrium'). Growth rates depend on the cell size, on the family of the phytoplankton clade (either diatoms, small eukaryotes, *Synechococcus* or *Prochlorococcus*) to which belongs the particle and on three physico-chemical environmental variables: local temperature, available light and available nutrients (nitrate) in the control grid cell (Figure 1a). The combination of the physiological and biological parameters of each particle (phytoplankton family, cell volume, maximum growth rate, half saturation constant for growth and optimal temperature) defined hereafter defines the species to which it belongs.

For the maximum growth rates, we follow the empiric allometric relationship with cell volume from Ward *et al.* 2012 [39] (Supplementary Figure 1a).

$$\mu_{max,i} = a_t \cdot V^b \quad (1.2)$$

a_t depends on the phytoplankton type (Diatoms: 3.8, Small eukaryotes: 2.1, *Synechococcus*: 1.4, *Prochlorococcus*: 1.0). b is equal to -0.15. Each particle is randomly assigned to a clade and to a volume within the range of observed sizes [39] for the given clade. This assignation fixes the particles' species.

For the light dependency $\gamma(I)$ following Evans and Parslow 1985 [40] and Krzystof 2016 [41], we use:

$$\gamma(I) = \frac{2}{H \cdot k} \cdot (F(\Psi \cdot \exp(k \cdot H), \frac{\tau_0}{2}) - F(\Psi, \frac{\tau_0}{2}) - F(\Psi \cdot \exp(k \cdot H), 0) + F(\Psi, \frac{\tau_0}{2})) \quad (1.3)$$

Where F is the following function:

$$F(y, t) = \sqrt{y^2 + t^2} - t \cdot \ln\left(\frac{t + \sqrt{y^2 + t^2}}{y}\right) \quad (1.4)$$

Ψ results from the transformation of vertical coordinate from depth to reciprocal light, given by $\Psi = \frac{\mu_{max,i} \cdot \frac{\tau_0}{2}}{\alpha \cdot J}$ and $y = \Psi \cdot \exp(kz)$ with $dz = \frac{dy}{k \cdot y}$ (see Parslow 1985 [40] and Krzystof 2016 [41] for discussion and details of the calculation). $\frac{\tau_0}{2}$ is half the duration of the day in day (*i.e.* the time in day between the sunrise and the zenith), α is the initial slope of the photosynthesis-irradiance (P-I) curve in $d^{-1}(W/m^2)^{-1}$, J is the light level at the surface at the zenith in $W \cdot m^{-2}$. k is the light attenuation coefficient in m^{-1} accounting for the water light absorption and for the phytoplankton self-shading:

$$k = e_p + e_w \cdot \sum_{jK} B_j(x, y, t) \quad (1.5)$$

e_p is the water light absorption coefficient in m^{-1} , e_w is the phytoplankton light absorption coefficient in $m^2(mmolN)^{-1}$ (self shading effect). Finally in (1.3) H is the Mixed layer depth (MLD) in m (see section Mixed layer depth field for details).

For the temperature dependency $\psi(T)$ (Supplementary Figure c-e) and $\eta(T)$ (grazing dependency to temperature), we follow Dutkiewicz *et al.* [5] by using the Eppley curve [42]:

$$\psi(T) = \tau_T \cdot \exp\left(-A_T \cdot \left(\frac{1}{T} - \frac{1}{T_N}\right)\right) \cdot \exp(-B_T \cdot |T - T_{opt,i}|^{b_T}); \quad (1.6)$$

$$\eta(T) = \tau_T \cdot \exp\left(-A_T \cdot \left(\frac{1}{T} - \frac{1}{T_N}\right)\right); \quad (1.7)$$

$T_{opt,i}$ is the optimal temperature of species i . τ_T normalizes the maximum value. A_T , T_N , B_T , and b_T regulate the sensitivity envelope (see Table 1 for values). More precisely, τ_T and T_N regulate the absolute values of $\psi(T)$, A_T regulates the sensitivity to temperature, B_T regulates the niche width and b_T controls the smoothness of the niche borders. For the simulations, each species is randomly assigned to one of ten thermal norms linearly spaced between -2 and 34°C.

For the nutrient dependency $\beta(N)$ (Supplementary Figure b,d), we use the Monod equation of growth [43]:

$$\beta(N) = \frac{N(x, y, t)}{N(x, y, t) + N_{c,i}} \quad (1.8)$$

Where $N(x, y, t)$ is the concentration in nitrate at location x, y and time t . $N_{c,i}$ the half saturation constant for growth of particle i (associated to a given species):

$$N_{c,i} = \frac{K_i \cdot \mu_{max,i} \cdot Q_{min,i}}{V_{max,i}} \quad (1.9)$$

Where K_i is the cell nutrient uptake half saturation constant, $Q_{min,i}$ is the cell minimum stoichiometric quota, $V_{max,i}$ is the maximal cell nutrient uptake rate. For all three parameters, an empirical allometric relationship with the cell volume V has been observed [5, 44]:

$$K_i = a_K \cdot V^{b_K}; \quad Q_{min,i} = a_Q \cdot V^{b_Q}; \quad V_{max,i} = a_V \cdot V^{b_V}; \quad (1.10, 1.11, 1.12)$$

Growth equilibrium

Assuming that the carrying capacity of the system is the absolute value of the local nutrient concentration minus the local theoretical nutrient equilibrium concentration [45], the equilibrium of (1) for any particle i in a control grid cell K is:

$$\sum_{Kj} B_{Kj} = \frac{\mu_{Kj,max}(N, T, I) - d}{\frac{1}{\tau(N,T,I) \cdot K} - g \cdot \eta(T)} = |N - N^*| \quad (1.13)$$

From (1.13) we deduce the characteristic time to reach equilibrium $\tau(N, T, I)$ in day:

$$\tau(N, T, I) = \frac{|N - N^*|}{K \cdot (\mu_{Kj,max}(N, T, I) - (d - g \cdot \eta(T) \cdot |N - N^*|))} \quad (1.14)$$

$\mu_{Kj,max}(N, T, I)$ is the maximum net growth rate among all particles in control grid cell kj that depends on $\mu_{max,i}$, $\beta(N)$, $\gamma(I)$ and $\psi(T)$. We use $\mu_{Kj,max}(N, T, I)$ as we assume that the phytoplankton type with the maximal growth rate drives the dynamic of nutrient consumption. In simulations we use $|N - N^*|$ to ensure that τ is never negative. The absolute value $|N - N^*|$ is used as the equilibrium because in the case nutrients are consumed, phytoplankton can not consume entirely nutrients but instead draw down the nutrient concentration to N^* . N^* is given by:

$$N^* = \frac{N_c \cdot (d + g \cdot \eta(T) \cdot \sum_{Kj} B_{Kj})}{\mu_{Kj,max} - (d + g \cdot \eta(T) \cdot \sum_{Kj} B_{Kj})} \quad (1.15)$$

Parameter	Symbol	Unit	Value	Reference
Basal mortality of phytoplankton	d	d^{-1}	0.1	-
Maximal grazing rate	g	$(mmolN \cdot m^{-2})^{-1} \cdot d^{-1}$	0.6 or 0	-
Water light absorption coefficient	e_p	m^{-1}	0.03	Fasham <i>et al.</i> 1990 [47]
Phytoplankton light absorption coefficient	e_w	$m^2(mmolN)^{-1}$	0.04	Fasham <i>et al.</i> 1990 [47]
Initial slope of the photosynthesis-irradiance (P-I) curve	α	$d^{-1}(W/m^2)^{-1}$	0.1	Geider <i>et al.</i> 1997 [48], Stawiarski <i>et al.</i> 2018 [49]
Activation temperature	A_T	K	4000	Dutkiewicz <i>et al.</i> 2020 [5]
Envelope sensitivity	T_N	K	293.15	Dutkiewicz <i>et al.</i> 2020 [5]
Temperature growth dependency normalization	τ_T	unitless	0.8	Dutkiewicz <i>et al.</i> 2020 [26]
Niche width regulator	B_T	K^{-1}	0.0003	Dutkiewicz <i>et al.</i> 2020 [5]
Niche skewness regulator	b_T	unitless	4	Dutkiewicz <i>et al.</i> 2020 [5]

Table 1 | Fixed parameters regulating plankton growth independently from their size.

Particle biomass concentration, duplication and death

The representative initial biomass concentration in a control grid cell K of particle i is:

$$B_{Ki0} = \frac{m_0}{S} \quad (1.17)$$

Parameter	Symbol	a_X	b_X	Unit	Reference
Maximum growth rate	μ_{max}			d^{-1}	
Diatoms		3.8	-0.15		Ward <i>et al.</i> 2012 [39]
Small eukaryotes		2.1	-0.15		Ward <i>et al.</i> 2012 [39]
<i>Synechococcus</i>		1.4	-0.15		Ward <i>et al.</i> 2012 [39]
<i>Prochlorococcus</i>		1.0	-0.15		Ward <i>et al.</i> 2012 [39]
Cell nitrate uptake half saturation constant	K	0.17	0.27	$mmolN.m^{-3}$	Zakem <i>et al.</i> 2018 [44]
Minimum cell quota relative to C	Q_{min}	0.07	-0.17	$mmolN$ $mmolC^{-1}$	Zakem <i>et al.</i> 2018 [44]
Maximum nutrient uptake rate	V_{max}	0.51	-0.27	$mmolN$ $mmolC^{-1}.d^{-1}$	Zakem <i>et al.</i> 2018 [44]

Table 2 | Size dependent parameters regulating plankton growth following allometric laws ($a_X V^{b_X}$).

S is the surface of the grid cell K in m^2 . Therefore in a control grid cell K the total biomass is defined as follow:

$$B_{Ktot} = \sum_i B_{Ki}(x, y, t) = \frac{1}{S} \sum_{Ki} m_{Ki}(x, y, t) \quad (1.18)$$

Integrating equation (1) using the forward Euler method with a biological time step of half a day, particles duplicate following the duplication condition:

$$B_i \geq \frac{2.m_0}{S} \quad (1.19)$$

A particle is erased from the set of particles following the death condition:

$$B_i \leq thresh \quad (1.20)$$

In practice we use $thresh = 1.5.10^{-5} mmolN.m^{-2}$ corresponding approximately to 100 cell of *Prochlorococcus* per liter [5].

Idealized North Atlantic double gyre flow

Main flow

We build an idealized double gyre flow consisting of two principle gyre (north and south) with the dimensions of the real North Atlantic double gyre and static eddies. The well known and extensively studied double gyre equation [46, 50, 51, 52, 53] is used to mimic the double gyre. The speed potential is given by:

$$\Psi_{dg}(x, y) = A_{dg} \cdot \sin(\pi \cdot \omega_x \cdot x) \cdot \sin(\pi \cdot \omega_y \cdot y) \quad (2.1)$$

From (2.1), respectively meridional and zonal flow speeds are derived:

$$u_{dg} = \frac{\partial \Psi_{dg}(x, y)}{\partial y} = -A_{dg} \cdot \pi \cdot \omega_y \cdot \sin(\pi \cdot \omega_x \cdot x) \cdot \cos(\pi \cdot \omega_y \cdot y) \quad (2.2)$$

$$v_{dg} = -\frac{\partial \Psi_{dg}(x, y)}{\partial x} = A_{dg} \cdot \pi \cdot \omega_x \cdot \cos(\pi \cdot \omega_x \cdot x) \cdot \sin(\pi \cdot \omega_y \cdot y) \quad (2.3)$$

ω_x is the spatial frequency of the gyre in m^{-1} (*i.e.* it characterize its size in the longitudinal direction), here $\omega_x = \frac{1}{5.10^6}$. It corresponds approximately to the width of the North Atlantic Ocean between the french city of Bordeaux to the Canadian province of Nova Scotia. ω_y is the equivalent in the latitudinal direction, here $\omega_y = \frac{1}{5.8.10^6}$. It corresponds approximately to half the distance of a transect between the Senegal upwelling east

of Dakar and the north of Iceland (54° of latitude). A_{dg} characterizes the maximum speed of the double gyre velocity field, $A_{dg} = \frac{u_{max}}{\pi \cdot \omega_x}$, we set $u_{max} = 0.15 \text{ m.s}^{-1}$. In addition to the main double gyre, we approximately model mesoscale eddies using the same equations with different scale parameters. Eddies are here static.

$$\Psi_e(x, y) = A_e \cdot \sin(\pi \cdot \omega_{ex} \cdot x) \cdot \sin(\pi \cdot \omega_{ey} \cdot y) \quad (2.4)$$

$$u_e = \frac{\partial \Psi_e(x, y)}{\partial y} = -A_e \cdot \pi \cdot \omega_{ey} \cdot \sin(\pi \cdot \omega_{ex} \cdot x) \cdot \cos(\pi \cdot \omega_{ey} \cdot y) \quad (2.5)$$

$$v_e = -\frac{\partial \Psi_e(x, y)}{\partial x} = A_e \cdot \pi \cdot \omega_{ex} \cdot \cos(\pi \cdot \omega_{ex} \cdot x) \cdot \sin(\pi \cdot \omega_{ey} \cdot y) \quad (2.6)$$

ω_{ex} and ω_{ey} are the spatial frequencies of the eddies in m^{-1} (*i.e.* it characterizes eddies size in the longitudinal and latitudinal direction), here $\omega_{ex} = \omega_{ey} = \frac{1}{250 \cdot 10^3} A_e$ characterizes the maximum speed of the eddies velocity field, $A_e = \frac{u_e \text{ max}}{\pi \cdot \omega_{ex}}$, we set $u_{max} = 0.15 \text{ m.s}^{-1}$. The global speed field is then the addition of the double gyre and the eddies.

$$u = u_{dg} + u_e \quad (2.7)$$

$$v = v_{dg} + v_e \quad (2.8)$$

Additioning the two field allows to reach a maximum speed of approximately 0.3 m.s^{-1} which is what is observed in the real oceans except in the Gulf Stream where higher speeds are reported [54].

Diffusion

In addition to the main velocity field, we model small scale turbulence by adding a 2 dimensional Brownian motion to the particles. Brownian motion is modeled as follows: at each time step the particle is transported in a random direction of the 2 dimensional grid according to the diffusion equation with a diffusion coefficient D . Accordingly, we have at any integration time step t :

$$x(t+1) = x(t) \pm U(0, 1) \cdot \sqrt{\frac{2 \cdot D \cdot dt}{r}} \quad (2.9)$$

$$y(t+1) = y(t) \pm U(0, 1) \cdot \sqrt{\frac{2 \cdot D \cdot dt}{r}} \quad (2.10)$$

Where $U(0, 1)$ is the uniform distribution between 0 and 1. r is equal to $\frac{1}{3}$ to account for the bi-dimensional nature of our field [55]. This renders the global field described in (2.7) and (2.8) stochastic and particle dependent. For a particle i we finally have:

$$u_{i(x,y)} = u_{dg} + u_e + u_b \quad (2.11)$$

$$v_{i(x,y)} = v_{dg} + v_e + v_b \quad (2.12)$$

By derivating over time (2.9) and (2.10) we have $|u_b| = |v_b| \sim U\left(\frac{1}{2 \cdot dt} \cdot \sqrt{\frac{2 \cdot D \cdot dt}{r}}, \frac{2 \cdot D}{12 \cdot r \cdot dt}\right) = U\left(\sqrt{\frac{3 \cdot D}{2 \cdot dt}}, \frac{D}{2 \cdot dt}\right)$, where U is the uniform distribution. With a diffusion coefficient of $1000 \text{ m}^2 \cdot \text{s}^{-1}$, the average speed due to diffusion is around 0.13 m.s^{-1} and the standard deviation is around 0.076 m.s^{-1} (for a time step of 1 day *i.e.* 86400 s). In practice we use 4 values of D : 10, 100, 1000 and $10000 \text{ m}^2 \cdot \text{s}^{-1}$.

Particles transport

Particles grow and are advected and diffused in a Lagrangian framework using the python package `Parcels v2.0` [55] within a 2 dimensional grid of 126 over 146 (square grid point of approximately 40x40 kms). For integration over time we build a custom kernel consisting of the addition of 3 kernels: `AdvectionRK4 Parcels v2.0` for the advection (RK4 standing for the 4th order Runge-Kutta scheme), the `2D BrownianMotion2D` diffusion kernel from Delandmeter 2019 [55] and a custom boundary kernel at the edges of the double gyre. This custom boundary kernel simply brings back the particle in the flow field by setting its coordinates 10 meters away from the boundary. We visualize example of particle trajectories for the four diffusivity kernel used (Figure 1c-d and Supplementary Figure 2).

Physical and geochemical forcings

Light field

For the light dependency of growth, three parameters are needed: the light irradiation J , the day length τ_0 and the Mixed layer depth (MLD). The two first depends only on latitude and day of the year while the last one depends both on the latitude, the longitude and the day of the year. We use the same radiation equation as Krzystof 2016 [41]:

$$J(t, lat) = I_0(t) \cdot \text{cosineeffect}(t, lat) \cdot f_{\text{par}} \cdot (1 - \text{albedo}) \cdot C_{\text{fac}} \quad (3.1)$$

Where $I_0(t)$ is the solar radiation at day t of the year that depends on the earth-sun distance [56]:

$$I_0(t) = 0.7 \cdot I_{\text{sc}} \cdot (1 + 0.034 \cdot \cos \frac{2 \cdot \pi \cdot t}{365}) \quad (3.2)$$

I_{sc} is the solar constant (1367 W/m²). The cosine effect $\text{cosineeffect}(t, lat)$ takes into account the variation inclination of irradiance rays with latitude and day of the year:

$$\text{cosineeffect}(t, lat) = \cos \frac{\alpha(t, lat) \cdot \pi}{180} \quad (3.3)$$

Where $\alpha(t)$ is the solar elevation of the sun at zenith dependent on day of the year and latitude [57]:

$$\alpha(t, lat) = 90 - \arcsin \left(\sin \frac{\delta(t) \cdot \pi}{180} \cdot \sin \frac{\text{lat} \cdot \pi}{180} + \cos \frac{\delta(t) \cdot \pi}{180} \cdot \cos \frac{\text{lat} \cdot \pi}{180} \right) \cdot \frac{180}{\pi} \quad (3.4)$$

$\delta(t)$ is the solar declination along the year from Spencer 1971 [58]:

$$\delta(t) = 0.39637 - 22.9133 \cos \psi_d + 4.02543 \sin \psi_d - 0.3872 \cos 2\psi_d + 0.052 \sin 2\psi_d \quad (3.5)$$

ψ_d is the day of the year expressed as an angle: $\psi_d = \frac{360 \cdot t \cdot \pi}{365 \cdot 180}$, t being the day of the year. $f_{\text{par}} = 0,43$ is the fraction of solar irradiance between 400 and 700 nm (used for photosynthesis). $\text{albedo} = 0.04$ is the albedo of the sea surface. C_{fac} is the effect of cloud on the atmosphere transmission.

$$C_{\text{fac}} = 1 - 0.0775W + 0.0019(90 - \alpha(t, lat)) \quad (3.6)$$

$W = 6$ is the cloud fraction in oktas.

Second the daylength [57] is formulated as:

$$\tau_0 = 0.133 \arccos \left(-\tan \frac{\text{lat} \cdot \pi}{180} \tan \frac{\delta(t) \cdot \pi}{180} \right) \frac{180}{\pi} \quad (3.7)$$

Mixed Layer Depth field

The biomass of a given particle in a grid square represents a fraction of the concentration of a species within this grid integrated over the Mixed Layer Depth. The oceanic Mixed Layer is the part of the surface ocean that is considered well mixed and homogeneous either for temperature or sigma-t (density). Here we use a regular 2°C×2°C grid temperature based global scale climatology from the IFREMER [59]. Monthly means in the considered area of the North Atlantic are used. For the southern gyre, we use datapoints within the rectangle delimited by isolatitudes 12°N to 38°N and by isolongitudes -16°E to -68°E. Similarly for the northern gyre, we use isolatitudes 40°N to 66°N and isolongitudes -10°E to -62°E. First to fill gaps where there is land in reality but not our gyre, we use the `na.approx` function from the R package `zoo` [60] latitudinally then longitudinally (using the transpose of the matrix). Then to increase the resolution to the 126x146 grid we linearly interpolate the whole grid using the `interp.surface.grid` function from the R package `fields` [61]. Finally, we use the `linspace` function from R package `pracma` [62] to interpolate linearly the MLD each day of the year and at each grid points. This Mixed Layer field is used to compute the temperature field, the nutrient field and the light dependency for growth.

Temperature and Nitrate fields

In our model, both temperature and nitrate are independent from the flow field. To better match real *in situ* data, we use the $1^\circ \times 1^\circ$ monthly data from the World Ocean Atlas 2018 [63, 64] as forcing fields. For temperature, we use the 2005-17 climatology, while for nitrate all decades are used. The same North Atlantic boundaries as for the MLD field are used. We use exactly the same steps as for MLD for each depth of the WOA18: first missing values from land are extrapolated, then the resolution is increased to the 126×146 grid and values are extrapolated to the whole year. Finally, weighted mean (taking into account in-homogeneous depth intervals from WOA18) are calculated using the MLD field which renders the final 2D fields of temperature and nitrate.

Sampling points

We sample the concentration of each phytoplankton type in the northern gyre, the southern gyre and in a total of 12 locations in the double gyre flow. Each sampling station is a square of 4×4 grid points (Figure 1b). The concentration of a given species at a sampling point is defined as the sum of the biomass concentration of each particle belonging to this species within this square and divided by the number of square in the sampling point (16).

Biogeographies: red green blue PCoA

We compute for each pair of grid point of the domain (restricted to a 73×63 grid to reduce computation time) a Bray-Curtis distance [65] on the vector of the biomass of each species. We then perform a Principal Coordinate Analysis (PCoA) of this dissimilarity matrix. Each grid point is then assigned a color whose red, green and blue components are based on its coordinates on the three first axis of the PCoA:

$$x_i = 128.(1 + \lambda_i.x_i/\max(\text{abs}(x_i))) \quad (4.1)$$

Where x_i is the coordinate vector on the i^{th} eigenvector of the PCoA and λ_i is the ratio of the eigenvalue of eigenvector i to the dominant eigenvalue.

Results

Global and local dynamics: influence of transport and diffusivity

In this section, we analyze the dynamics of the ecological behavior of the model *i.e.* the temporal variation of modeled species concentrations. We analyze this behavior for the no transport simulation and the four diffusivity kernel used (D10, D100, D1000 and D10000) at the scale of whole gyres (north and south) (Figure 2 and Supplementary Figure 3-4) and at localized sampling points (Figure 1b, Supplementary Figure 5-6). These results are for the simulations with the highest number of particle we could model. Dynamics over the 10 years of simulation is shown in supplementary Fig. 3. Most species are out competed as soon as the first year and some more slowly especially in high diffusivity simulation (Supplementary Fig. 3). First, we comment the global annual cycle of each gyre during the last year of simulation (10^{th} year).

In the no transport simulation (Figure 2a), in the northern gyre, we observe the beginning of the bloom *i.e.* a rise in phytoplankton biomass around the 55th day of the year (end of February). We then observe three successive peaks of phytoplankton concentration respectively around the 120th, 150th and 170th day of the year (respectively end of April, May and mid-end of June). Each successive peak is higher than its predecessor. The first peak is dominated by *Prochlorococcus* (purples) and *Synechococcus* species (reds), the second and third by diatoms (greens) followed by small eukaryotes (blues) for the third one. Finally at the end of the year, after two small autumn phytoplankton concentration peaks, the cyanobacteria species that dominated the spring bloom become dominant again.

In the three simulations with transport (main double gyre circulation) and low diffusivity (D10, Figure 2b, D100, Figure 2c and D1000, Figure 2d), we observe a synchronization in the mean concentration of all present species in the northern gyre. The phytoplankton is dominated by a single *Prochlorococcus* species followed by a mix of diatoms, small eukaryotes and other *Prochlorococcus*. In the southern gyre, the same rank of abundance in phytoplankton species is observed (Supplementary Fig. 4b-d) so there is a global homogenization of the phytoplankton community in these three simulations. The *Prochlorococcus* species takes over at the end of the first year of simulation as it peaks in autumn and then takes over the northern gyre (Supplementary Fig. 3a). This single event might explain the synchronization observed afterwards. We will further discuss this result but we basically propose that it is due to (1) this invasion event which might be linked to (2) the number of particles modeled (see below) and generating (3) a sort of "snowball" effect where the dominant species "forces" the dynamics of the others (and so on with the second dominant species). In addition, such a mechanism would allow species to grow with low net growth rates and survive in environments that don't correspond to their

niches. In the northern gyre, the various peaks in phytoplankton concentration across the year are also visible though less marked as all species are synchronized. We also notice that all peaks are slightly delayed in time (of around 15 days) and around ten times higher in mean concentration of phytoplankton compared to the no transport simulation. Consequently the summer bloom peaks at the beginning of July. These results suggests that transport allows escaping grazing resulting in a better utilization of resources.

In the higher diffusivity simulations (D10000), Figure 2e), fewer species survive and their concentration dynamics are not all synchronized. The first two peaks of phytoplankton concentration are dominated by two diatoms. The summer peak is due to a small eukaryote species (though not dominant in the north) that dominate in the summer gyre at this time (Supplementary Fig. 3e) suggesting an effect of the strong diffusivity from one gyre to another. The small eukaryote that dominated this peak in the D1000 simulation is also present here but at low concentrations. In the southern gyre, as expected, a lesser seasonality is observed and phytoplankton mean concentration is globally lower compared to the bloom peaks in the north (Supplementary Fig. 4). In opposition to the low diffusivity simulations, different species alternatively dominate: in winter, spring and autumn, a *Prochlorococcus* (that globally dominate in D10, D100 and D1000) dominates while in summer the small eukaryote species that peaks in the north at this time dominates. Overall, these results suggests that diffusion strength allows "winner" species to colonize large parts of the double gyre domain resulting in the exclusion of other species. This hypothesis is reinforced by an overall increased biomass when diffusivity strength increases. Indeed, it shows that nutrient are better utilized and escaping from grazing is also enhanced when diffusion is stronger.

Zooming on specific points of the double gyre (points 4 and 7, Figure 1c), the same characteristics as the whole gyre mean concentrations are observable although with much more noise in the dynamic (Supplementary Fig. 5-6). For D10, D100 and D1000, the *Prochlorococcus* species globally dominates while other species show some turnover between each other (Supplementary Fig. 5b-c and 6b-c). For the D10000, in the southern gyre point (point 4), the *Prochlorococcus* species dominate and coexist with the small eukaryote species, the latter following a north-like seasonal cycle (Supplementary fig. 6d-e). In the north gyre point (point 7), in D10000, the same diatom species is always dominant. This strengthens the idea that high diffusivity allows 'winner' species to colonize and dominate in large areas.

Simulations with no grazing display similar behaviors than when grazing is on with a few changes. In the D10, D100 and D1000 simulations, all species dynamics are also synchronized (Supplementary Fig. 7b-d) but the relative difference in abundance between the first and following species is less marked. This suggests that when grazing is on, the 'winning' species is also imposing a strong grazing on other species (given our simple grazing formulation, see Methods). In addition, much higher biomass concentration are observed, the spring peak is less visible while the autumn peak is more marked compared to the simulations with grazing (Supplementary Fig. 7). Finally, in the D10000 simulation, a small eukaryote instead of a diatom dominates the northern gyre (Supplementary Fig. 7e).

Ecological characteristics of the model: diversity, dominant species, biogeography and biomass

With regard to the previous analyses, we now explore the mean ecological characteristics of the model throughout the last year of simulation (Figure 3). First, we comment the local diversity patterns *i.e.* the number of coexisting species at a given location. In the no transport simulation, up to 6 species are coexisting at a single location (in the northern gyre) and the global pattern of diversity is mainly latitudinal and doesn't follow the shape of the gyre (Figure 3a). At the extreme north of the domain, no species survives no matter whether grazing is turned on or not although grazing increases this area (Figure 3a-b and Supplementary Figure 8a-b). In the D10, D100 and D1000 simulations, from up to 33 (D10) to 28 (D1000) species are coexisting together at a given location and at several diversity hot spots of the double gyre. These hot spots include the center of the northern gyre, the front between the two gyres and the Senegal upwelling. Globally, the diversity pattern seem to be shaped by the double gyre circulation and the strength of the diffusivity (Figure 3 e,i,m). Interestingly, even though almost all species are synchronized in these simulation, we still find a diversity pattern so each species is not present over the whole double gyre domain. This reinforces the idea of a snowball effect where the dominating species (Figure 3f,j,n) "controls" or "forces" the global dynamics of each species. In the D10000 simulation, the mean maximum number of coexisting species is found at the center of the northern gyre and reaches 6.5. while the minimum is found at the center of the southern gyre and is around (2.2) (Figure 3s).

We now comment mean dominant species patterns *i.e.* the most abundant species at each grid point. In the no transport simulation (Figure 3b), the observed pattern globally follows the physiology of the organisms. Diatoms and small eukaryotes, which have the lowest nutrient affinities but the highest maximum growth rates, are found dominant in the north while a mix of *Prochlorococcus* and *Synechococcus* species dominate in the south. This result is clearer in the simulations with no grazing (Supplementary Fig. 7b). In the D10, D100 and D1000 simulations, as already observed, a single *Prochlorococcus* dominates the whole domain (Figure 3f,j,n). In the D10000 simulation, a clear pattern of dominant species is observed. A diatom species dominates the northeastern part of the northern gyre while a small eukaryote dominates its northwestern part. From the Senegal upwelling

to the double gyre front, a small eukaryote also dominate. Finally, the *Prochlorococcus* species that takes over the whole domain in the other simulations, dominates the major part of the southern gyre.

Using a red/green/blue PCoA (rgb PCoA) visualization method (see Methods), we now investigate whether biogeographical pattern at the community level emerges from our simulations. In the no transport simulation, while there is a strong regional heterogeneity, a fully latitudinal biogeography emerges (Figure 3c). In the D10, D100 and D1000 simulations (Figure 3g,k,o), the pattern mainly reflects variations in biomass (Figure 3h,i,p) because the community is relatively homogeneous and dominated by a single species. Therefore, it is impossible to distinguish a clear biogeographical pattern here. In the D1000 simulation (Figure 3s), the rgb PCoA globally recapitulates the dominant species map (Figure 3r) with a grey zone (the center of the PCoA axis) found at the heart of the northern gyre where high local diversity is found (Figure 3q). In addition, the eastern part of the domain, from the Senegal upwelling to the east of Iceland is relatively homogeneous. Overall these results are surprising as we find a clear biogeographical pattern only in this D10000 simulation (in simulations with transport). However, patterns of biogeography are found in other transport configurations (lower diffusivity) when the number of particles modeled is lower (see below). In terms of biomass, the highest phytoplankton concentration is found in the Senegal upwelling zone (Figure 3d,h,l,p,t) and transport and increased diffusivity allows an enhanced spread and increase of the biomass (Figure 3h,l,p,t versus Figure 3d). We note that in no grazing simulations, total biomass is largely increased (by a factor of 10, Figure 3 versus Supplementary Fig. 8). In addition, when grazing is turned off, total biomass is the highest in the no transport simulation. This comfort the idea that diffusion and transport allows escaping grazing in the no transport simulation (with grazing turned on). We compile the seasonal dynamics of the ecological characteristics we just described on a website (https://gigaplankton.shinyapps.io/gifs_phytoIBM_double_gyre/). The visualization of the invasion of a single dominant species and synchronization with other species over the whole double gyre domain at the end of the first year and during the second year of simulation is available (Invasion GIF tab of the website). We now discuss the effect of the number of particles in the simulations on the emerging diversity of the model as well as local coexistence of species.

The number of particles: a driver of diversity and coexistence?

In this Individual Based Model (IBM), we can vary the number of particles that thrive in the double gyre by changing the basal biomass m_0 that each particle represent (and so the biomass at which a particle divide). We investigate the effect of increasing the number of particles in total and local diversity that emerges from the model. When m_0 is high (and consequently the number of particles is low), a biogeographical pattern (Supplementary Fig. 9) is observed for the simulations with low diffusivity kernel (D10, D100 and D1000). However when the number of modeled particles is increased (by decreasing the basal biomass m_0), we observe a non linear effect on the total number of species modeled and the biogeographical pattern observed (Figure 4a,b). This threshold seem to be diffusivity dependent *i.e.* it increases as diffusivity increases. It is not reached for the D10000 simulation (Figure 4a,b). Increasing the number of particles is therefore very different than increasing the diffusivity of the particles and this effect is even counterbalancing the effect of diffusivity as it increases when diffusivity is lower. Increasing the number of particles also increases the maximum and minimum local richness (species coexistence) observed (Figure 4c,d). We also note that these latter local maximum and minimum richness is higher in the D10000 simulation (at high m_0). The non linear transition in the behavior of the system from low richness to high richness is very hard to explain as we fixed a temperature niche to each species so they should not even spread to the whole domain as observed (see next session for discussion). Intuitively, increasing the number of particles result in an increase in the number of possible trajectories to which the particle will survive. We hypothesize that this is the mechanism that allows a few species to spread all over the double gyre. In addition, if one species type is able to colonize the whole gyre then as all other species depend on it, it will "force" the global behavior of all other species at the global scale of the double gyre. Similar results are obtained when the grazing term is removed except that total diversity starts decreasing for the highest number of particles modeled while maximum and local richness still increase or stay steady (Supplementary Fig. 10). Overall these results are nevertheless very surprising. We further try to explain it by analyzing the temperature niches of surviving species and the effect of currents on the temperature niche of these species.

Temperature niche selection and influence of currents on the temperature niche

We now examine the temperature niches of the organisms that compose the biogeographies of the different simulations. In the D10, D100 and D1000 simulation, the *Prochlorococcus* species that dominate the whole domain has an optimal temperature growth of 26°C explaining why it dominates in the southern gyre in the first year of these simulations and in the D10000 simulation (only in the southern gyre). Over the global gyre, the mean temperature over the year is 17.3°C and the mean temperature limitation for this species is 0.49, close to the maximum mean temperature limitation (0.53) for the species with an optimal temperature growth of 22°C (which is the case for the species ranked second and 4th globally). Therefore, optimal growth temperature of the

winning species seem to be particularly well adapted for the double gyre domain used here. The distribution of the selected maximum growth rates and optimal temperature growth for the different simulations are plotted in Supplementary Figure 10. In every simulation, species with the maximum growth rates are selected in each taxonomic subcategories (Supplementary Figure 10a,c,e,g,i). Lower growth rate species only survive in high diversity simulation (no transport and low diffusivity). However, in terms of optimal temperature growth, a bias towards low temperatures (D10 and D100) and towards intermediate-low temperatures (D1000) is observed. This result might be due to the higher nutrient supply in the northern gyre supporting more growth and diversity in the north. We note that species with extreme values of optimal growth temperature are rarely selected. In our double gyre set up, the particles trajectories are probably more often in the center of the domain (between the two gyres) so towards intermediate temperature values. This latter observation probably participate in the homogenization of the phytoplankton community observed for the low diffusivity particles (D10, D100, D1000). This is confirmed by plotting the niches centroids of the different species (Figure 5a,d,g,j,m). In the D10, D100 and D1000 simulations, almost all centroids are located at the same place around the center of the double gyre domain (Figure 5a,d,g). This renders niches centroids far apart from the no transport simulations (Figure 5b,e,h), high temperature difference compared to the no transport (Figure 5c,f,i) and optimal growth temperature (Supplementary Fig. 11). Therefore advection here seem to be the principal driver of the niches of surviving species with a flow of particles (in all directions, as the number of particle is high) sufficient to maintain species present over the whole double gyre domain, far from their optimal niches. Even in the no transport simulation, high difference of temperature up to 6.3°C (and -6.3°C) between the observed temperature niche and the optimal temperature growth is observed (Supplementary Figure 11a). This is however allowed by or temperature niche parametrization (Supplementary Figure 1c) and optimal growth temperature follow the temperature gradient of the double gyre (Figure 5m). In the D10000, these differences are surprisingly lower and range only from -2.4°C to +2.0°C (Supplementary Figure 11e) so in this case the niche centroids follow the temperature gradient of the double gyre (Figure 5j,k,l).

Discussion

We successfully implemented an Individual Based Model of a phytoplankton community representing major taxons (cyanobacteria, small eukaryotes and diatoms) with a physical (temperature, light) and geochemical (nutrient limitation) forcing in an idealized double gyre circulation of the North Atlantic. Interestingly, we observe that the model display a non-linear transition in its behavior when the number of particles is increased and the diffusivity of particles is sufficiently low. Indeed, when the number of particles is low, the model display a clear north-south and east-west (especially in the northern gyre) biogeographical pattern with a few small eukaryotes and diatoms dominating in the northern gyre and a few cyanobacteria dominating in the southern gyre). However when the number of particles is increased, a high number of species coexist and show a synchronicity in their dynamical behavior at the scale of the whole double gyre. This renders a globally homogeneous phytoplankton community over the whole double gyre domain. According to the literature, stable biogeographical patterns are expected for plankton communities at the global scale [12, 13, 14, 5], so intuitively the latter behavior of the model does not seem to give the expected structure. Several aspect of the model might explain this behavior: (1) the double gyre domain is closed so there is no outflows at the north and south of the domain which probably favors species with intermediate-low temperature optimal growth (which is observed), (2) the integration time step (1 day) might not be sufficient to capture a sufficiently precise dynamics and the spatial resolution of bottom up and top down control might be too low, (3) the model runs on a constant seasonal cycle and (4) the model is relatively simple and could be complexified with additional nutrient dependencies for instance or a different grazing term. Regarding (1), this seems to be an important issue as the species that invades the whole domain becomes quite abundant in the northern gyre as particles have looped through the northern gyre (with no outflows) in the north and so lot of particles are found around the middle of the double gyre domain at the end of the first year of simulation with favorable condition for its growth. Regarding point (3), this constant seasonal cycle might bias the 'winning' species as a few species of the total species pool (when no adaptation is possible) are probably on average the best adapted to this seasonal cycle.

Despite these limitations, our results nevertheless suggests that the number of particles represented in our IBM is critical. We propose that the number of supercells and by extension the number of plankton cells can be a driver of total richness as well as of local coexistence in plankton population. Intuitively increasing the number of particles augments the chance for a particle to follow a trajectory that will allow its survival. In population ecology the paradox of the plankton is a long standing issue [66, 67, 68] and several explanations have been proposed [5, 67, 68]. To our knowledge the number of plankton particles has never been proposed as a mechanism of coexistence despite being well known to be a driver of intra-species diversity in the field of population genetics [70, 69]. Our suggestion might also further explain the latitudinal diversity gradient observed in plankton communities [9]. Indeed, smaller cells are found more abundant in low latitudes, they are however much more numerous in number than bigger cells found at high latitudes [20]. Therefore this large number

of cells might be a driver of high diversity in low latitudes. Here using our model we don't reconstruct the latitudinal diversity gradient though we find higher diversity at fronts as expected [5]. Nevertheless, imposing a size dependent basal biomass for the particles of our model might further confirm this hypothesis. This might however be tricky to handle as it would probably further increase the advantage of cyanobacteria over larger small eukaryotes or diatom supercells.

Another limitation of our approach is the advection field within which the supercells are advected. It is only two dimensional, there is no northern and southern outflows and inflows (unlike in the real ocean), eddies are static and finally diffusion is equal over the whole domain although it is well known that it is not the case [71]. Therefore the model could be refined by imposing a spatial heterogeneity in the diffusivity constant [71], modelling moving eddies [72] or even advect the particles in a global circulation model [5, 73].

Finally, we demonstrated that our ecological IBM framework could be a strong tool to answer many questions of interests regarding phytoplankton ecology. In particular, the ecological model could be refined by complexifying the grazing module (by explicitly representing the zooplankton or add size dependent grazing [29]), add other limiting nutrients [5], add a viral lysis model [74, 75] and finally add an adaptation module to the phytoplankton [29, 28]. The latter could be coded as simple diffusion in the trait space [29, 28] or even by coding mutation in a genome [34]. Indeed to our knowledge, the effect of adapting species in a global phytoplankton community has never been addressed. Finally, in the context of a changing climate [14, 9, 10, 76], a temperature increase as well as changes in the geochemistry projections of the double gyre could be imposed by projections from global biogeochemical models [77].

References

- [1] Field, C. B., Behrenfeld, M. J., Randerson, J. T., Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science (80-.)*. (1998). doi:10.1126/science.281.5374.237
- [2] Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* (2016). doi:10.1038/nature16942
- [3] Henson, S. A., Sanders, R., Madsen, E. Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochem. Cycles* (2012). doi:10.1029/2011GB004099
- [4] Schobert, H. Chemistry of fossil fuels and biofuels. *Chemistry of Fossil Fuels and Biofuels* 9780521114004, (2010).
- [5] Dutkiewicz, S. *et al.* Dimensions of marine phytoplankton diversity. *Biogeosciences* (2020). doi:10.5194/bg-17-609-2020
- [6] Flombaum, P., Wang, W. L., Primeau, F. W., Martiny, A. C. Global picophytoplankton niche partitioning predicts overall positive response to ocean warming. *Nat. Geosci.* (2020). doi:10.1038/s41561-019-0524-2
- [7] Buitenhuis, E. T. *et al.* MAREDAT: Towards a world atlas of MARine Ecosystem DATA. *Earth Syst. Sci. Data* 5, (2013).
- [8] Buitenhuis, E. T. *et al.* Picophytoplankton biomass distribution in the global ocean. *Earth Syst. Sci. Data* 4, 37–46 (2012).
- [9] Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* (2019). doi:10.1016/j.cell.2019.10.008
- [10] Busseni, G. *et al.* Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean. *Glob. Ecol. Biogeogr.* (2020). doi:10.1111/geb.13161
- [11] Longhurst, A. R. Ecological Geography of the Sea. *Ecological Geography of the Sea* (2007). doi:10.1016/B978-0-12-455521-1.X5000-1
- [12] Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem. Cycles* (2013). doi:10.1002/gbc.20089
- [13] Richter, D. J. *et al.* (2022) Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems *eLife* 11:e78129
- [14] Frémont, P. *et al.* Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat. Clim. Chang.* (2022). doi:10.1038/s41558-022-01314-8
- [15] Benedetti, F. *et al.* Major restructuring of marine plankton assemblages under global warming. *Nat. Commun.* 12, (2021).

- [16] Guilhem, S.-K. *et al.* Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science* (80-). 374, 594–599 (2021).
- [17] Thomas, M. K., Kremer, C. T., Klausmeier, C. A., Litchman, E. A global pattern of thermal adaptation in marine phytoplankton. *Science* (80-). (2012). doi:10.1126/science.1224836
- [18] Hellweger, F. L. *et al.* The role of ocean currents in the temperature selection of plankton: Insights from an individual-based model. *PLoS One* 11, (2016).
- [19] Ward, B. A., Marañón, E., Sauterey, B., Rault, J., Claessen, D. The Size Dependence of Phytoplankton Growth Rates: A Trade-Off between Nutrient Uptake and Metabolism. *Am. Nat.* 189, 170–177 (2016).
- [20] Barton, A. D. *et al.* The biogeography of marine plankton traits. *Ecology Letters* (2013). doi:10.1111/ele.12063
- [21] Mock, T. *et al.* Evolutionary genomics of the cold-Adapted diatom *Fragilariopsis cylindrus*. *Nature* 541, (2017).
- [22] Delmont, T. O. *et al.* Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *Elife* (2019). doi:10.7554/eLife.46497
- [23] Laso-Jadart, R. *et al.* Investigating population-scale allelic differential expression in wild populations of *Oithona similis* (Cyclopoida, Claus, 1866). *Ecol. Evol.* (2020). doi:10.1002/ece3.6588
- [24] Irwin, A. J., Finkel, Z. V., Müller-Karger, F. E., Troccoli Ghinaglia, L. Phytoplankton adapt to changing ocean environments. *Proc. Natl. Acad. Sci.* (2015). doi:10.1073/pnas.1414752112
- [25] Bopp, L. *et al.* Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences* (2013). doi:10.5194/bg-10-6225-2013
- [26] Dutkiewicz, S., Hickman, A.E., Jahn, O., Gregg, W.W., Mouw, C.B. and Follows, M.J. Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeoscience*, 12, 4447-4481 doi:10.5194/bg-12-4447-2015 (2015)
- [27] Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., Gehlen, M. PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* (2015). doi:10.5194/gmd-8-2465-2015
- [28] Beckmann, A., Schaum, C. E., Hense, I. Phytoplankton adaptation in ecosystem models. *J. Theor. Biol.* 468, (2019).
- [29] Sauterey, B., Ward, B., Rault, J., Bowler, C., Claessen, D. The Implications of Eco-Evolutionary Processes for the Emergence of Marine Plankton Community Biogeography. *Am. Nat.* (2017). doi:10.1086/692067
- [30] Ward, B. A., Cael, B. B., Collins, S., Robert Young, C. Selective constraints on global plankton dispersal. *Proc. Natl. Acad. Sci. U. S. A.* 118, (2021).
- [31] Clark, J. R., Daines, S. J., Lenton, T. M., Watson, A. J., Williams, H. T. P. Individual-based modelling of adaptation in marine microbial populations using genetically defined physiological parameters. *Ecol. Modell.* 222, (2011).
- [32] Clark, J. R., Lenton, T. M., Williams, H. T. P., Daines, S. J. Environmental selection and resource allocation determine spatial patterns in picophytoplankton cell size. *Limnol. Oceanogr.* 58, (2013).
- [33] Hellweger, F. L., Van Sebille, E., Fredrick, N. D. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* (80-). (2014). doi:10.1126/science.1254421
- [34] Hellweger, F. L. *et al.* The role of ocean currents in the temperature selection of plankton: Insights from an individual-based model. *PLoS One* 11, (2016).
- [35] Zaiss, J., Boyd, P. W., Doney, S. C., Havenhand, J. N., Levine, N. M. Impact of Lagrangian Sea Surface Temperature Variability on Southern Ocean Phytoplankton Community Growth Rates. *Global Biogeochem. Cycles* 35, (2021).
- [36] Wu *et al.*, (2022). PlanktonIndividuals.jl: A GPU supported individual-based phytoplankton life cycle model. *Journal of Open Source Software*, 7(73), 4207, <https://doi.org/10.21105/joss.04207>
- [37] Case T.J. 2000 *An illustrated guide to theoretical ecology*. Oxford, UK: Oxford University Press.
- [38] Pontarp Mikael and Petchey Owen L. 2018. Ecological opportunity and predator–prey interactions: linking eco-evolutionary processes and diversification in adaptive radiations *Proc. R. Soc. B.* **285**:20172550 <http://doi.org/10.1098/rspb.2017.2550>

- [39] A size-structured food-web model for the global ocean, B. Ward *et al.* *Limnology and Oceanography*, 2012, 57, 1877-1891
- [40] Geoffrey T. Evans and John S. Parslow (1985) A Model of Annual Plankton Cycles, *Biological Oceanography*, 3:3, 327-347, DOI: 10.1080/01965581.1985.10749478
- [41] Stec, Krzysztof Franciszek (2016). Role of Life Strategies in Diatom Biodiversity: A Modelling Approach. PhD thesis The Open University. doi: <https://doi.org/10.21954/ou.ro.0000ef90>
- [42] Eppley, R. W. 1972. Temperature and phytoplankton growth in the sea. *Fish. Bull.* 70: 1063–1085.
- [43] Monod, Jacques (1949). "The growth of bacterial cultures". *Annual Review of Microbiology*. 3: 371–394. doi:10.1146/annurev.mi.03.100149.002103
- [44] Zakem, E.J., Al-Haj, A., Church, M.J. *et al.* Ecological control of nitrite in the upper ocean. *Nat Commun* 9, 1206 (2018). <https://doi.org/10.1038/s41467-018-03553-w>
- [45] Tilman, D. 1982. Resource competition and community structure. *Princeton University Press*.
- [46] Shadden, Shawn C. and Lekien, Francois and Marsden, Jerrold E. (2005) Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Physica D*, 212 (3-4). pp. 271-304. ISSN 0167-2789. <https://resolver.caltech.edu/CaltechAUTHORS:20100914-104231365>
- [47] M.J.R. Fasham, Hugh Ducklow, Hugh Ducklow, S.M. McKelvie. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. August 1990 *Journal of Marine Research* 48(3):591-639. DOI: 10.1357/002224090784984678.
- [48] Richard J. Geider, Richard J. Geider, Hugh L. Macintyre, Hugh L. Macintyre, Todd Kana, Todd Kana. Dynamic model of phytoplankton growth and acclimation: Responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature March 1997 *Marine Ecology Progress Series* 148(1-3):187-200 DOI: 10.3354/meps148187
- [49] Stawiarski B, Buitenhuis ET, Fallens M (2018) The physiological response of seven strains of picophytoplankton to light, and its representation in a dynamic photosynthesis model. *Limnol Oceanogr* 63:S367–S380
- [50] Froyland, G. and Padberg, K. (2009) Almost-Invariant Sets and Invariant Manifolds-Connecting Probabilistic and Geometric Descriptions of Coherent Structures in Flows. *Physica D*, 238, 1507-1523.
- [51] Gary Froyland and Kathrin Padberg-Gehle. A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data *Chaos* 25, 087406 (2015); <https://doi.org/10.1063/1.4926372>
- [52] R Banisch, P Koltai. Understanding the geometry of transport: Diffusion maps for Lagrangian trajectory data unravel coherent sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27 (3), 035804, 2017
- [53] Reijnders, D and van, Leeuwen EJ and van, Sebille E *Journal of Geophysical Research: Oceans* Ocean Surface Connectivity in the Arctic: Capabilities and Caveats of Community Detection in Lagrangian Flow Networks. <http://dx.doi.org/10.1029/2020jc016416>, 1262021
- [54] Alice Marzocchi, Joël J.-M. Hirschi, N. Penny Holliday, Stuart A. Cunningham, Adam T. Blaker, Andrew C. Coward, The North Atlantic subpolar circulation in an eddy-resolving global ocean model, *Journal of Marine Systems*, Volume 142, 2015, 126-143, ISSN 0924-7963, <https://doi.org/10.1016/j.jmarsys.2014.10.007>.
- [55] Delandmeter, P. and van Sebille, E.: The Parcels v2.0 Lagrangian framework: new field interpolation schemes, *Geosci. Model Dev.*, 12, 3571–3584, <https://doi.org/10.5194/gmd-12-3571-2019>, 2019.
- [56] <https://www.itacanet.org/the-sun-as-a-source-of-energy/part-2-solar-energy-reaching-the-earths-surface/>
- [57] Kirk, J.T.O., Light and Photosynthesis in Aquatic Ecosystems, 401 pp, *Cambridge University Press*, Cambridge, 1983
- [58] Spencer, J. W. (1971), Fourier series representation of the position of the sun, *Search*, 2(5), 172
- [59] de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology, *J. Geophys. Res.*, 109, C12003. doi:10.1029/2004JC002378, pdf
- [60] Zeileis A, Grothendieck G (2005). "zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software*, 14(6), 1–27. doi: 10.18637/jss.v014.i06.

- [61] Douglas Nychka, Reinhard Furrer, John Paige, Stephan Sain (2017). “fields: Tools for spatial data.” doi: 10.5065/D6W957CT, R package version 11.6, <https://github.com/NCAR/Fields>.
- [62] Borchers, H. W. 2017. R package “pracma.”
- [63] Locarnini, R. A., A. V. Mishonov, O. K. Baranova, T. P. Boyer, M. M. Zweng, H. E. Garcia, J. R. Reagan, D. Seidov, K. Weathers, C. R. Paver, and I. Smolyar, 2018. *World Ocean Atlas 2018, Volume 1: Temperature*. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 81, 52 pp.
- [64] Garcia, H. E., K. Weathers, C. R. Paver, I. Smolyar, T. P. Boyer, R. A. Locarnini, M. M. Zweng, A. V. Mishonov, O. K. Baranova, D. Seidov, and J. R. Reagan, 2018. *World Ocean Atlas 2018, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate)*. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 84, 35 pp.
- [65] Somerfield, Paul. (2008). Identification of the Bray-Curtis similarity index: Comment on Yoshioka (2008). Marine Ecology-progress Series - MAR ECOL-PROGR SER. 372. 303-306. 10.3354/meps07841.
- [66] Hutchinson, G. E. (1961) The paradox of the plankton. American Naturalist 95, 137-145. Shovonlal Roy, J. Chattopadhyay,
- [67] Shovonlal Roy, J. Chattopadhyay, Towards a resolution of ‘the paradox of the plankton’: A brief overview of the proposed mechanisms, Ecological Complexity, Volume 4, Issues 1–2, 2007, Pages 26-33, ISSN 1476-945X, <https://doi.org/10.1016/j.ecocom.2007.02.016>.
- [68] Kléparski Loïck, Beaugrand Grégory and Kirby Richard R. 2022How do plankton species coexist in an apparently unstructured environment?Biol. Lett.182022020720220207. <http://doi.org/10.1098/rsbl.2022.0207>
- [69] Luca Ferretti, Sebastián E. Ramos-Onsins, A generalized Watterson estimator for next-generation sequencing: From trios to autopolyploids, Theoretical Population Biology, Volume 100, 2015, Pages 79-87, ISSN 0040-5809, <https://doi.org/10.1016/j.tpb.2015.01.001>.
- [70] Watterson, G.A. (1975), On the number of segregating sites in genetical models without recombination., Theoretical Population Biology, 7 (2): 256–276, doi:10.1016/0040-5809(75)90020-9
- [71] Bates, M., Tulloch, R., Marshall, J., Ferrari, R. (2014). Rationalizing the Spatial Distribution of Mesoscale Eddy Diffusivity in Terms of Mixing Length Theory, Journal of Physical Oceanography, 44(6), 1523-1540.
- [72] Abraham, E. The generation of plankton patchiness by turbulent stirring. Nature 391, 577–580 (1998). <https://doi.org/10.1038/35361>
- [73] Clayton, S. *et al.* Biogeochemical versus ecological consequences of modeled ocean physics. Biogeosciences 14, (2017).
- [74] Weitz, J. S., Wilhelm, S. W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol. Rep. 4, (2012).
- [75] Demory, D. *et al.* A thermal trade-off between viral production and degradation drives virus-phytoplankton population dynamics. Ecol. Lett. 24, (2021).
- [76] Bopp, L. *et al.* Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. Biogeosciences (2013). doi:10.5194/bg-10-6225-2013
- [77] Dufresne, J. L. *et al.* Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. Clim. Dyn. (2013). doi:10.1007/s00382- 012-1636-1

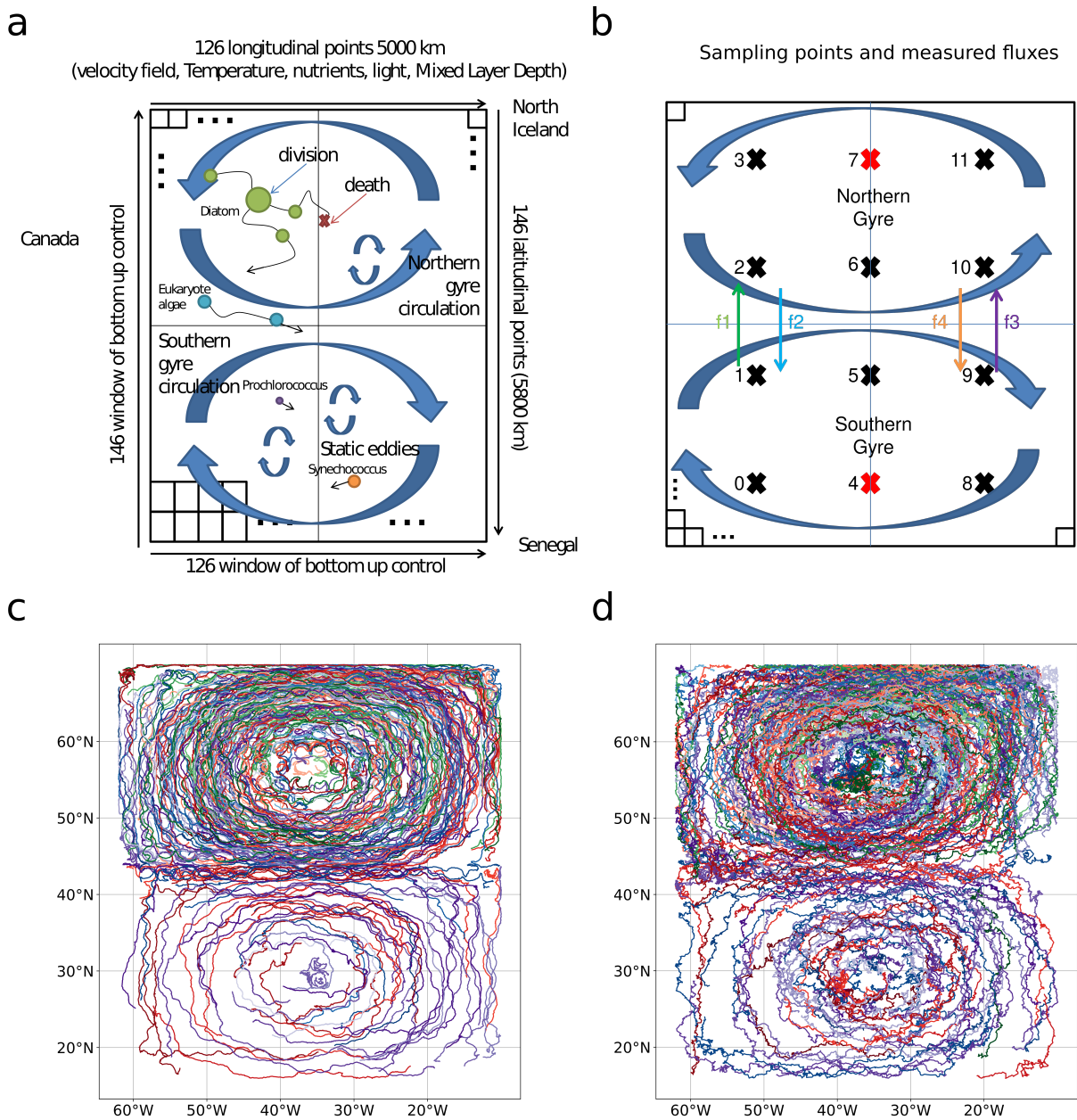


Figure 1 | Scheme of the double gyre flow and individual based model (IBM) representation of a size structured phytoplankton community. (a) A set of particles randomly assigned to a taxonomy, size (which determines the maximum growth rate) and thermal niche are advected within a constant double gyre flow that include static mesoscale eddies. Particles grow, divide and die according to mean temperature and nutrient level and integrated light limitation over the mixed layer depth at each grid point. Particles divide when they reach twice a prefixed biomass m_0 and die when their biomass is reaches a biomass equivalent to 100 cells per m^2 of *Prochlorococcus*. Four types of phytoplankton are represented in the model in increasing sizes: *Prochlorococcus* (purple cells), *Synechococcus* (orange cells), Other eukaryotes (blue cells) and diatoms (green cells). (b) Sampling points and measured particle and biomass flows in the double gyre scheme. (c) Example of particle trajectories within the double gyre flow with a diffusivity constant of $100 m^2 \cdot s^{-1}$ and (d) $1000 m^2 \cdot s^{-1}$.

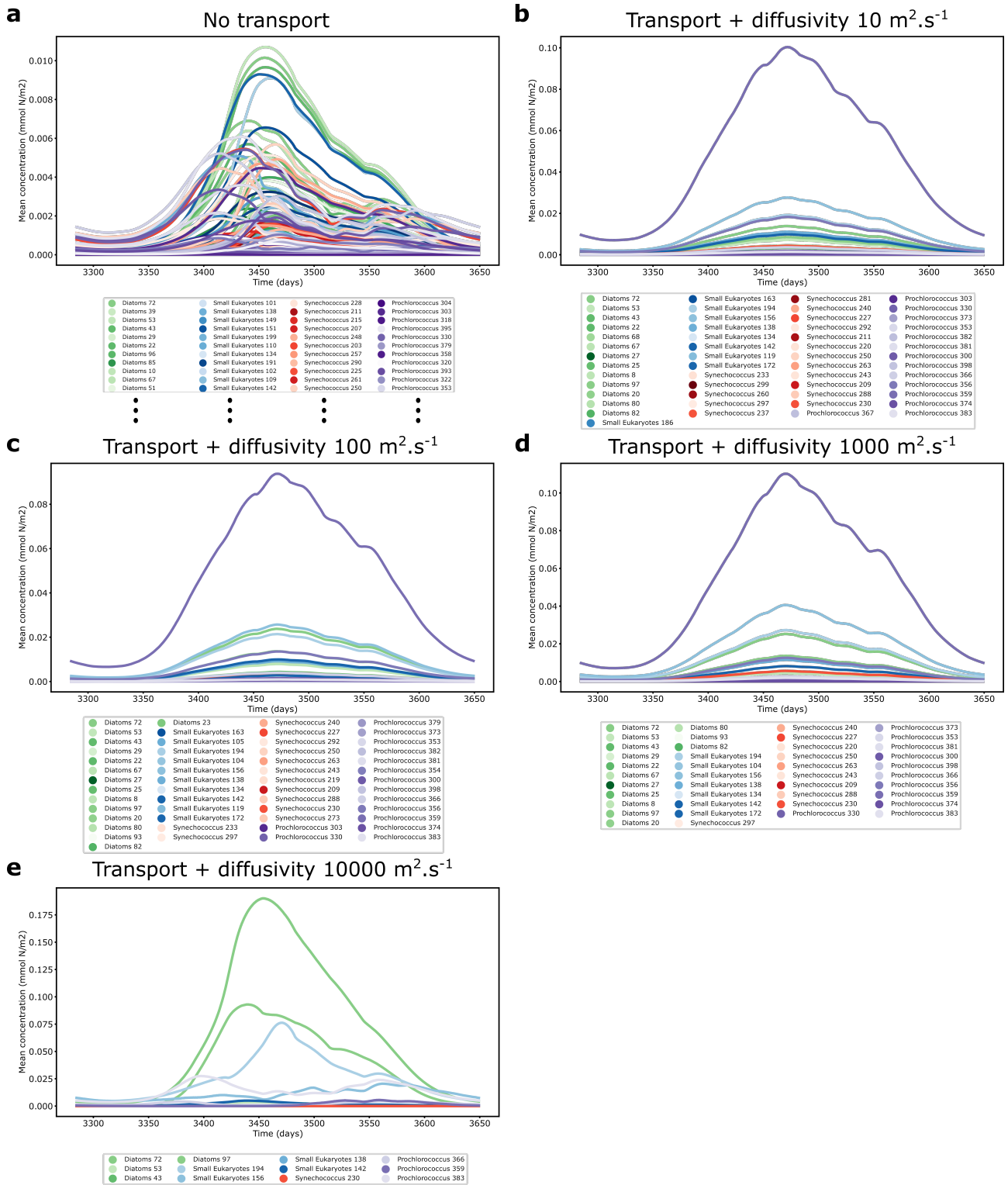


Figure 2 | Average biomass concentration of each phytoplankton type in the northern gyre over the 10^4 th year of simulation (with a grazing term) using different transport and diffusivity parametrization. For each type of simulation, only surviving species concentration are represented. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2 \cdot \text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2 \cdot \text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2 \cdot \text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2 \cdot \text{s}^{-1}$.

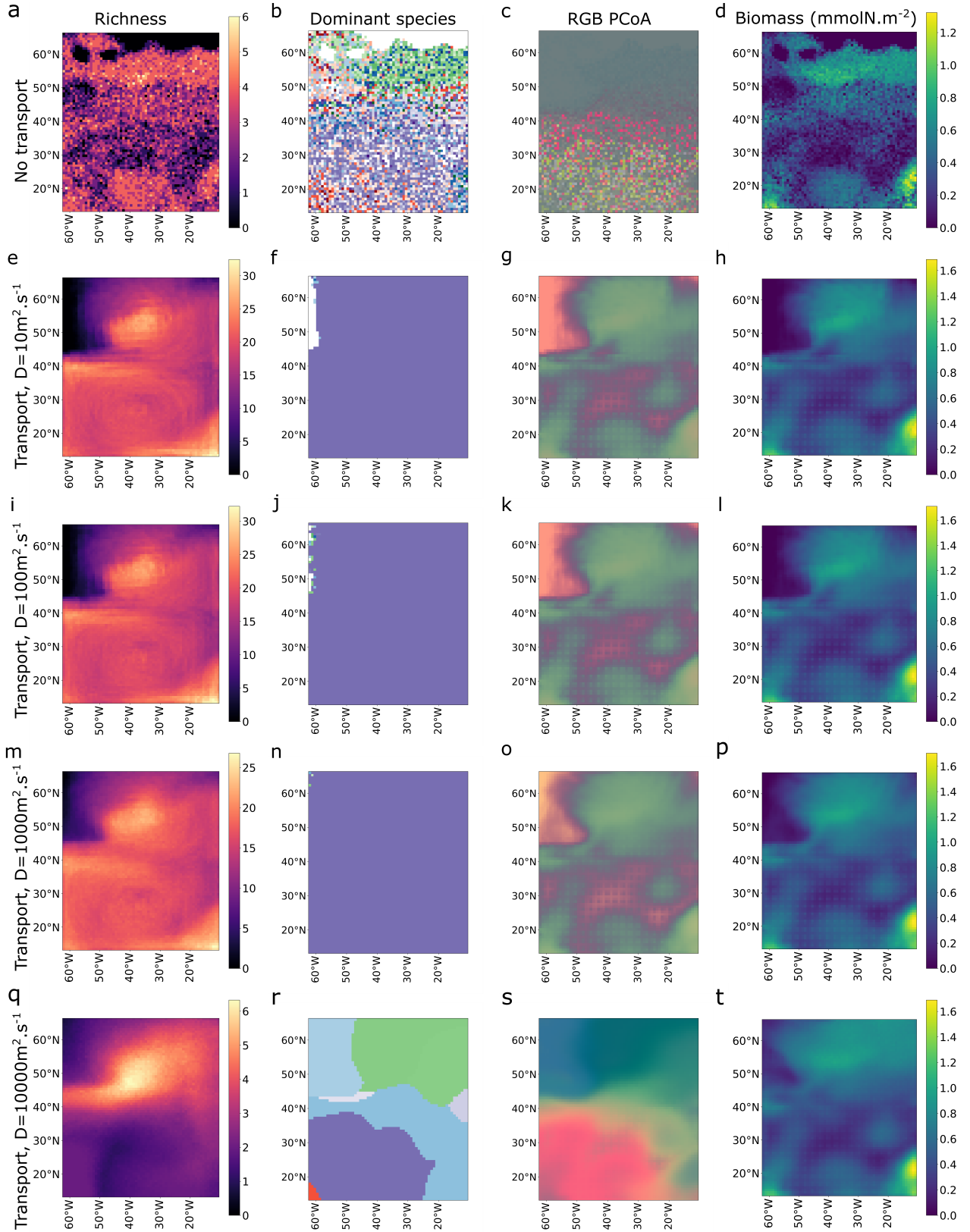


Figure 3 | Mean basic ecological properties of the model (with a grazing term) term over the 10th year of simulation for different transport and diffusivity parametrization. (a-d) No transport, (e-h) Transport and diffusivity kernel $D = 10m^2.s^{-1}$, (i-l) Transport and diffusivity kernel $D = 100m^2.s^{-1}$, (m-p) Transport and diffusivity kernel $D = 1000m^2.s^{-1}$, (q-t) Transport and diffusivity kernel $D = 10000m^2.s^{-1}$, (a, e, i, m, q) Richness defined as the mean number of coexisting species at each grid point, (b, f, j, n, r) Mean dominant species. Diatoms are in greens, other Eukaryotes in blues, *Prochlorococcus* in purples and *Synechococcus* in reds/oranges. The lighter the color is the larger the cell size of the species. (c, g, k, o, s) Biogeography: Principal Coordinate Analysis (PCoA) of the mean Bray-Curtis distance matrix of each grid point's species biomass over the year. (d, h, l, p, t) Mean integrated biomass over the year.

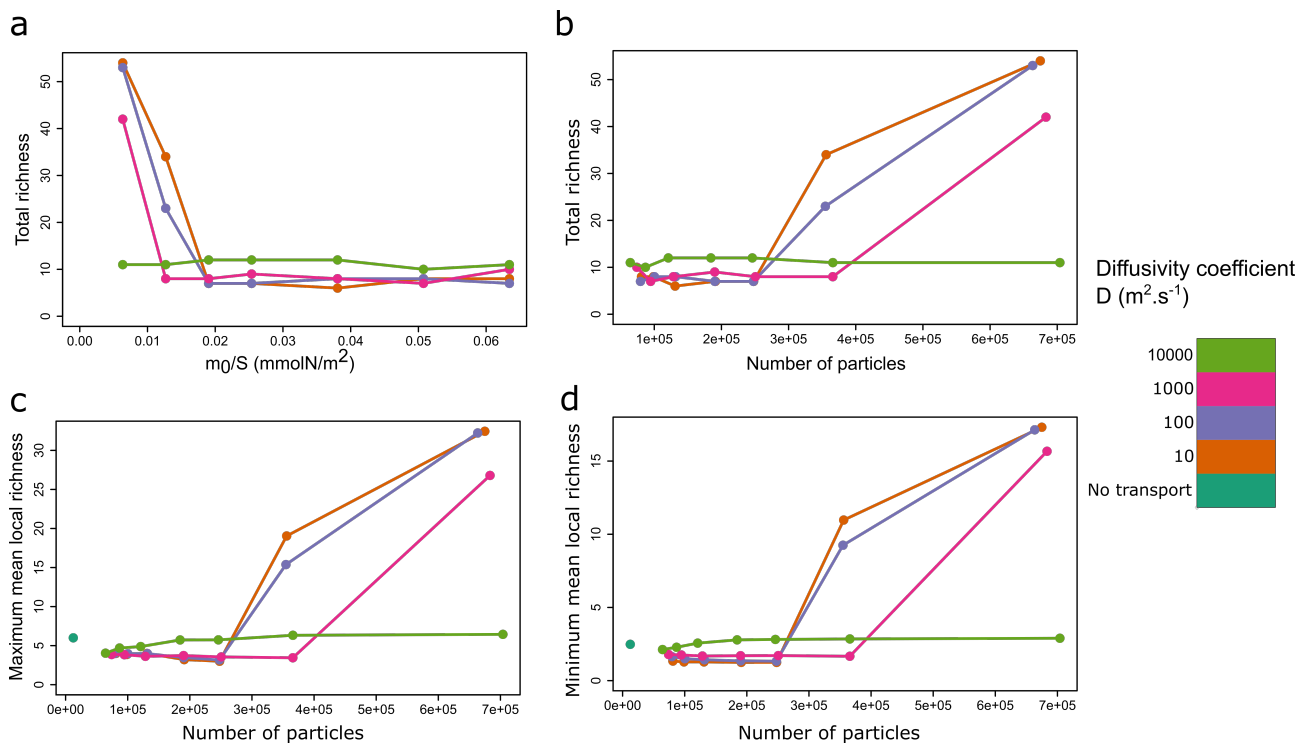


Figure 4 | Modeled diversity in function of the number of particles or the reference biomass concentration of a particle for the 4 diffusivity kernels with a grazing term. (a) Total diversity modeled in function of the reference biomass concentration of a particle. (b) Total diversity modeled in function of the mean number of particles in the last year of simulation. (c) Maximum local richness modeled in function of the mean number of particles in the last year of simulation. (d) Minimum local richness modeled in function of the mean number of particles in the last year of simulation.

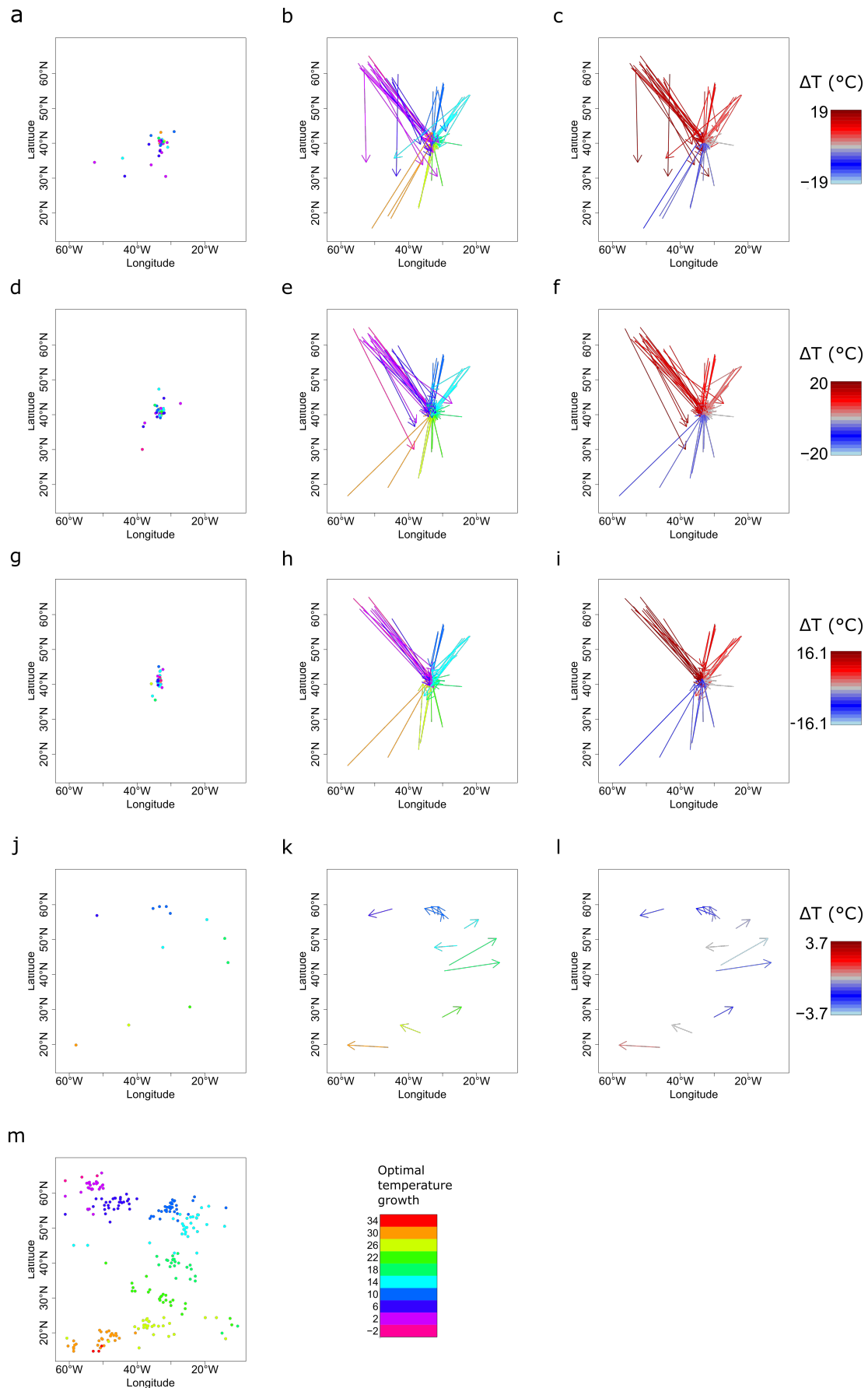
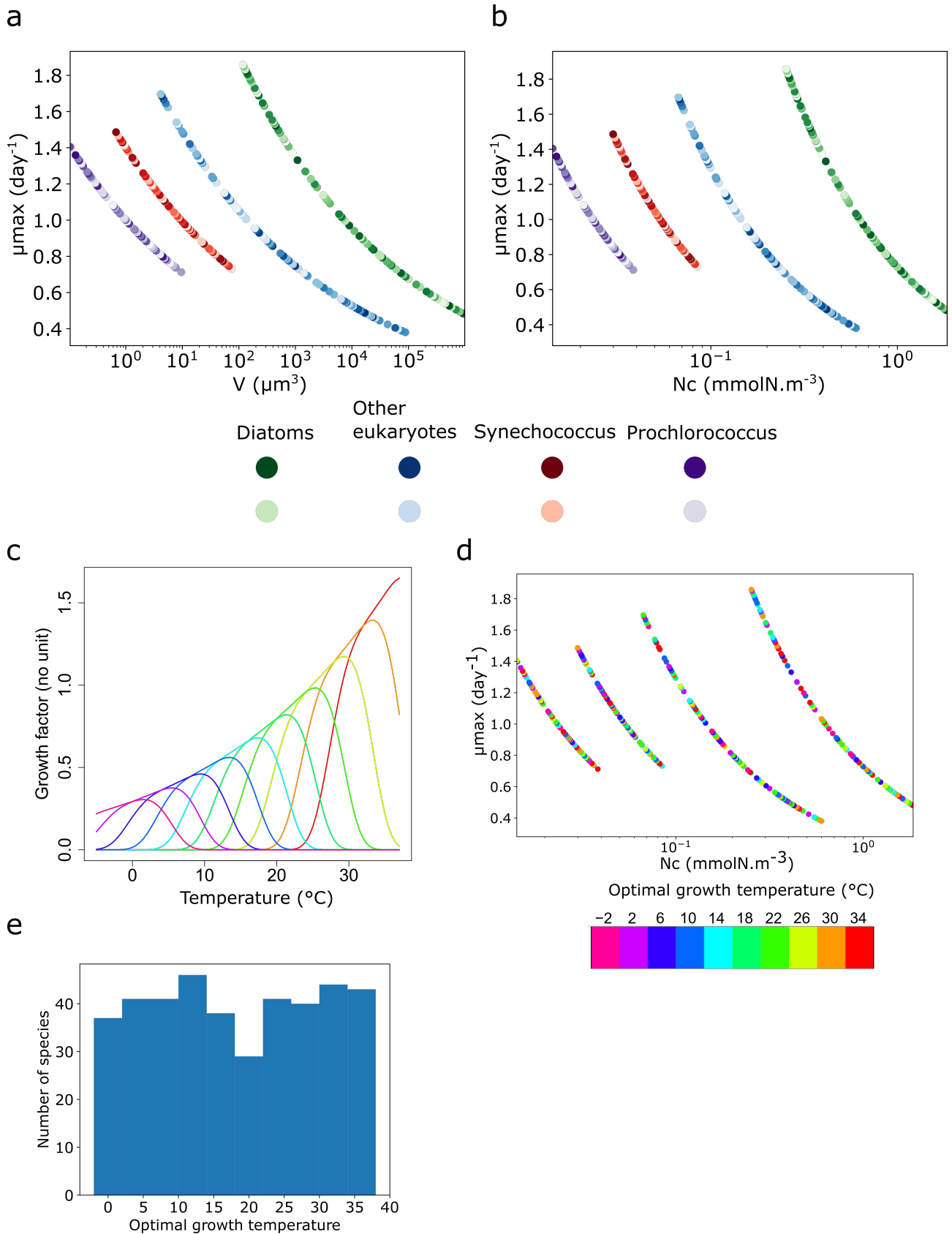
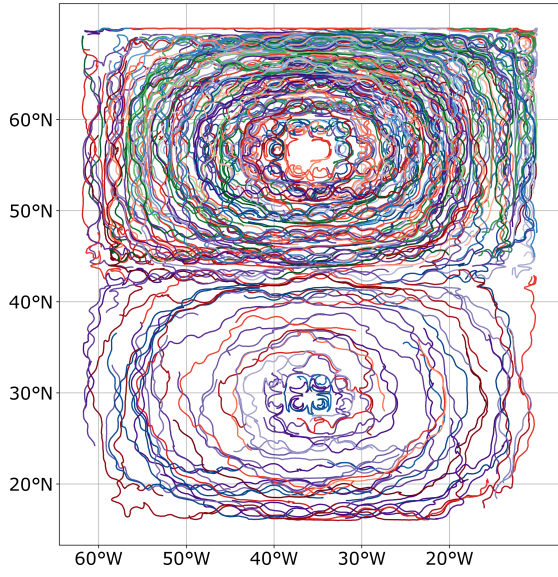
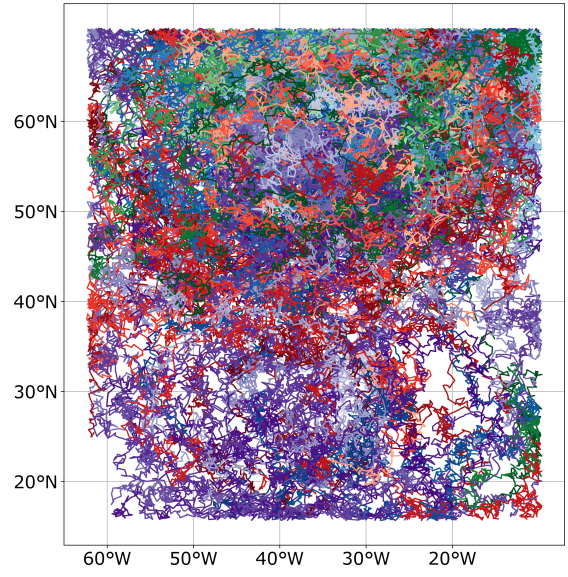


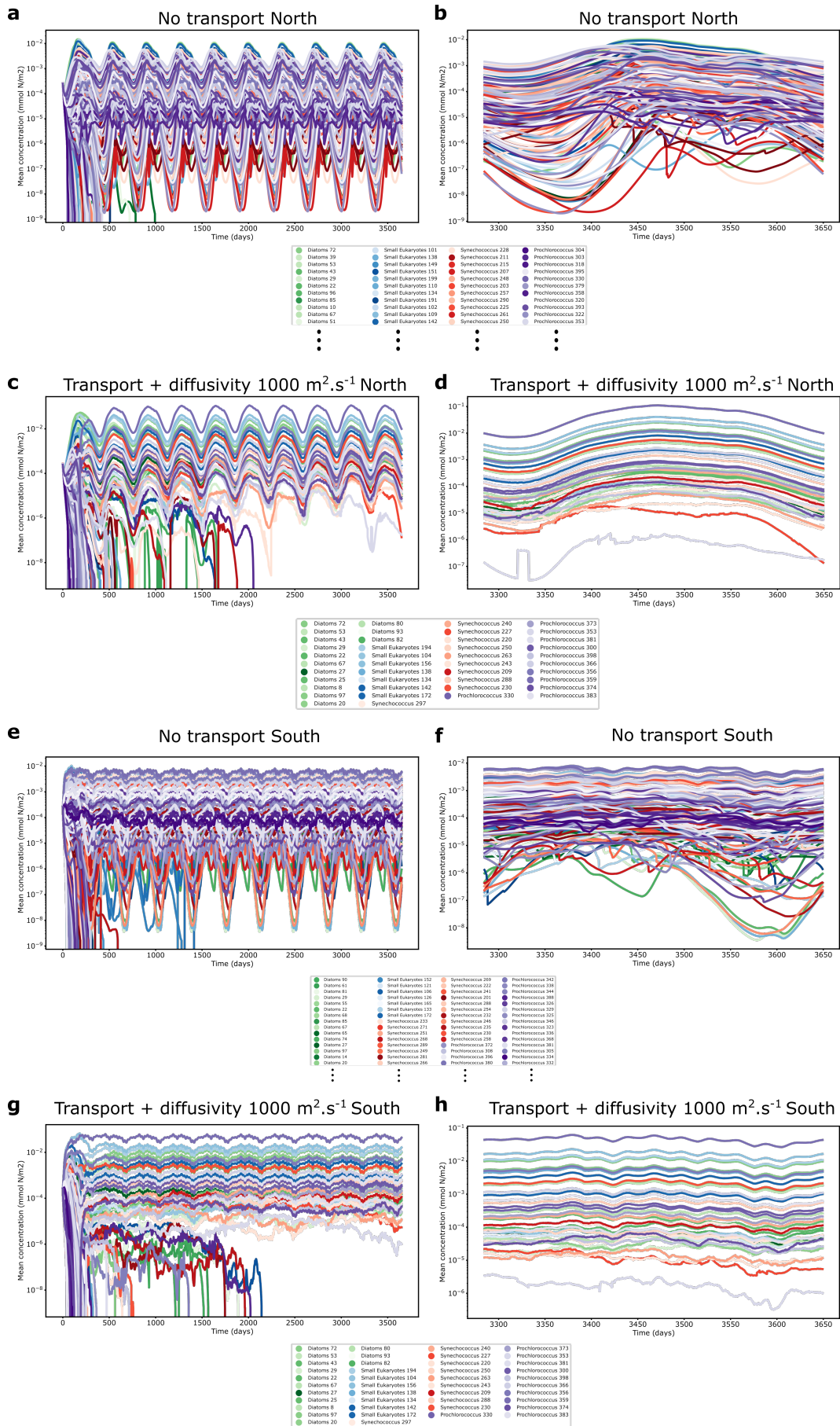
Figure 5 | Influence of currents and temperature shifts in niches centroids of modeled species for different transport and diffusivity parametrization. (a-c) Transport and diffusivity kernel $D = 10m^2.s^{-1}$, (d-f) Transport and diffusivity kernel $D = 100m^2.s^{-1}$, (g-i) Transport and diffusivity kernel $D = 1000m^2.s^{-1}$, (j-l) Transport and diffusivity kernel $D = 10000m^2.s^{-1}$, (m) No transport. (a, d, g, j, m) Niches centroids of modeled species colored according to their optimal growth temperature. (b, e, h, k) Displacement of niches centroids in simulations with transport taking as reference the no transport simulation (m). (c, f, i, l) Changes in mean temperature between the niches centroids with transport and the mean temperature of centroids for the no transport simulation (m).



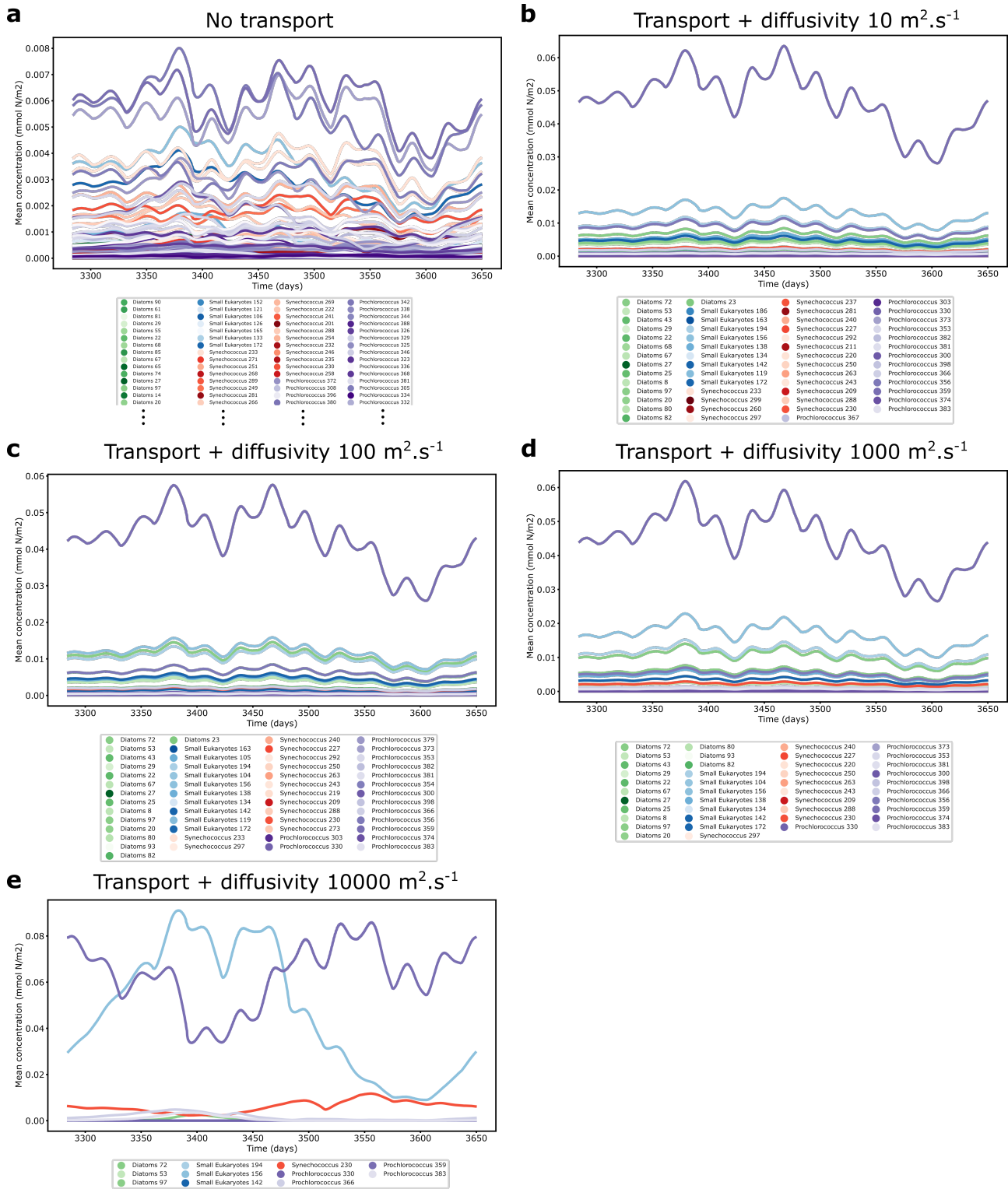
Supplementary Figure 1 | Initial ecological characteristics of the phytoplankton community model For a given simulation, 100 species of each phytoplankton type are randomly initiated with a size and a thermal niche. Here is presented the initial state of the simulation with the same seed that is used throughout the study. (a) Trade-off between maximal growth rate and size (cellular volume) across the different modeled phytoplankton groups. (b) Trade-off between maximal growth rate and nutrient half saturation constant for growth across the different modeled phytoplankton groups. (b) Growth response curves of the ten modeled thermal niches. (d) Same as (b) with each species colored according to its optimal growth temperature. (e) Initial distribution of optimal growth temperature.

a**b**

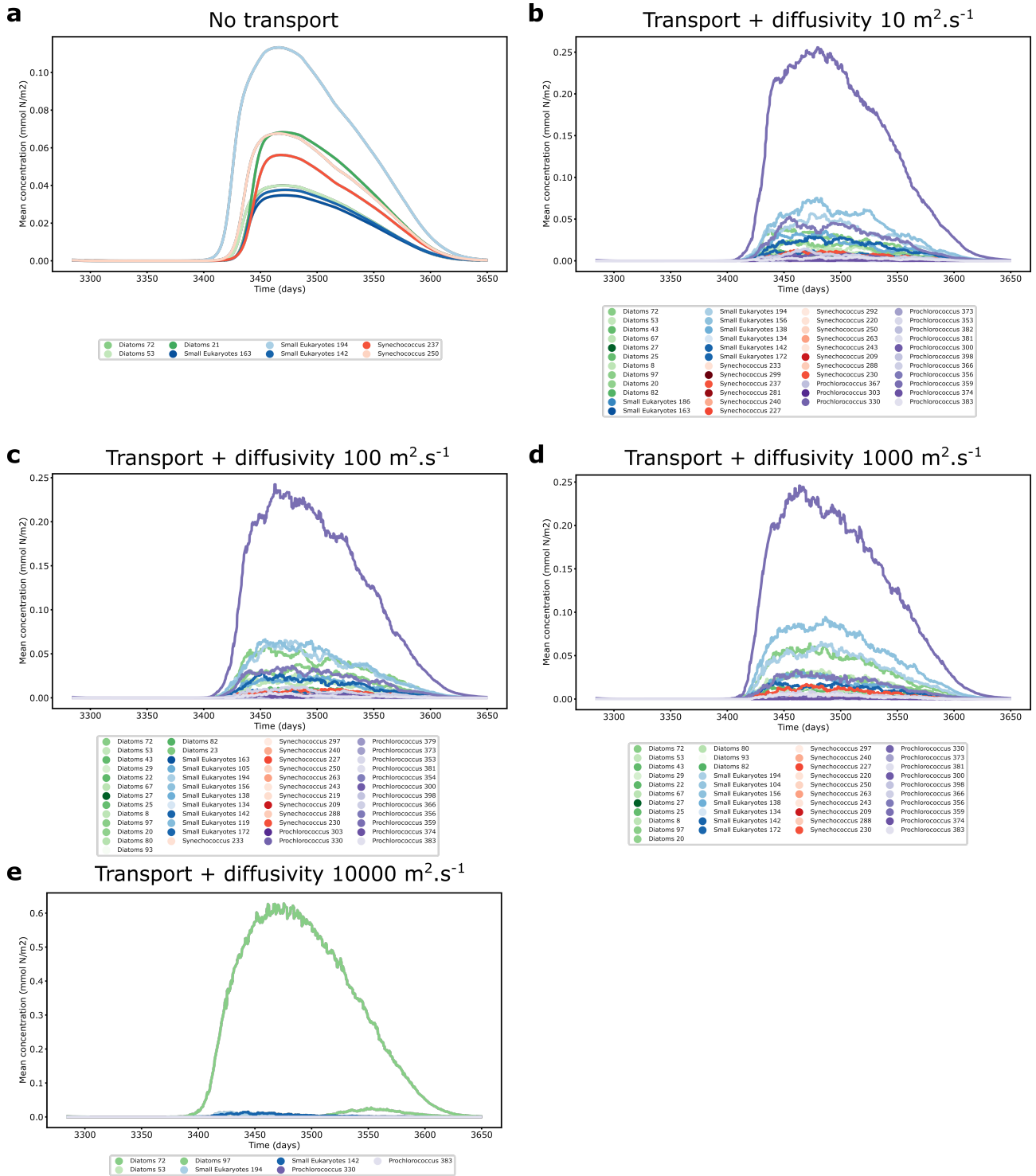
Supplementary Figure 2 | Example of particle trajectories within the double gyre flow with two different diffusivity constant. (a) $10 \text{ m}^2.\text{s}^{-1}$ and (b) $10000 \text{ m}^2.\text{s}^{-1}$



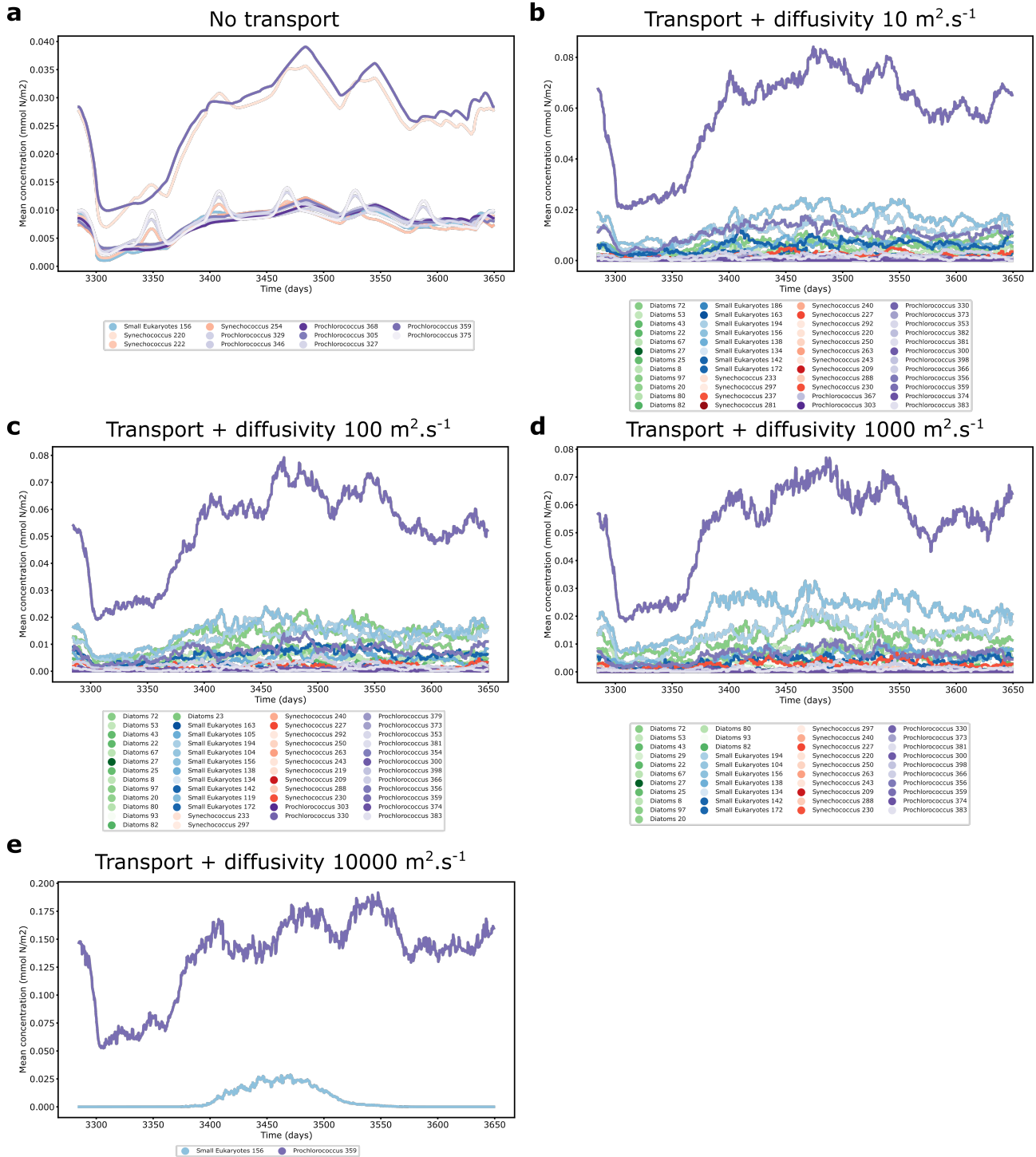
Supplementary Figure 3 | Average biomass concentration in log scale of each phytoplankton type in both gyres over the whole simulation and for the 10th year using different transport and diffusivity parameterization and a grazing term. For each type of simulation, only surviving species concentration are represented. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2 \cdot \text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2 \cdot \text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2 \cdot \text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2 \cdot \text{s}^{-1}$.



Supplementary Figure 4 | Average biomass concentration of each phytoplankton type in the southern gyre over the 10th year of simulation using different transport and diffusivity parametrization and a grazing term. For each type of simulation, only surviving species concentration are represented. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2 \cdot \text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2 \cdot \text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2 \cdot \text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2 \cdot \text{s}^{-1}$.

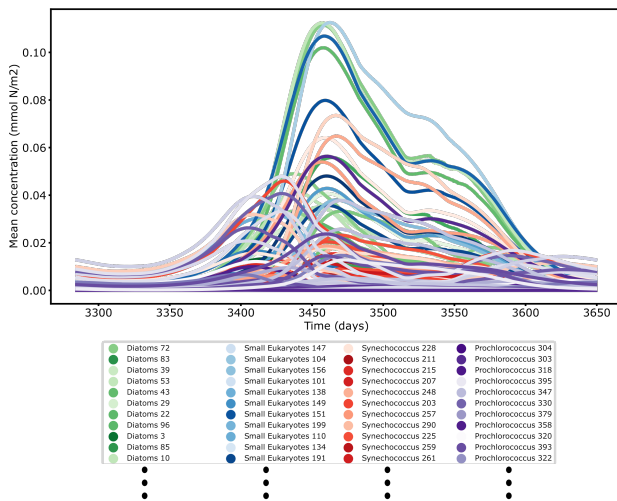


Supplementary Figure 5 | Average biomass concentration of each phytoplankton type in the point 7 of the northern gyre over the 10th year of simulation using different transport and diffusivity parametrization and a grazing term. For each type of simulation, only surviving species concentration are represented. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2 \cdot \text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2 \cdot \text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2 \cdot \text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2 \cdot \text{s}^{-1}$.

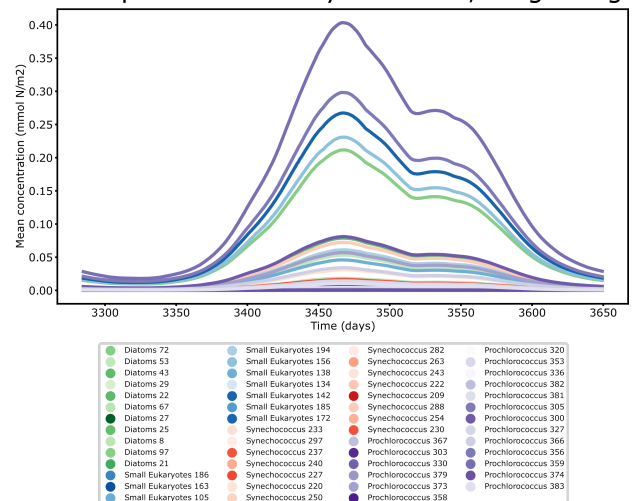


Supplementary Figure 6 | Average biomass concentration of each phytoplankton type in the point 4 of the southern gyre over the 10th year of simulation using different transport and diffusivity parametrization. For each type of simulation, only surviving species concentration are represented. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2 \cdot \text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2 \cdot \text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2 \cdot \text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2 \cdot \text{s}^{-1}$.

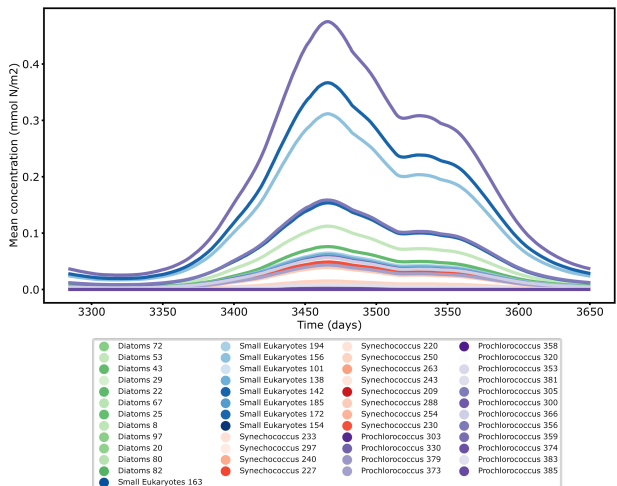
a No transport, no grazing



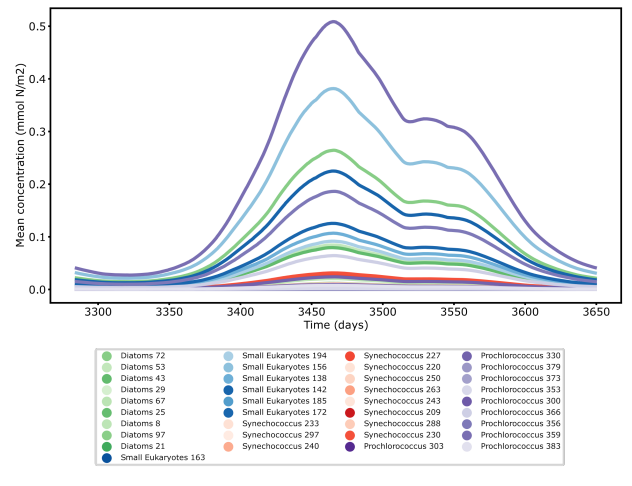
b Transport + diffusivity $10 \text{ m}^2 \cdot \text{s}^{-1}$, no grazing



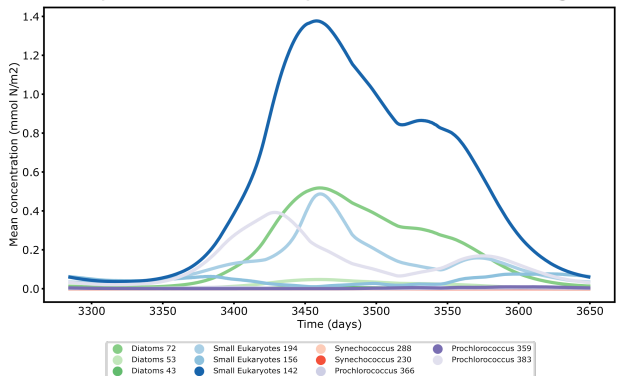
c Transport + diffusivity $100 \text{ m}^2 \cdot \text{s}^{-1}$, no grazing



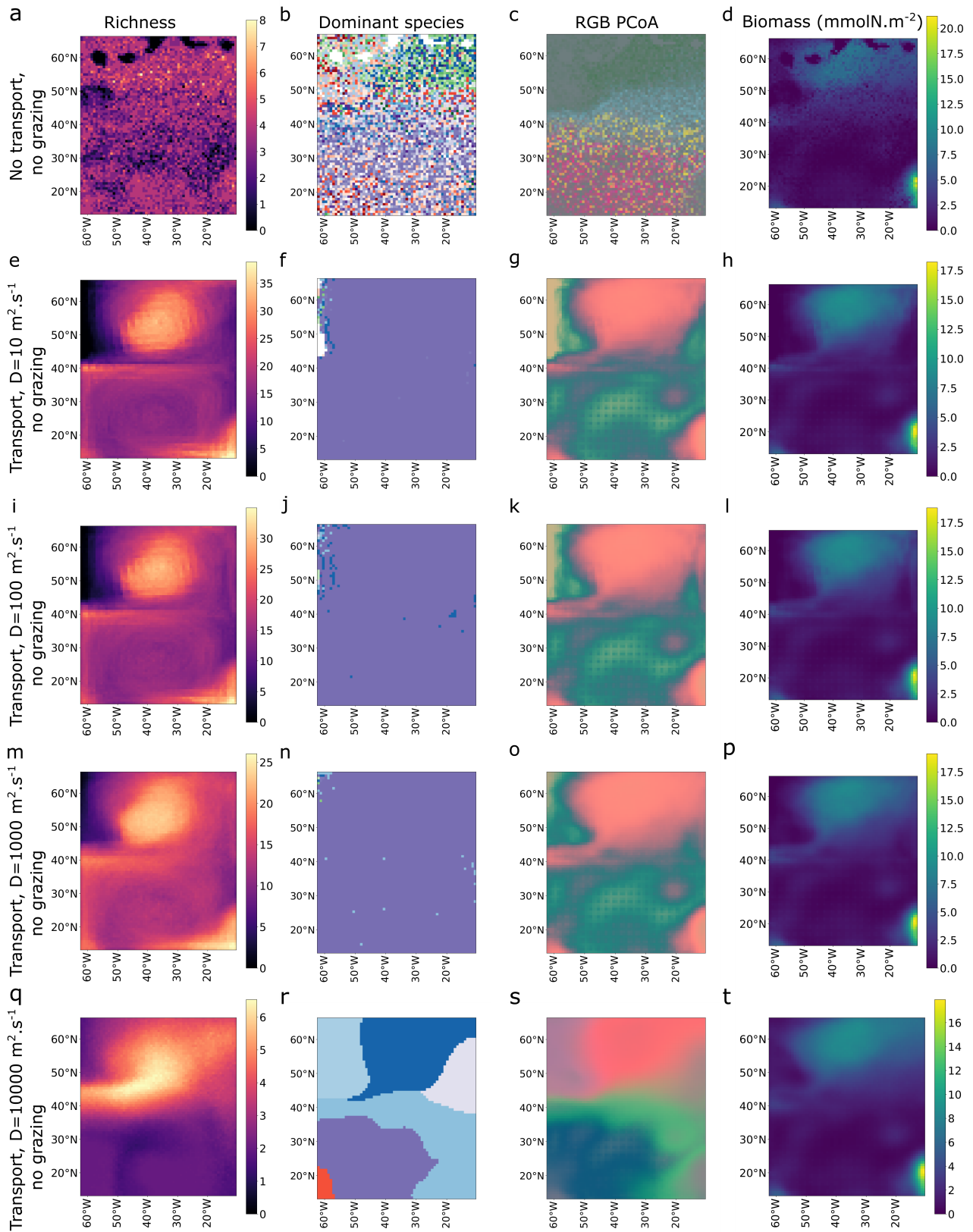
d Transport + diffusivity $1000 \text{ m}^2 \cdot \text{s}^{-1}$, no grazing



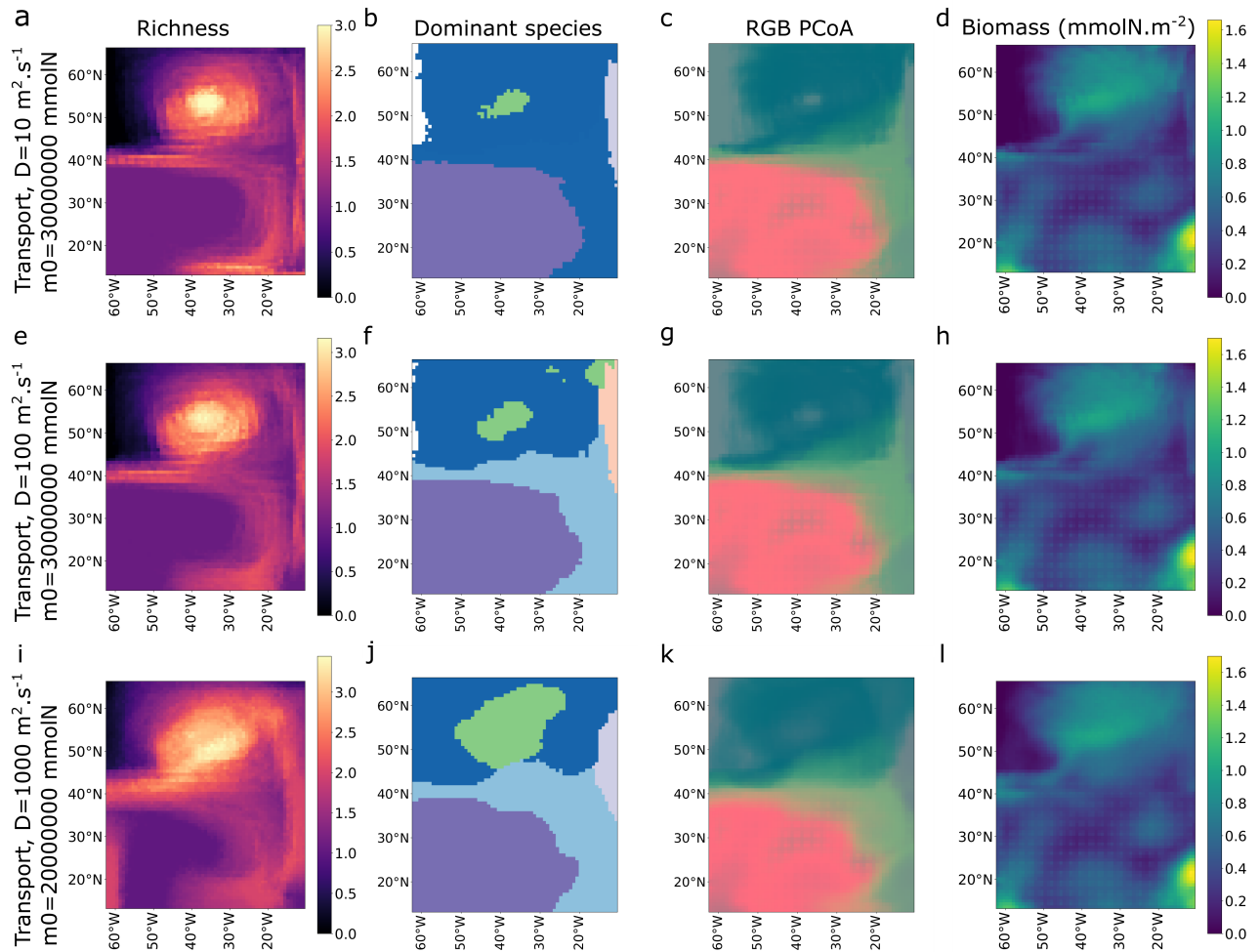
e Transport + diffusivity $10000 \text{ m}^2 \cdot \text{s}^{-1}$, no grazing



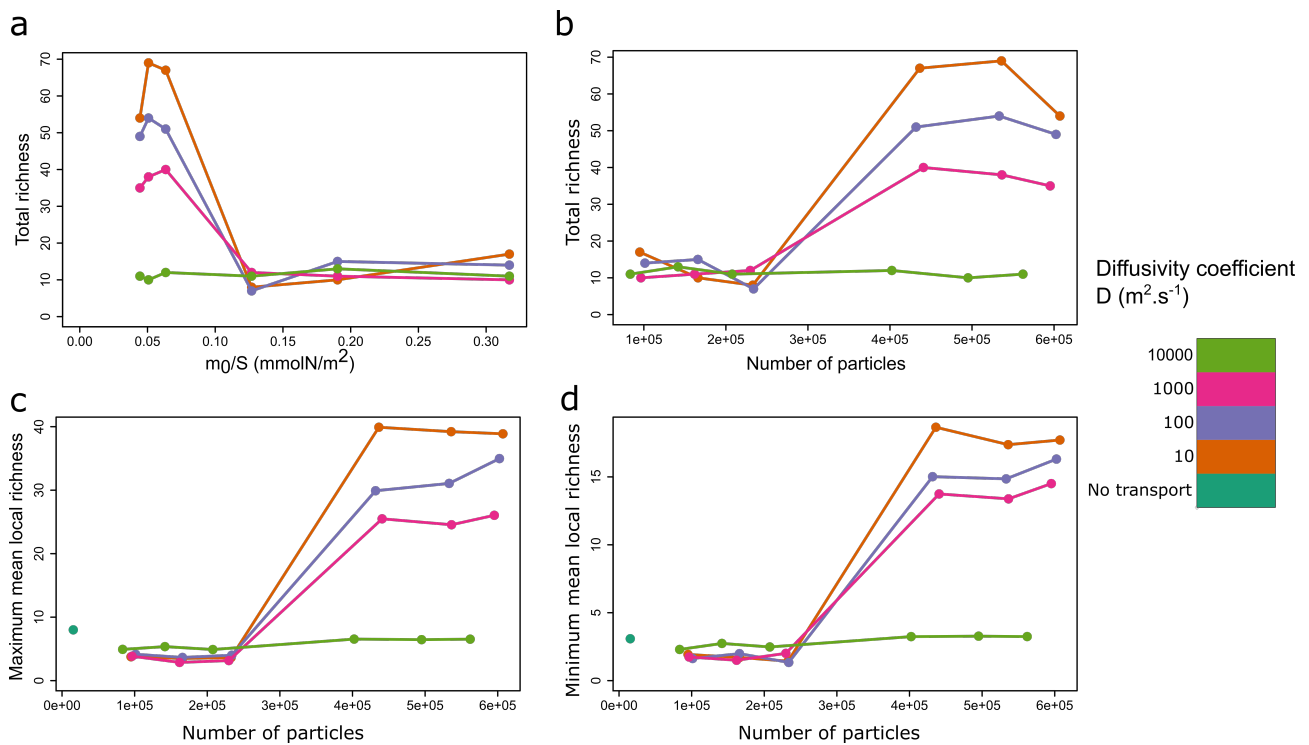
Supplementary Figure 7 | Average biomass concentration of each phytoplankton type in the northern gyre over the 10th year of simulation using different transport and diffusivity parametrization with no grazing. For each type of simulation, only surviving species concentration are represented. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2 \cdot \text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2 \cdot \text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2 \cdot \text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2 \cdot \text{s}^{-1}$.



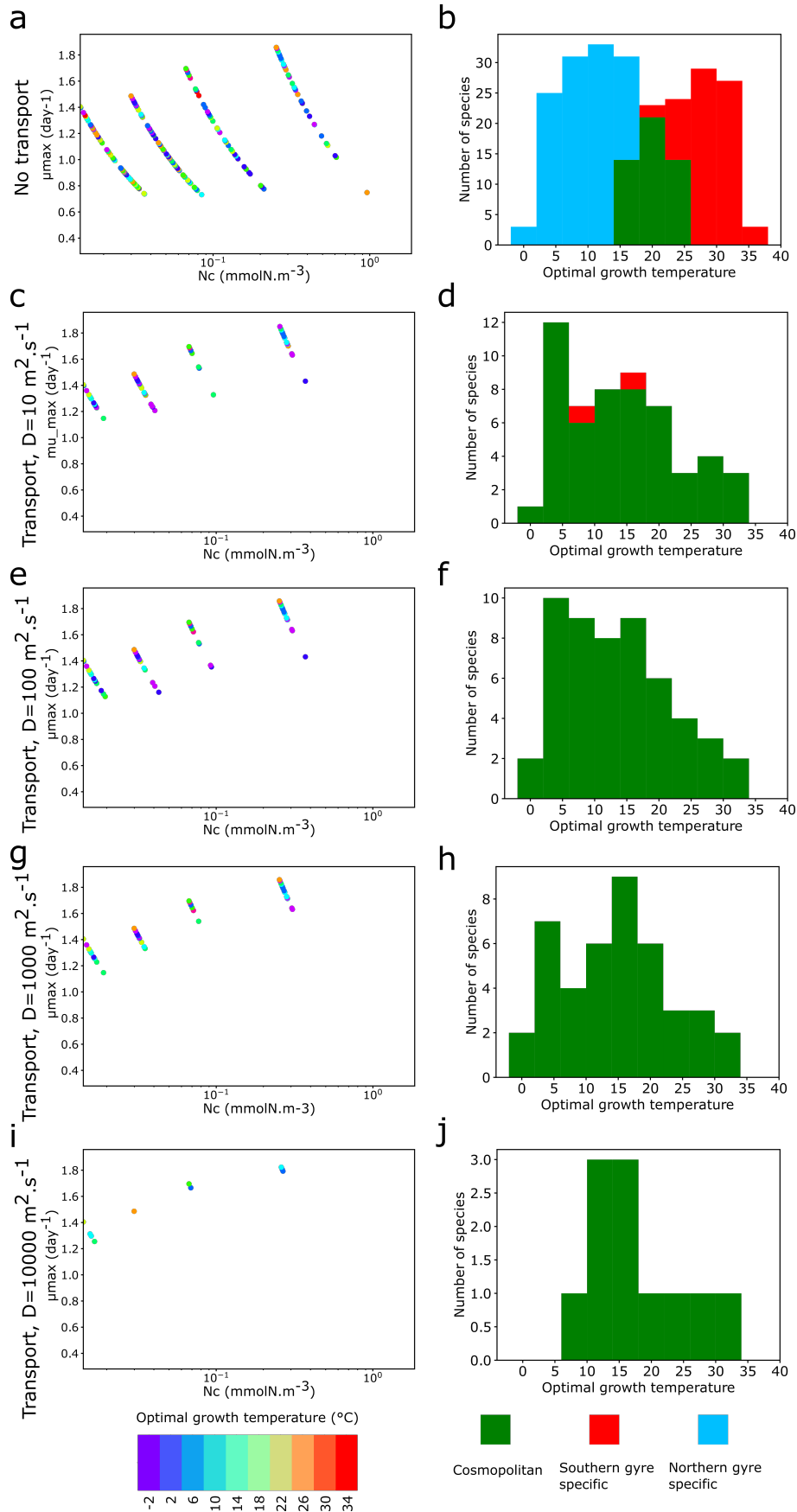
Supplementary Figure 8 | Mean basic ecological properties of the model over the 10th year of simulation for different transport and diffusivity parametrization with no grazing (a-d) No transport and diffusivity kernel $D = 10m^2.s^{-1}$, (e-h) Transport and diffusivity kernel $D = 100m^2.s^{-1}$, (i-l) Transport and diffusivity kernel $D = 1000m^2.s^{-1}$, (m-p) Transport and diffusivity kernel $D = 10000m^2.s^{-1}$, (q-t) Transport and diffusivity kernel $D = 10000m^2.s^{-1}$, (a, e, i, m, q) Richness defined as the mean number of coexisting species at each grid point, (b, f, j, n, r) Mean dominant species. Diatoms are in greens, other Eukaryotes in blues, *Prochlorococcus* in purples and *Synechococcus* in reds/oranges. The lighter the color is the larger the cell size of the species. (c, g, k, o, s) Biogeography: Principal Coordinate Analysis (PCoA) of the mean Bray-Curtis distance matrix of each grid point's species biomass over the year. (d, h, l, p, t) Mean integrated biomass over the year.



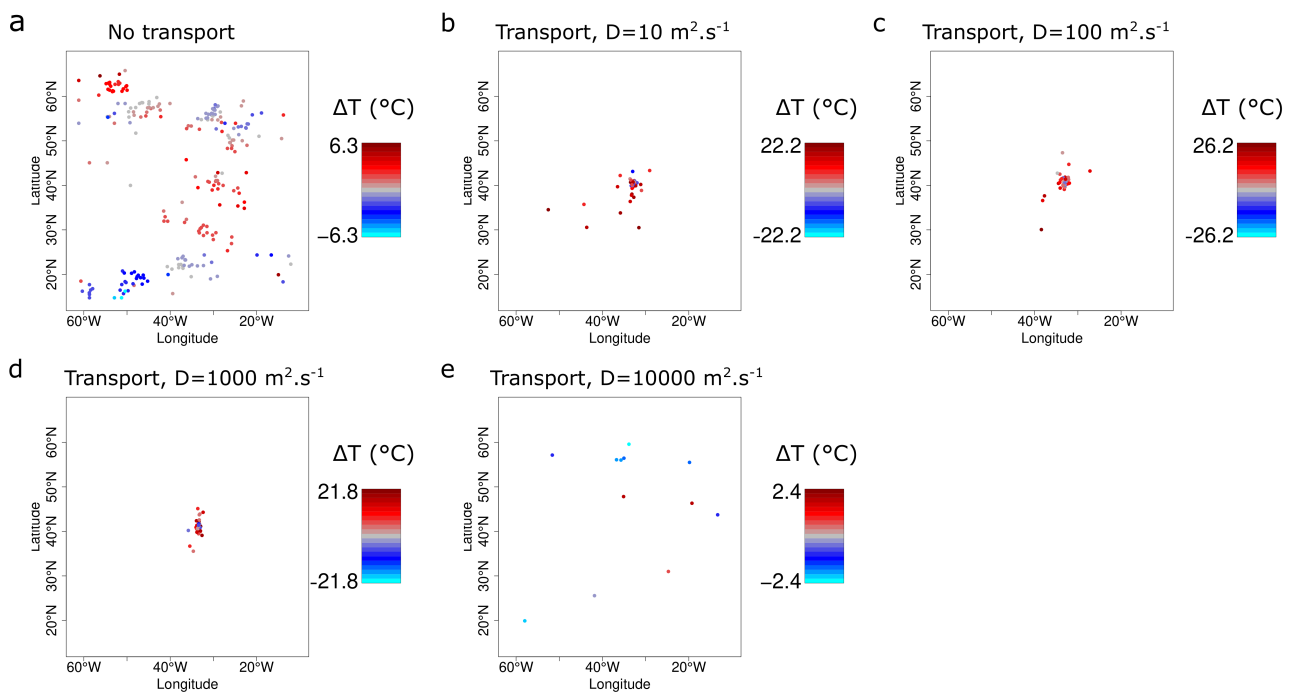
Supplementary Figure 9 | Mean basic ecological properties of the model over the 10th year of simulation for different transport and diffusivity parametrization and high basal biomass m_0 . For each low diffusivity simulations (D10, D100 and D1000), we select the highest m_0 for which we find a biogeographical structure (Figure 4a). **(a-d)** Transport and diffusivity kernel $D = 10m^2.s^{-1}$, **(e-h)** Transport and diffusivity kernel $D = 100m^2.s^{-1}$, **(i-l)** Transport and diffusivity kernel $D = 1000m^2.s^{-1}$. **(a, e, i)** Richness defined as the mean number of coexisting species at each grid point, **(b, f, j)** Mean dominant species. Diatoms are in greens, other Eukaryotes in blues, *Prochlorococcus* in purples and *Synechococcus* in reds/oranges. The lighter the color is the larger the cell size of the species. **(c, g, k)** Biogeography: Principal Coordinate Analysis (PCoA) of the mean Bray-Curtis distance matrix of each grid point's species biomass over the year. **(d, h, l, p, t)** Mean integrated biomass over the year.



Supplementary Figure 10 | Modeled diversity in function of the number of particles or the reference biomass concentration of a particle for the 4 diffusivity kernels and with no grazing term. (a) Total diversity modeled in function of the reference biomass concentration of a particle. (b) Total diversity modeled in function of the mean number of particles in the last year of simulation. (c) Maximum local richness modeled in function of the mean number of particles in the last year of simulation. (d) Minimum local richness modeled in function of the mean number of particles in the last year of simulation.



Supplementary Figure 11 | Maximum growth rate and distribution of optimal growth temperature of surviving species for different transport and diffusivity parametrization and a grazing term. (a-b) No transport. (c-d) Transport and diffusivity constant of $10 \text{ m}^2.\text{s}^{-1}$. (e-f) Transport and diffusivity constant of $100 \text{ m}^2.\text{s}^{-1}$. (g-h) Transport and diffusivity constant of $1000 \text{ m}^2.\text{s}^{-1}$. (i-j) Transport and diffusivity constant of $10000 \text{ m}^2.\text{s}^{-1}$



Supplementary Figure 12 | Temperature differences between niches centroids mean temperature and optimal growth temperature for different transport and diffusivity parametrization and a grazing term. (a) No transport. (b) Transport and diffusivity constant of $10 \text{ m}^2.\text{s}^{-1}$. (c) Transport and diffusivity constant of $100 \text{ m}^2.\text{s}^{-1}$. (d) Transport and diffusivity constant of $1000 \text{ m}^2.\text{s}^{-1}$. (e) Transport and diffusivity constant of $10000 \text{ m}^2.\text{s}^{-1}$

CONCLUSION

Dans cette conclusion, je rappelle d'abord brièvement le contexte, les enjeux, la motivation et les questions scientifiques de cette thèse. Je résume ensuite les principaux résultats ainsi que les principaux points de discussion et perspectives qu'ils ont soulevés.

I. CONTEXTE, ENJEUX ET QUESTIONS SCIENTIFIQUES

Le plancton marin est un élément crucial de l'écosystème terrestre par son rôle dans la régulation du climat, la production primaire permettant le soutien de la chaîne alimentaire¹³ et par son rôle de pompe à carbone biologique contribuant au stockage de carbone dans les sédiments marins^{238,239,321}. Ces derniers sont notamment à l'origine du pétrole qui se forme sur des échelles de millions d'années. La combustion et l'utilisation pour diverses applications (plastique, industries...) à outrance des matières organiques fossiles, dont notamment le pétrole, sont à l'origine du changement climatique actuelle par effet de serre, effet connu depuis le XIX^{ème} siècle. Ce changement climatique se traduit à l'échelle planétaire notamment par un réchauffement de la température moyenne planétaire de l'atmosphère et des océans³⁰⁸ et l'augmentation en fréquence et intensité de nombreux événements climatiques extrêmes. Ces bouleversements sont aujourd'hui une menace pour l'écosystème terrestre et marin, les espèces qui les forment, allant des insectes à l'homme, et du plancton aux grands mammifères marins. Les activités humaines en particuliers sont grandement impactées par ces changements car elles dépendent à la fois du climat et des services écosystémiques directement liés au climat et à la biodiversité. Certains écosystèmes sauvages comme les océans participent à ces services comme la pêche qui dépend directement du plancton marin. Aujourd'hui pas moins de 3 milliards d'humains dépendent des activités de pêche³⁰¹. C'est dans ce contexte et face à ces enjeux que j'ai réalisé cette thèse.

Bien que le plancton ait été étudié dans le passé¹⁸, ce contexte de changement climatique rend plus que jamais nécessaire une meilleure compréhension du

plancton au sein du système planétaire. Ce contexte, l'avènement du séquençage d'ADN à haut débit et des omiques environnementales ont ainsi permis un apport de connaissances important sur le plancton et son importance écosystémique et climatique⁷⁰. Ainsi plusieurs grandes expéditions d'envergures^{35,39,40} dont certaines à l'échelle de l'ensemble des océans ont été menées comme les expéditions *Tara Océans*³⁹. C'est à l'aide des données omiques (métagénomiques et métatranscriptomiques) de cette expédition que j'ai pu soulever plusieurs questions d'importance toutes centrées sur la biogéographie du plancton et l'impact du réchauffement climatique sur les communautés de plancton. C'est donc ce fil rouge de la répartition des espèces planctoniques à différentes échelles biologiques et écologiques, l'impact du réchauffement climatique sur celle-ci et les rétroactions potentielles sur le climat qui ont guidés l'ensemble de cette thèse.

J'ai d'abord posé la question de la biogéographie du plancton à une échelle globale de l'ensemble des océans^{85,121,321}. J'ai ainsi abordé l'impact du réchauffement climatique sur la biogéographie (déplacements de communautés, impacts globaux) ainsi que les changements biogéochimiques potentielles de cet impact et pouvant avoir un effet de rétroaction sur le climat³²¹. J'ai ensuite abordé la question des changements fonctionnels et transcriptionnels du plancton eucaryote dans la transition de l'Atlantique Nord vers l'océan Arctique dans le contexte de la biogéographie génomique de ces deux régions et de l'Atlantification, phénomène lié au réchauffement climatique dans cette région. Je me suis particulièrement intéressé aux fonctions semblant jouer un rôle important dans l'accoutumance par changement d'expression de certains gènes des espèces persistantes dans le milieu Arctique. Enfin, j'ai posé la question d'une explication mécanistique de la biogéographie du plancton en développant un modèle d'écologie des communautés de phytoplancton à l'échelle du double gyre océanique de l'Atlantique Nord. J'ai ainsi exploré le comportement dynamique du modèle et les principales structures biogéographiques qui en émergeaient ainsi que l'influence des courants et de la diffusivité sur ces derniers avec notamment la sélection de la niche de température

des espèces constituant cette biogéographie. Je résumerai maintenant les résultats, discussion et les perspectives que ces questions ont amenées.

II. PRINCIPAUX RESULTATS, DISCUSSIONS ET PERSPECTIVES

Le premier chapitre a permis de mettre en évidence pour la première fois une biogéographie génomique du plancton à l'échelle des océans globaux^{85,321} (hormis l'océan Arctique). Un résultat majeur et méconnu par le passé est le découplage partiel entre les biogéographies des plus petits organismes planctoniques tels que les virus, bactéries et petits eucaryotes et les biogéographies des plus grands organismes planctoniques, principalement le zooplancton (les copépodes majoritairement). Globalement les biogéographies de tous les organismes planctoniques sont organisées de manière latitudinale et suivent le gradient latitudinal de température. Les plus petits organismes ont cependant des biogéographies plus complexes que celles des grands organismes avec une organisation à l'échelle des grandes structures océaniques tel que les gyres oligotrophes ou encore l'upwelling équatoriale pacifique. En revanche, les grands organismes, jusqu'à 2mm, ont une organisation plus globale et transocéanique à l'échelle du globe entier. Ce résultat implique nécessairement un découplage trophique avec l'existence d'espèces zooplanctoniques généralistes se dispersant sur de vaste régions océaniques et trans-bassins. Ce résultat va à l'encontre de l'hypothèse que les petits organismes ont une plus grande capacité de diffusion en raison de leur plus grands nombre d'individu^{159,347}. Nous suggérons plutôt que ce résultat est lié aux différences entre les cycles de vie des grands et petits organismes, les premiers ayant des durées de vies de plusieurs mois voire années et les seconds ayant des cycles très courts de l'ordre du jour ou quelques jours⁸⁵. Le zooplancton peut ainsi parcourir de plus grandes distances lors de son transport par les courants. Un second résultat majeur est la mise en évidence des espèces, et associées à des génomes¹²¹, structurant globalement ces biogéographies et ce pour les différentes tailles de plancton. Ce résultat rend concret les biogéographies observées en leur donnant un support biologique et écologique. Il ouvre ainsi à de nombreuses

questions écologiques allant de l'étude des différences fonctionnelle^{125,127} (contenu génomique, activité transcriptionnelle), de microévolution¹⁹⁶⁻¹⁹⁸ (analyses de polymorphismes génétiques), de biologie structurale³⁴⁸ (biogéographie des repliements protéiques), de la biomasse de ces organismes, ou encore d'analyse des niches métaboliques de ces organismes^{162,163} ou de leurs capacités biosynthétiques¹³⁴. Le second chapitre explore notamment l'activité transcriptionnelle d'organismes persistant le long d'un fort gradient environnemental bien qu'avec une référence de génomes légèrement différente.

Un second développement majeur du premier chapitre est l'étude de l'impact du changement climatique sur les biogéographies que je viens de décrire. Il a ainsi été mis en évidence que selon un scénario de réchauffement climatique important (scénario RCP8.5⁶¹), jusqu'à 50% de l'océan étudié pourrait connaître un changement important de communautés de plancton. Ce résultat implique notamment l'apparition potentiel de nouvelles interactions biotique par le découplage déjà existant entre petits et grands organismes planctoniques impliquant nécessairement un découplage dans leur réponse au changement climatique. De manière intéressante, les taux de déplacement des communautés de plancton (provinces génomiques du chapitre I), sont similaires aux taux de déplacement de certains biomes phytoplanctoniques calculés par les modèles biogéochimiques seuls³⁴⁹. Ce résultat n'était pas trivial et renforce la confiance dans les modèles biogéochimiques mais aussi l'utilisation de données génomiques en couplage avec ces derniers. Un second résultat majeur est la possibilité de changements biogéochimiques liés à la restructuration de la biogéographie du plancton aussi en accord avec les modèles biogéochimiques. Un premier changement est l'augmentation en diazotrophes dans les zones tropico-équatoriales pacifiques, effet prévu par les modèles²⁷¹ et conforté ici par la génomique. Un second changement projeté via la génomique est une baisse d'environ 4% de la pompe à carbone biologique d'ici la fin du siècle effet aussi prévu par les modèles bien qu'avec de grandes incertitudes^{255,350}. Un tel changement aurait un effet aggravant sur le réchauffement climatique. Pour conclure sur les

résultats de ce premier chapitre, la capacité de la génomique à nous informer sur la biogéographie du plancton et sur l'impact du changement climatique sur celle-ci a été démontrée. Pour ouvrir des perspectives plus larges, on peut penser à réaliser une biogéographie mieux résolue spatialement et temporellement ouvrant la perspective à de nouvelles grandes expéditions ou même de génomique *in situ* (type « genomic ARGO »). Deuxièmement, sur le changement climatique, les questions de la réponse des communautés de plancton dans les zones dépassant les extrêmes environnementaux actuelles (brièvement abordé dans la discussion de l'article) ou encore en réponse aux points de basculement du changement climatique restent ouvertes³⁵¹.

Dans le second chapitre, j'ai exploré les différences fonctionnelles et transcriptionnelles des communautés de plancton eucaryotes le long de la transition entre l'océan Atlantique Nord et l'océan Arctique. La biogéographie métagénomique des deux bassins a d'abord été décrites et mises en lien avec la connectivité océanique de la région. L'approche génomique a ainsi permis de mettre en évidence une forte différenciation entre les deux bassins et une transition au niveau du front polaire enrichie en espèces communes aux deux bassins comme cela est connu aux niveaux de tels fronts océaniques. Un premier résultat majeur est la mise en évidence d'une réponse transcriptionnelle forte et commune d'espèces de grands clades d'algues planctoniques eucaryotes abondants dans les deux bassins et donc persistants au front polaire. Cette réponse se traduit par l'augmentation de l'expression de certains gènes le long du gradient de température (paramètre expliquant le mieux ces changements par rapports aux autres gradients physico-chimiques) du front polaire séparant les deux bassins océaniques. Cet ensemble de gènes est apparu enrichi en un grand nombre de fonctions dont un grand nombre étaient préalablement décrites comme liées directement à l'acclimatation au froid, la régulation de l'expression génétique (de la voie de signalisation jusqu'aux modifications post traductionnelles), la division cellulaire, la consommation d'énergie, le repliement et le recyclage des protéines, la réparation de l'ADN ou encore la

division cellulaire. De manière intéressantes les mécanismes d'acclimatation au froid se sont avérées assez variées et sont des fonctions connues chez certaines plantes terrestres. Ce résultat montre la conservation de mécanismes ancestraux et donc de stratégies communes à l'accoutumance au froid chez les organismes photosynthétiques à la fois terrestre et planctoniques marins. D'autre part l'enrichissement en fonctions liées à la régulation et à des processus d'expression génétique suggère que l'entrée dans l'arctique provoque d'important changements transcriptionnels chez les organismes de l'Atlantique nord qui sont persistants dans l'Arctique. Les mécanismes de régulation et liés à l'expression génétique sont apparus aussi comme des fonctions plus abondantes à l'échelle de l'ensemble de la communauté dans les métatranscriptomes Arctiques par rapport à ceux de l'Atlantique soulignant l'importance de ces mécanismes dans cet environnement. Parmi ces mécanismes, la méthylation des ARNs semble particulièrement importante en réponse au froid chez les plantes³³² et a déjà été relié à la stabilité des ARNs³⁵² et la fidélité de la traduction des protéines³⁵³. D'autres fonctions plus abondantes dans les métatranscriptomes Arctiques ont été mises en évidence et discutées tel que des fonctions liées au métabolisme des lipides, mis en lien avec une réponse au froid (catabolisme des lipides³³⁷) et la fluidité membranaire^{335,336}, à la maintenance des télomères, la rotation du pool protéique cellulaire.

Pour conclure sur ce deuxième chapitre, la mise en évidence *in situ* d'une réponse transcriptionnelle à l'échelle de la communauté eucaryote est, sans être une première, un apport important permis par les données génomiques environnementales. En effet, elles mettent en évidence des mécanismes fondamentaux permettant de mieux expliquer la niche écologique de certains organismes. Le contexte de l'Atlantification^{326,327,329,354} lié au changement climatique, rendant le bassin Arctique eurasiatique plus similaire au bassin Atlantique, se traduit par un plus grand rôle des influx par les courants marins Atlantiques³²⁶ se réchauffant et entraînant l'invasion d'espèces tempérées^{329,330,354}. L'échelle génomique permet d'apporter des réponses sur les changements physiologiques de ces espèces et sur

leur adaptation à ce nouveau climat. Au-delà des changements transcriptionnels étudiés ici, un couplage avec une étude des changements au niveau de la séquence de l'ADN et *a fortiori* des protéines^{196,198} permettraient aussi de mieux comprendre les mécanismes globaux d'adaptation des espèces persistantes ou dont la biogéographie s'étend vers l'Arctique avec le changement climatique. Enfin le lien entre les changements des courants liés au changement climatique comme le Gulf Stream dont la circulation pourrait ralentir voire s'effondrer^{180,181,351,355} est à investiguer. En effet, hypothétiquement, un effondrement de la circulation pourrait entraîner une isolation des communautés planctoniques arctique par une diminution de la connectivité physique par les courants.

Enfin dans le troisième chapitre, nous avons développé les prémices d'un modèle basé individu de communauté phytoplanctonique advectée dans une circulation océanique idéalisée de l'Atlantique Nord. Le modèle basé individu a été couplé avec un forçage physique (température, lumière) et géochimique (un nutritif limitant, le nitrate). Les structures biogéographiques émergents de la dynamique du modèle ont été explorées mettant en évidence un rôle important des courants (circulation du double gyre) et de la diffusivité. Une structuration nord-sud a émergé du modèle avec des espèces de diatomées et petits eucaryotes dominantes dans le gyre nord et des cyanobactéries dominantes dans le gyre sud ce qui est observé dans l'océan. En effet, les plus petites cellules (cyanobactéries) ont des meilleures affinités pour les éléments nutritifs leur permettant de mieux survivre dans les régions oligotrophes. La circulation océanique en double gyre a aussi permis de recréer une forte diversité locale au niveau des fronts océaniques. Par ailleurs, un effet inattendu non-linéaire du nombre de particules utilisées dans le modèle a été mis en évidence. L'augmentation du nombre de particules engendre la déstructuration de la biogéographie observée et donc une homogénéisation de la communauté phytoplanctonique à l'échelle de l'ensemble du double gyre de l'Atlantique Nord et accompagné d'une augmentation importante de la diversité totale d'espèces représentées dans le modèle. Ce résultat semble s'expliquer en grande partie par

l'invasion d'une espèce à l'ensemble du gyre à la fin de la première année de simulation et forçant une synchronisation avec la dynamique des autres espèces. Ce résultat est potentiellement artefactuel et lié à certaines limites du modèle comme l'absence d'efflux aux limites nord et sud du double gyre, le pas de temps et la résolution spatiale du modèle, la dépendance à un seul nutritif ou encore un cycle saisonnier constant. Néanmoins nous suggérons que ce résultat reste intéressant notamment au regard du paradoxe du plancton^{356,357}. Ce paradoxe s'étonne en effet de la grande diversité observée du plancton par rapport au nombre faible de ressources pour lesquelles les différentes espèces sont en compétition et en vertu du principe d'exclusion compétitive. Je suggère ainsi que le nombre de particules dans notre modèle et par extension le nombre de cellules planctoniques peut influencer grandement la diversité totale d'espèces observées et la diversité locale d'espèces pouvant coexister dans un système de circulation océanique. Intuitivement, l'augmentation du nombre de particules permet d'augmenter le nombre de chemins possibles pour la survie de la particule (dans mon modèle) ou d'une cellule planctonique (dans l'océan). On peut mettre en lien cette diversité émergente avec la diversité intra-espèce déterminé par le taux de mutation d'une espèce mais aussi la taille de la population de cette espèce.

Pour conclure, ce modèle offre des premiers résultats prometteurs et offre de multiples voies de développement comme la complexification du terme de broutage¹⁴⁹, l'ajout de nouveaux nutritifs limitants¹⁴⁹, l'ajout d'un module de lyse virale³⁵⁸⁻³⁶¹, d'un module d'adaptation^{150,160,340}, l'advection dans un modèle de circulation globale^{188,191} ou encore le couplage avec un modèle biogéochimique de changement climatique³⁶². Tous ces développements permettront d'explorer de nombreuses questions théoriques écologiques comme l'impact de la température sur la niche écologique³⁶³, l'importance de la lyse virale³⁵⁸⁻³⁶¹ dans le contrôle des blooms océaniques ou encore l'adaptation à l'échelle d'une communauté composée de plusieurs espèces en compétition et face au réchauffement climatique. Ces explorations théoriques devront aussi toujours être comparées aux observations et

notamment aux apports importants des données génomiques.

Pour les derniers mots de mon mémoire de thèse, je conclurais que la recherche fondamentale en écologie, génomique et en climatologie, qu'elle soit terrestre ou marine, semble plus que jamais nécessaire face au défi de ce siècle qu'est le changement climatique. Le travail et le dialogue constant avec les climatologues mais aussi les politiques semblent aussi primordiaux. Cette thèse est, je pense, un très bon exemple des liens forts entre la biogéochimie, qui est une science de la terre, l'écologie, qui est la science des organismes vivants en interaction avec leur milieu, et la génomique, qui est la science des génomes. La recherche dans ces domaines permet de mieux comprendre le monde qui nous entoure et le système Terre en général, de mieux anticiper ses changements, *a fortiori* de mieux s'y adapter et peut être tout simplement de mieux y vivre.

REFERENCES

1. Hensen, V. *Über die Bestimmung des Planktons oder des im Meere treibenden Materials an Pflanzen und Thieren. V. Bericht der Commission zur Wissenschaftlichen Untersuchung der Deutschen Meere, Jahrgang* vol. 53 (1887).
2. Lalli, C. M. & Parsons, T. R. *Biological oceanography: an introduction*. (1993) doi:10.1016/0022-0981(96)02604-4.
3. Brock, T. D. SIEBURTH, J. McN. 1979. *Sea microbes*. Oxford University Press, New York, x + 491 p. *Limnol Oceanogr* **25**, (1980).
4. Sullivan M *et al.* Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLoS Biol* **3**, null (2005).
5. Alsante, A. N., Thornton, D. C. O. & Brooks, S. D. Ocean Aerobiology. *Front Microbiol* **12**, (2021).
6. Korunic, Z. Diatomaceous Earths - Natural Insecticides. *Pesticidi i fitomedicina* **28**, 77–95 (2013).
7. Maranda, L. Infection of *Prorocentrum minimum* (Dinophyceae) by the parasite *Amoebophrya* sp (Dinoflagellea). *J Phycol* **37**, 245–248 (2001).
8. Hansen, B. *et al.* Acute toxicity of dispersed crude oil on the cold-water copepod *Calanus finmarchicus*: Elusive implications of lipid content. *J Toxicol*

- Environ Health A* **79**, 549–557 (2016).
9. Bentlage, B. & Lewis Ames, C. An illustrated key and synopsis of the families and genera of carybdeid box jellyfishes (Cnidaria: Cubozoa: Carybdeida), with emphasis on the "Irukandji family" (Carukiidae). *J Nat Hist* **46**, (2012).
 10. Walsh, E. J. Biochemistry (Garrett, Reginald H.; Grisham, Charles M.). *J Chem Educ* **74**, (1997).
 11. Fenchel, T. Marine plankton food chains. *Annual review of ecology and systematics*. Vol. 19 (1988) doi:10.1146/annurev.ecolsys.19.1.19.
 12. Roig, X. Aquatic Photosynthesis. Paul G. Falkowski, John A. Raven. *International microbiology: official journal of the Spanish Society for Microbiology* **3**, (2000).
 13. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* (1979) (1998) doi:10.1126/science.281.5374.237.
 14. Steinberg, D. K. & Landry, M. R. Zooplankton and the Ocean Carbon Cycle. *Ann Rev Mar Sci* **9**, (2017).
 15. Wiebe, P. H. & Benfield, M. C. From the Hensen net toward four-dimensional biological oceanography. *Progress in Oceanography* vol. 56 Preprint at [https://doi.org/10.1016/S0079-6611\(02\)00140-4](https://doi.org/10.1016/S0079-6611(02)00140-4) (2003).
 16. Damkaer, D. M. John Vaughan Thompson (1779-1847), pioneer planktonologist: a life renewed. *Journal of Crustacean Biology* **36**, (2016).
 17. Caspers, H. Unesco: Zooplankton Sampling. - Monographs on oceanographic methodology 2. - 174 pp. Paris: The Unesco Press 1974 (second impression). ISBN 92-3-101 194-4 \$ 9.50. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* **65**, (1980).
 18. Fleming, R. H. Oceanography: The past. *Deep Sea Research Part A. Oceanographic Research Papers* **28**, (1981).
 19. (1872-1876), C. E. et al. *Report on the scientific results of the voyage of H.M.S. Challenger during the years 1873-76 under the command of Captain George S. Nares ... and the late Captain Frank Tourle Thomson, R.N.* vol. Narrative (Edinburgh:Neill, 1885).
 20. Cleve, P. T. & Cleve, P. T. *Plankton Collected by the Swedish Expedition to Spitzbergen in 1898.* (P.A. Norstedt & Söner, 1899).
 21. T. Editors of Encyclopedia Britannica. Per Teodor Cleve. *Encyclopedia Britannica*.

- (2022).
22. Bjørklund, K. R., Itaki, T. & Dolven, J. K. Per Theodor Cleve: a short résumé and his radiolarian results from the Swedish Expedition to Spitsbergen in 1898. *J Micropalaeontol* **33**, (2014).
 23. Groeben, C. The Stazione Zoologica Anton Dohrn as a place for the circulation of scientific ideas: vision and management. In *Information for Responsible Fisheries: Libraries as Mediators: proceedings of the 31st Annual Conference* (2006).
 24. Barton, A. D., Irwin, A. J., Finkel, Z. V. & Stock, C. A. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proceedings of the National Academy of Sciences* (2016) doi:10.1073/pnas.1519080113.
 25. Jonas, T. D., Walne, A., Beaugrand, G., Gregory, L. & Hays, G. C. The volume of water filtered by a Continuous Plankton Recorder sample: The effect of ship speed. *Journal of Plankton Research* vol. 26 Preprint at <https://doi.org/10.1093/plankt/fbh137> (2004).
 26. Defant, A. *Das Kaltwasserauftriebsgebiet vor der Küste Südwes Afrikas, Landerkundliche Forschung, Festschrift Nobert Krebs zur Vollendung des 60 Lebensjahres dargebracht.*, (1936).
 27. Sverdrup, H. U. On the Process of Upwelling. *J Mar Res* **1**, 155–164 (1938).
 28. Sverdrup, H. U. and Fleming, R. H. The waters of the coast of southern California, March to July 1937. *J Mar Res* **4**, (1941).
 29. Barnes, H. On the volume measurement of water filtered by a plankton pump, with some observations on the distribution of planktonic animals. *Journal of the Marine Biological Association of the United Kingdom* **28**, (1949).
 30. Hopkins, T. L. The variation in the catch of plankton nets in a system of estuaries. *Deep Sea Research and Oceanographic Abstracts* **11**, (1964).
 31. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, (1953).
 32. «Rosalind Franklin, génie incompris réhabilitée par Google».
 33. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, (1977).
 34. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & Delong, E. F. Characterization of

- uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178**, (1996).
35. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**, (2007).
 36. Yooseph, S. *et al.* The Sorcerer II global ocean sampling expedition: Expanding the universe of protein families. *PLoS Biol* **5**, (2007).
 37. Barberán, A., Fernández-Guerra, A., Bohannon, B. J. M. & Casamayor, E. O. Exploration of community traits as ecological markers in microbial metagenomes. *Mol Ecol* **21**, (2012).
 38. Gianoulis, T. A. *et al.* Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc Natl Acad Sci U S A* **106**, (2009).
 39. Karsenti, E. *et al.* A holistic approach to marine Eco-systems biology. *PLoS Biol* (2011) doi:10.1371/journal.pbio.1001177.
 40. Duarte, C. M. Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin* Preprint at <https://doi.org/10.1002/lob.10008> (2015).
 41. Murphy, R. E., Ardanuy, P., Deluccia, F. J., Clement, J. E. & Schueler, C. F. The Visible Infrared Imaging Radiometer Suite. In *Earth Science Satellite Remote Sensing: Vol. 1: Science and Instruments* (eds. Qu, J. J., Gao, W., Kafatos, M., Murphy, R. E. & Salomonson, V. V) 199–223 (Springer Berlin Heidelberg, 2006). doi:10.1007/978-3-540-37293-6_11.
 42. Yala, K. *et al.* Estimation of phytoplankton pigments from ocean-color satellite observations in the Senegalo-Mauritanian region by using an advanced neural classifier. *Ocean Science* **16**, (2020).
 43. Cokacar, T., Oguz, T. & Kubilay, N. Satellite-detected early summer coccolithophore blooms and their interannual variability in the Black Sea. *Deep Sea Res 1 Oceanogr Res Pap* **51**, (2004).
 44. Villar, E. *et al.* Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science (1979)* **348**, 1261447 (2015).
 45. Wong, A. P. S. *et al.* Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats. *Frontiers in Marine Science* vol. 7 Preprint at <https://doi.org/10.3389/fmars.2020.00700> (2020).
 46. Claustre, H., Johnson, K. S. & Takeshita, Y. Observing the Global Ocean with

- Biogeochemical-Argo. *Annual Review of Marine Science* vol. 12 Preprint at <https://doi.org/10.1146/annurev-marine-010419-010956> (2020).
47. Locarnini, R. A. *et al.* World Ocean Atlas 2018, Volume 1: Temperature. *NOAA Atlas NESDIS 81* **1**, (2018).
 48. Zweng, M. M. *et al.* World Ocean Atlas 2018 Volume 2: Salinity. *NOAA Atlas NESDIS* vol. 82 (2019).
 49. Garcia, H. E. *et al.* World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. *NOAA Atlas NESDIS* **3**, (2019).
 50. Garcia, H. *et al.* World Ocean Atlas 2018. Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). *NOAA Atlas NESDIS* **84**, (2018).
 51. Gross, L. A New Human Genome Sequence Paves the Way for Individualized Genomics. *PLoS Biol* **5**, 1–2 (2007).
 52. <https://www.embl.org/news/science/we-all-come-from-the-ocean/>.
 53. <https://www.flickr.com/photos/arselectronica/10398819016>.
 54. <https://ethz.ch/en/news-and-events/eth-news/news/2015/10/computing-for-climate-part-2-how-climate-models-work.html>.
 55. Hjfreland, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons.
 56. Lesbats Stephane (2019). Mise en oeuvre de la CTD rosette ou bathysonde. Ifremer. <https://image.ifremer.fr/data/00621/73314/>.
 57. [https://commons.wikimedia.org/wiki/File:The_schooner_TARA_\(Port_Lay,_%C3%84Ele_de_Groix,_2009\).jpg](https://commons.wikimedia.org/wiki/File:The_schooner_TARA_(Port_Lay,_%C3%84Ele_de_Groix,_2009).jpg).
 58. https://upload.wikimedia.org/wikipedia/commons/2/23/201511_genomesequencer9.png.
 59. Longhurst, A. R. *Ecological Geography of the Sea. Ecological Geography of the Sea* (2007). doi:10.1016/B978-0-12-455521-1.X5000-1.
 60. Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci Model Dev* (2015) doi:10.5194/gmd-8-2465-2015.

61. van Vuuren, D. P. *et al.* The representative concentration pathways: An overview. *Clim Change* (2011) doi:10.1007/s10584-011-0148-z.
62. Hutchinson, G. E. Concluding remarks. *Cold Spring Harb Symp Quant Biol* **22**, 415–427 (1957).
63. Brun, P. *et al.* Ecological niches of open ocean phytoplankton taxa. *Limnol Oceanogr* **60**, (2015).
64. Benedetti, F., Guilhaumon, F., Adloff, F. & Ayata, S. D. Investigating uncertainties in zooplankton composition shifts under climate change scenarios in the Mediterranean Sea. *Ecography* (2018) doi:10.1111/ecog.02434.
65. Benedetti, F. *et al.* Major restructuring of marine plankton assemblages under global warming. *Nat Commun* **12**, (2021).
66. Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. & Levin, S. A. Marine taxa track local climate velocities. *Science* (1979) (2013) doi:10.1126/science.1239352.
67. Hofmann Elizondo, U., Righetti, D., Benedetti, F. & Vogt, M. Biome partitioning of the global ocean based on phytoplankton biogeography. *Prog Oceanogr* **194**, (2021).
68. Alberti, A. *et al.* Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci Data* (2017) doi:10.1038/sdata.2017.93.
69. <https://www.nzherald.co.nz/nz/sir-peter-blake-killed-in-amazon-pirate-attack/XJPSKTOY5EN2C5VGG2SMWZJ2PY/>.
70. Sunagawa, S. *et al.* Tara Oceans: towards global ocean ecosystems biology. *Nature Reviews Microbiology* Preprint at <https://doi.org/10.1038/s41579-020-0364-5> (2020).
71. Karsenti, E. *Aux sources de la vie.* (Flammarion, 2018).
72. Rigonato, J. *et al.* Insights into biotic and abiotic modulation of ocean mesopelagic communities. *bioRxiv* (2021).
73. Gorsky, G. *et al.* Digital zooplankton image analysis using the ZooScan integrated system. *J Plankton Res* **32**, 285–303 (2010).
74. Sieracki, C. K., Sieracki, M. E. & Yentsch, C. S. An imaging-in-flow system for automated analysis of marine microplankton. *Mar Ecol Prog Ser* **168**, (1998).
75. Gendron, M.-C. Cytométrie en flux. *EMC - Biologie Médicale* **1**, (2006).

76. Picheral, M. *et al.* The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnol Oceanogr Methods* **8**, 462–473 (2010).
77. Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* (2015) doi:10.1038/sdata.2015.23.
78. Peters, A., Hothorn, T. & Lausen, B. ipred: Improved Predictors. *R News* **2**, (2002).
79. Breiman, L. & Cutler, A. Breiman and Cutler's random forests for classification and regression. *Package 'randomForest'* (2012) doi:10.5244/C.22.54.
80. Wallace, C. S. & Patrick, J. D. Coding Decision Trees. *Mach Learn* **11**, (1993).
81. Kitahashi, T., Watanabe, H. K., Tsuchiya, M., Yamamoto, H. & Yamamoto, H. A new method for acquiring images of meiobenthic images using the FlowCAM. *MethodsX* **5**, (2018).
82. Sun, L., Wu, H., Pan, B., Wang, B. & Guo, W. Evaluation and validation of a novel 10-color flow cytometer. *J Clin Lab Anal* **35**, (2021).
83. Flombaum, P., Wang, W. L., Primeau, F. W. & Martiny, A. C. Global picophytoplankton niche partitioning predicts overall positive response to ocean warming. *Nat Geosci* (2020) doi:10.1038/s41561-019-0524-2.
84. Busseni, G. *et al.* Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean. *Global Ecology and Biogeography* (2020) doi:10.1111/geb.13161.
85. Richter, D. J. *et al.* Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *Elife* **11**, e78129 (2022).
86. Sieracki, M. E. *et al.* Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. *Sci Rep* **9**, (2019).
87. Seeleuthner, Y. *et al.* Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat Commun* **9**, (2018).
88. Guidi, L. *et al.* Environmental context of all samples from the Tara Oceans Expedition (2009-2013), about nutrients in the targeted environmental feature. in (PANGAEA, 2017). doi:10.1594/PANGAEA.875575.
89. Guidi, L., Picheral, M., Pesant, S., Tara Oceans Consortium, C. & Tara Oceans Expedition, P. Environmental context of all samples from the Tara Oceans

- Expedition (2009-2013), about sensor data in the targeted environmental feature. in (PANGAEA, 2017). doi:10.1594/PANGAEA.875576.
90. Speich, S. *et al.* Environmental context of all samples from the Tara Oceans Expedition (2009-2013), about the water column features at the sampling location. in (PANGAEA, 2017). doi:10.1594/PANGAEA.875579.
 91. Tara Oceans Consortium, C. & Tara Oceans Expedition, P. Registry of all samples from the Tara Oceans Expedition (2009-2013). Preprint at <https://doi.org/10.1594/PANGAEA.875582> (2017).
 92. Lacour, L. *et al.* Unexpected winter phytoplankton blooms in the North Atlantic subpolar gyre. *Nat Geosci* **10**, (2017).
 93. <https://www.illumina.com/science/technology/next-generation-sequencing/illumina-sequencing-history.html>.
 94. <http://news.bbc.co.uk/2/hi/science/nature/2940601.stm>.
 95. Nurk, S. *et al.* The complete sequence of a human genome. *Science (1979)* **376**, 44–53 (2022).
 96. Dulbecco, R. A turning point in cancer research: sequencing the human genome. *Science (1979)* **231**, (1986).
 97. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, (2012).
 98. Wohns, A. W. *et al.* A unified genealogy of modern and ancient genomes. *Science (1979)* **375**, eabi8264 (2022).
 99. Kaul, S. *et al.* Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, (2000).
 100. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, (2002).
 101. Jaillon, O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, (2007).
 102. Gupta, N. & Verma, V. K. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. in (2019). doi:10.1007/978-981-13-8844-6_15.
 103. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* vol. 449 Preprint at <https://doi.org/10.1038/nature06244> (2007).

104. Van Rossum, T. V. *et al.* Year-long metagenomic study of river microbiomes across land use and water quality. *Front Microbiol* **6**, (2015).
105. Daniel, R. The metagenomics of soil. *Nature Reviews Microbiology* vol. 3 Preprint at <https://doi.org/10.1038/nrmicro1160> (2005).
106. Schneider, D., Aßmann, N., Wicke, D., Poehlein, A. & Daniel, R. Metagenomes of Wastewater at Different Treatment Stages in Central Germany. *Microbiol Resour Announc* **9**, (2020).
107. Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270**, (2003).
108. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol* **5**, (1998).
109. Linné Carl von. *Species plantarum: exhibentes plantas rite cognitatas ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. vol. t.1 (1753) (Berlin:Junk, 1753).
110. Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol Ecol Resour* **15**, (2015).
111. Sipos, R. *et al.* Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol Ecol* **60**, (2007).
112. McLaren, M. R., Willis, A. D. & Callahan, B. J. Consistent and correctable bias in metagenomic sequencing experiments. *Elife* **8**, (2019).
113. de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science (1979)* **348**, 1261605 (2015).
114. Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* (2019) doi:10.1016/j.cell.2019.10.008.
115. Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science (1979)* (2015) doi:10.1126/science.1262073.
116. Chaffron, S. *et al.* Environmental vulnerability of the global ocean epipelagic plankton community interactome. *Sci Adv* **7**, (2021).
117. <https://www.france-genomique.org/technological->

- expertises/metagenomics/metagenomics-using-targeted-amplicon-barcoding-metabarcoding/?lang=en.
118. <https://www.france-genomique.org/technological-expertises/metagenomics/shotgun-metagenomics/?lang=en>.
 119. Gweon, H. S. *et al.* The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environ Microbiome* **14**, 7 (2019).
 120. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology* vol. 8 Preprint at <https://doi.org/10.3389/fmicb.2017.02224> (2017).
 121. Delmont, T. O. *et al.* Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics* **2**, 100123 (2022).
 122. Coelho, L. P. *et al.* Towards the biogeography of prokaryotic genes. *Nature* **601**, (2022).
 123. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* vol. 10 Preprint at <https://doi.org/10.1038/nrg2484> (2009).
 124. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science (1979)* **348**, (2015).
 125. Salazar, G. *et al.* Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* (2019) doi:10.1016/j.cell.2019.10.014.
 126. Carradec, Q. *et al.* A global ocean atlas of eukaryotic genes. *Nat Commun* (2018) doi:10.1038/s41467-017-02342-1.
 127. Caputi, L. *et al.* Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Global Biogeochem Cycles* **33**, (2019).
 128. Busseni, G. *et al.* Meta-Omics Reveals Genetic Flexibility of Diatom Nitrogen Transporters in Response to Environmental Changes. *Mol Biol Evol* (2019) doi:10.1093/molbev/msz157.
 129. Vorobev, A. *et al.* Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities via high-throughput metagenomics and metatranscriptomics. *Genome Res* **30**, (2020).

130. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
131. Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**, 533–538 (2013).
132. Delmont, T. O. *et al.* Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *ISME J* (2021) doi:<https://doi.org/10.1038/s41396-021-01135-1>.
133. Royo-Llonch, M. *et al.* Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol* **6**, 1561–1574 (2021).
134. Paoli, L. *et al.* Biosynthetic potential of the global ocean microbiome. *Nature* **607**, 111–118 (2022).
135. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* (2018) doi:[10.1038/s41564-018-0176-9](https://doi.org/10.1038/s41564-018-0176-9).
136. Faure, E., Ayata, S.-D. & Bittner, L. Towards omics-based predictions of planktonic functional composition from environmental data. *Nat Commun* **12**, 4361 (2021).
137. Frémont, P. *et al.* Restructuring of genomic provinces of surface ocean plankton under climate change. *bioRxiv* 2020.10.20.347237 (2021) doi:[10.1101/2020.10.20.347237](https://doi.org/10.1101/2020.10.20.347237).
138. Nelson, W. C., Tully, B. J. & Mobberley, J. M. Biases in genome reconstruction from metagenomic data. *PeerJ* **8**, (2020).
139. Costea, P. I. *et al.* Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol* **35**, (2017).
140. Istace, B. *et al.* de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* (2017) doi:[10.1093/gigascience/giw018](https://doi.org/10.1093/gigascience/giw018).
141. Bickhart, D. M. *et al.* Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol* (2022) doi:[10.1038/s41587-021-01130-z](https://doi.org/10.1038/s41587-021-01130-z).
142. Sahlin, K., Sipos, B., James, P. L. & Medvedev, P. Error correction enables use of

- Oxford Nanopore technology for reference-free transcriptome analysis. *Nat Commun* **12**, (2021).
143. Wang, L., Qu, L., Yang, L., Wang, Y. & Zhu, H. NanoReviser: An Error-Correction Tool for Nanopore Sequencing Based on a Deep Learning Algorithm. *Front Genet* **11**, (2020).
 144. Humboldt, A. von. *Cosmos: a sketch of a physical description of the universe*. vol. 1 (Harper & Brothers, 1866).
 145. Steele, J. H. The ocean 'landscape'. *Landsc Ecol* (1989) doi:10.1007/BF00131537.
 146. Kavanaugh, M. T. *et al.* Seascapes as a new vernacular for pelagic ocean monitoring, management and conservation. *ICES Journal of Marine Science* (2016) doi:10.1093/icesjms/fsw086.
 147. Okubo, A. Oceanic diffusion diagrams. *Deep Sea Research and Oceanographic Abstracts* **18**, 789–802 (1971).
 148. Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J. & Follows, M. J. Patterns of diversity in marine phytoplankton. *Science* (1979) **327**, (2010).
 149. Dutkiewicz, S. *et al.* Dimensions of marine phytoplankton diversity. *Biogeosciences* (2020) doi:10.5194/bg-17-609-2020.
 150. Sauterey, B., Ward, B., Rault, J., Bowler, C. & Claessen, D. The Implications of Eco-Evolutionary Processes for the Emergence of Marine Plankton Community Biogeography. *Am Nat* (2017) doi:10.1086/692067.
 151. Steele, J. H. A comparison of terrestrial and marine ecological systems. *Nature* **313**, 355–358 (1985).
 152. Caswell, H. & Cohen, J. E. Red, white and blue: environmental variance spectra and coexistence in metapopulations. *J Theor Biol* **176**, (1995).
 153. Stokes, G. G. On the Effect of the Internal Friction of Fluids on the Motion of Pendulums. *Transactions of the Cambridge Philosophical Society* **9**, 8–106 (1851).
 154. Abbott, M. R. & Letelier, R. M. Decorrelation scales of chlorophyll as observed from bio-optical drifters in the California Current. *Deep Sea Research Part II: Topical Studies in Oceanography* **45**, 1639–1667 (1998).
 155. Lovejoy, S. *et al.* Universal multifractals and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity. *J Plankton Res* **23**, 117–141 (2001).
 156. KEISTER, J. E., DI LORENZO, E., MORGAN, C. A., COMBES, V. & PETERSON, W. T.

- Zooplankton species composition is linked to ocean transport in the Northern California Current. *Glob Chang Biol* **17**, (2011).
157. Lindsey, B. J. & Batchelder, H. P. Cross-shelf distribution of *Euphausia pacifica* in the Oregon coastal upwelling zone: field evaluation of a differential transport hypothesis. *J Plankton Res* **33**, (2011).
 158. Kiørboe, T. & Jiang, H. To eat and not be eaten: optimal foraging behaviour in suspension feeding copepods. *J R Soc Interface* **10**, (2013).
 159. Finlay, B. J. Global Dispersal of Free-Living Microbial Eukaryote Species. *Science (1979)* **296**, 1061 LP – 1063 (2002).
 160. Hellweger, F. L., van Sebille, E. & Fredrick, N. D. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science (1979)* (2014) doi:10.1126/science.1254421.
 161. Prigent, S. *et al.* The genome-scale metabolic network of *Ectocarpus siliculosus* (EctoGEM): A resource to study brown algal physiology and beyond. *Plant Journal* **80**, (2014).
 162. Budinich, M., Bourdon, J., Larhlimi, A. & Eveillard, D. A multi-objective constraint-based approach for modeling genome-scale microbial ecosystems. *PLoS One* **12**, (2017).
 163. Régimbeau, A. *et al.* Contribution of genome scale metabolic modeling to niche theory. *bioRxiv* 2021.07.21.453190 (2021) doi:10.1101/2021.07.21.453190.
 164. Doblin, M. A. & Van Sebille, E. Drift in ocean currents impacts intergenerational microbial exposure to temperature. *Proc Natl Acad Sci U S A* **113**, (2016).
 165. Zaiss, J., Boyd, P. W., Doney, S. C., Havenhand, J. N. & Levine, N. M. Impact of Lagrangian Sea Surface Temperature Variability on Southern Ocean Phytoplankton Community Growth Rates. *Global Biogeochem Cycles* **35**, (2021).
 166. Cushman-Roisin, B. & Beckers, J. M. *Introduction to Geophysical Fluid Dynamics: Physical and Numerical Aspects*. (Elsevier Science, 2011).
 167. Sarmiento, J. L. *Ocean Biogeochemical Dynamics*. (Princeton University Press, 2013). doi:10.1515/9781400849079.
 168. Hogg, N. G. On the transport of the gulf stream between cape hatteras and the grand banks. *Deep Sea Research Part A. Oceanographic Research Papers* **39**, 1231–1246 (1992).
 169. Hendry, R. M. On the structure of the deep Gulf Stream. *J Mar Res* (1982).

170. Rossby, T. The North Atlantic Current and surrounding waters: At the crossroads. *Reviews of Geophysics* **34**, (1996).
171. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Research Part I* (1995) doi:10.1016/0967-0637(95)00021-W.
172. Abraham, E. R. The generation of plankton patchiness by turbulent stirring. *Nature* **391**, (1998).
173. McGillicuddy, J. J., Anderson, L. A., Doney, S. C. & Maltrud, M. E. Eddy-driven sources and sinks of nutrients in the upper ocean: Results from a 0.1° resolution model of the North Atlantic. *Global Biogeochem Cycles* **17**, (2003).
174. Li, Q. P. & Hansell, D. A. Nutrient distributions in baroclinic eddies of the oligotrophic North Atlantic and inferred impacts on biology. *Deep Sea Res 2 Top Stud Oceanogr* **55**, (2008).
175. Biastoch, A., Böning, C. W. & Lutjeharms, J. R. E. Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation. *Nature* **456**, (2008).
176. Rühls, S., Zhurbas, V., Koszalka, I. M., Durgadoo, J. V. & Biastoch, A. Eddy diffusivity estimates from Lagrangian trajectories simulated with ocean models and surface drifter data-A case study for the greater Agulhas system. *J Phys Oceanogr* **48**, (2018).
177. Durand, M.-H. *et al.* *Global versus Local Changes in Upwelling Systems. Local And Global Change in Upwelling Systems* (1998).
178. Carl, W. What Is the Thermohaline Circulation? *Science* (1979) **298**, 1179–1181 (2002).
179. Wunsch, C. Oceanography: Moon, tides and climate. *Nature* **405**, 743+ (2000).
180. Johannes, L. & D., D. P. Risk of tipping the overturning circulation due to increasing rates of ice melt. *Proceedings of the National Academy of Sciences* **118**, e2017989118 (2021).
181. Kilbourne, K. H. *et al.* Atlantic circulation change still uncertain. *Nat Geosci* **15**, 165–167 (2022).
182. Thornalley, D. J. R. *et al.* Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years. *Nature* **556**, (2018).
183. Srokosz, M. A. & Bryden, H. L. Observing the Atlantic Meridional Overturning

- Circulation yields a decade of inevitable surprises. *Science* vol. 348 Preprint at <https://doi.org/10.1126/science.1255575> (2015).
184. Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N. & Rahmstorf, S. Current Atlantic Meridional Overturning Circulation weakest in last millennium. *Nat Geosci* **14**, (2021).
 185. Machado, J. P., Justino, F. & Pezzi, L. P. Impacts of Wind Stress Changes on the Global Heat Transport, Baroclinic Instability, and the Thermohaline Circulation. *Advances in Meteorology* **2016**, (2016).
 186. Zhu, J., Liu, Z., Zhang, X., Eisenman, I. & Liu, W. Linear weakening of the AMOC in response to receding glacial ice sheets in CCSM3. *Geophys Res Lett* **41**, (2014).
 187. McLean, D. Continuum fluid mechanics and the Navier-Stokes equations. *Aerospace Series* 13–77 (2012) doi:10.1002/9781118454190.ch3.
 188. Adcroft, A., Hill, C. & Marshall, A. J. Representation of topography by shaved cells in a height coordinate ocean model. *Mon Weather Rev* **125**, (1997).
 189. Reijnders, D., van Leeuwen, E. J. & van Sebille, E. Ocean Surface Connectivity in the Arctic: Capabilities and Caveats of Community Detection in Lagrangian Flow Networks. *J Geophys Res Oceans* **126**, (2021).
 190. Ser-Giacomi, E. *et al.* Lagrangian betweenness as a measure of bottlenecks in dynamical systems with oceanographic examples. *Nat Commun* **12**, (2021).
 191. Clayton, S. *et al.* Biogeochemical versus ecological consequences of modeled ocean physics. *Biogeosciences* **14**, (2017).
 192. Sommeria-Klein, G. *et al.* Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science (1979)* **374**, 594–599 (2021).
 193. Marañón, E., Lorenzo, M. P., Cermeño, P. & Mouriño-Carballido, B. Nutrient limitation suppresses the temperature dependence of phytoplankton metabolic rates. *ISME Journal* **12**, (2018).
 194. Eppley, Richard. Temperature and phytoplankton growth in the sea. *Fishery Bulletin* (1972).
 195. Arrhenius, S. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Zeitschrift für Physikalische Chemie* **4U**, 96–116 (1889).
 196. Delmont, T. O. *et al.* Single-amino acid variants reveal evolutionary processes

- that shape the biogeography of a global SAR11 subclade. *Elife* (2019) doi:10.7554/eLife.46497.
197. Leconte, J. *et al.* Equatorial to Polar genomic variability of the microalgae *Bathycoccus prasinos*; *bioRxiv* 2021.07.13.452163 (2021) doi:10.1101/2021.07.13.452163.
 198. Laso-Jadart, R. *et al.* Investigating population-scale allelic differential expression in wild populations of *Oithona similis* (Cyclopoida, Claus, 1866). *Ecol Evol* (2020) doi:10.1002/ece3.6588.
 199. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. In *Ecology* vol. 85 (2004).
 200. Droop, M. R. Vitamin B12 and Marine Ecology. IV. The Kinetics of Uptake, Growth and Inhibition in *Monochrysis Lutheri*. *Journal of the Marine Biological Association of the United Kingdom* **48**, (1968).
 201. Ward, B. A., Marañón, E., Sauterey, B., Rault, J. & Claessen, D. The Size Dependence of Phytoplankton Growth Rates: A Trade-Off between Nutrient Uptake and Metabolism. *Am Nat* **189**, 170–177 (2016).
 202. Monod, J. THE GROWTH OF BACTERIAL CULTURES. *Annu Rev Microbiol* **3**, 371–394 (1949).
 203. Sommer, U. A Comparison of the Droop and the Monod Models of Nutrient Limited Growth Applied to Natural Populations of Phytoplankton. *Funct Ecol* **5**, 535–544 (1991).
 204. Sakka, A. *et al.* Nitrate, phosphate, and iron limitation of the phytoplankton assemblage in the lagoon of Takapoto Atoll (Tuamotu Archipelago, French Polynesia). *Aquatic Microbial Ecology* **19**, (1999).
 205. Twining, B. S. *et al.* Taxonomic and nutrient controls on phytoplankton iron quotas in the ocean. *Limnology And Oceanography Letters* vol. 6 Preprint at <https://doi.org/10.1002/lol2.10179> (2021).
 206. McCain, J. S. P. *et al.* Cellular costs underpin micronutrient limitation in phytoplankton. *Sci Adv* **7**, (2021).
 207. Flynn, K. J. & Martin-Jézéquel, V. Modelling Si-N-limited growth of diatoms. *J Plankton Res* **22**, (2000).
 208. Sverdrup, H. U. On conditions for the vernal blooming of phytoplankton. *ICES Journal of Marine Science* **18**, (1953).

209. Huisman, J., Van Oostveen, P. & Weissing, F. J. Critical depth and critical turbulence: Two different mechanisms for the development of phytoplankton blooms. *Limnol Oceanogr* **44**, (1999).
210. Huisman, J., Arrayás, M., Ebert, U. & Sommeijer, B. How do sinking phytoplankton species manage to persist? *American Naturalist* **159**, (2002).
211. Behrenfeld, M. J., Doney, S. C., Lima, I., Boss, E. S. & Siegel, D. A. Annual cycles of ecological disturbance and recovery underlying the subarctic Atlantic spring plankton bloom. *Global Biogeochem Cycles* **27**, (2013).
212. Behrenfeld, M. J. & Boss, E. S. Resurrecting the ecological underpinnings of ocean plankton blooms. *Ann Rev Mar Sci* **6**, (2014).
213. Ward, B. A. & Waniek, J. J. Phytoplankton growth conditions during autumn and winter in the Irminger Sea, North Atlantic. *Mar Ecol Prog Ser* **334**, (2007).
214. Dale, T., Rey, F. & Heimdal, B. R. Seasonal development of phytoplankton at a high latitude oceanic site. *Sarsia* **84**, (1999).
215. Vardi, A. *et al.* Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science (1979)* **326**, (2009).
216. Bidle, K. D. & Vardi, A. A chemical arms race at sea mediates algal host-virus interactions. *Current Opinion in Microbiology* vol. 14 Preprint at <https://doi.org/10.1016/j.mib.2011.07.013> (2011).
217. Briggs, N. *et al.* High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom. *Deep Sea Research Part I: Oceanographic Research Papers* **58**, 1031–1039 (2011).
218. Nielsdóttir, M. C., Moore, C. M., Sanders, R., Hinz, D. J. & Achterberg, E. P. Iron limitation of the postbloom phytoplankton communities in the Iceland Basin. *Global Biogeochem Cycles* **23**, (2009).
219. Mahadevan, A., D'Asaro, E., Lee, C. & Perry, M. J. Eddy-driven stratification initiates North Atlantic spring phytoplankton blooms. *Science (1979)* **336**, (2012).
220. de Baar, H. *et al.* On iron limitation of the Southern Ocean: experimental observations in the Weddell and Scotia Seas. *Mar Ecol Prog Ser* **65**, (1990).
221. Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).

222. Cavender-Bares, K. K., Mann, E. L., Chisholm, S. W., Ondrusek, M. E. & Bidigare, R. R. Differential response of equatorial Pacific phytoplankton to iron fertilization. *Limnol Oceanogr* **44**, (1999).
223. Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, (2000).
224. Ustick, L. J. *et al.* Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. *Science (1979)* **372**, (2021).
225. Luo, Y.-W. *et al.* Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data* **4**, 47–73 (2012).
226. Fernández-Castro, B., Pahlow, M., Mouriño-Carballido, B., Marañón, E. & Oschlies, A. Optimality-based *Trichodesmium* diazotrophy in the North Atlantic subtropical gyre. *J Plankton Res* **38**, (2016).
227. Cavan, E. L., Trimmer, M., Shelley, F. & Sanders, R. Remineralization of particulate organic carbon in an ocean oxygen minimum zone. *Nat Commun* **8**, 14847 (2017).
228. Black, E. E., Buesseler, K. O., Pike, S. M. & Lam, P. J. ²³⁴Th as a tracer of particulate export and remineralization in the southeastern tropical Pacific. *Mar Chem* **201**, (2018).
229. Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem Cycles* (2013) doi:10.1002/gbc.20089.
230. Reygondeau, G. *et al.* Climate Change-Induced Emergence of Novel Biogeochemical Provinces. *Front Mar Sci* **7**, (2020).
231. Fay, A. R. & McKinley, G. A. Global open-ocean biomes: Mean and temporal variability. *Earth Syst Sci Data* (2014) doi:10.5194/essd-6-273-2014.
232. Oliver, M. J. & Irwin, A. J. Objective global ocean biogeographic provinces. *Geophys Res Lett* (2008) doi:10.1029/2008GL034238.
233. Beaugrand, G. *et al.* Prediction of unprecedented biological shifts in the global ocean. *Nature Climate Change* Preprint at <https://doi.org/10.1038/s41558-019-0420-1> (2019).
234. Beaugrand, G. Decadal changes in climate and ecosystems in the North Atlantic Ocean and adjacent seas. *Deep Sea Res 2 Top Stud Oceanogr* (2009) doi:10.1016/j.dsr2.2008.12.022.
235. Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C. E., Balch, W. M. & Guikema, S.

- D. Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO₂. *Science (1979)* (2015) doi:10.1126/science.aaa8026.
236. Buitenhuis, E. T. *et al.* MAREDAT: Towards a world atlas of MARine Ecosystem DATA. *Earth Syst Sci Data* **5**, (2013).
237. Friedlingstein, P. *et al.* Global Carbon Budget 2020. *Earth Syst Sci Data* **12**, (2020).
238. Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochem Cycles* (2012) doi:10.1029/2011GB004099.
239. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* (2016) doi:10.1038/nature16942.
240. Bopp, L., Aumont, O., Cadule, P., Alvain, S. & Gehlen, M. Response of diatoms distribution to global warming and potential implications: A global model study. *Geophys Res Lett* **32**, (2005).
241. Richardson, T. L. Mechanisms and pathways of small-phytoplankton export from the surface ocean. *Annual Review of Marine Science* vol. 11 Preprint at <https://doi.org/10.1146/annurev-marine-121916-063627> (2019).
242. Tréguer, P. *et al.* Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience* vol. 11 Preprint at <https://doi.org/10.1038/s41561-017-0028-x> (2018).
243. Richardson, A. J. In hot water: zooplankton and climate change. *ICES Journal of Marine Science* **65**, 279–295 (2008).
244. Halfter, S., Cavan, E. L., Swadling, K. M., Eriksen, R. S. & Boyd, P. W. The Role of Zooplankton in Establishing Carbon Export Regimes in the Southern Ocean – A Comparison of Two Representative Case Studies in the Subantarctic Region. *Frontiers in Marine Science* vol. 7 837 Preprint at (2020).
245. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science (1979)* (1998) doi:10.1126/science.281.5374.237.
246. Schlitzer, R. Export production in the equatorial and North Pacific derived from dissolved oxygen, nutrient and carbon data. *J Oceanogr* **60**, (2004).
247. Laws, E. A., Falkowski, P. G., Smith, W. O., Ducklow, H. & McCarthy, J. J. Temperature effects on export production in the open ocean. *Global Biogeochem Cycles* **14**, (2000).

248. Lutz, M. J., Caldeira, K., Dunbar, R. B. & Behrenfeld, M. J. Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean. *J Geophys Res Oceans* **112**, (2007).
249. Jahnke, R. A. The global ocean flux of particulate organic carbon: Areal distribution and magnitude. *Global Biogeochem Cycles* **10**, (1996).
250. Klaas, C. & Archer, D. E. Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochem Cycles* **16**, (2002).
251. Muller-Karger, F. E. *et al.* The importance of continental margins in the global carbon cycle. *Geophys Res Lett* **32**, (2005).
252. Smith, K. L., Ruhl, H. A., Huffard, C. L., Messié, M. & Kahru, M. Episodic organic carbon fluxes from surface ocean to abyssal depths during long-term monitoring in NE Pacific. *Proc Natl Acad Sci U S A* **115**, (2018).
253. Bopp, L. *et al.* Potential impact of climate change on marine export production. *Global Biogeochem Cycles* (2001) doi:10.1029/1999GB001256.
254. Steinacher, M. *et al.* Projected 21st century decrease in marine productivity: A multi-model analysis. *Biogeosciences* (2010) doi:10.5194/bg-7-979-2010.
255. Laufkötter, C. *et al.* Projected decreases in future marine export production: The role of the carbon flux through the upper ocean ecosystem. *Biogeosciences* **13**, (2016).
256. Hanrahan, G. & Chan, G. NITROGEN. in (eds. Worsfold, P., Townshend, A. & Poole, C. B. T.-E. of A. S. (Second E.) 191–196 (Elsevier, 2005). doi:<https://doi.org/10.1016/B0-12-369397-7/00401-5>.
257. Sorai, M., Yoshida, N. & Ishikawa, M. Biogeochemical simulation of nitrous oxide cycle based on the major nitrogen processes. *J Geophys Res Biogeosci* **112**, (2007).
258. Dugdale, R. C. & Goering, J. J. UPTAKE OF NEW AND REGENERATED FORMS OF NITROGEN IN PRIMARY PRODUCTIVITY. *Limnol Oceanogr* **12**, (1967).
259. Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, (1979).
260. Bronk, DA, Steinberg, DK. Chapter 8-Nitrogen regeneration. In *Nitrogen in the marine environment. 2nd Edn.* (ed. Capone DG, Bronk DA, Mulholland MR, Carpenter EJ, Editors.) 385–467 (San Diego: Academic Press, 2008).

261. Ploug, H. *et al.* Carbon, nitrogen and O₂ fluxes associated with the cyanobacterium *Nodularia spumigena* in the Baltic Sea. *ISME J* **5**, 1549–1558 (2011).
262. Ploug, H. *et al.* Carbon and nitrogen fluxes associated with the cyanobacterium *Aphanizomenon* sp. in the Baltic Sea. *ISME Journal* **4**, (2010).
263. Berthelot, H., Bonnet, S., Camps, M., Grosso, O. & Moutin, T. Assessment of the dinitrogen released as ammonium and dissolved organic nitrogen by unicellular and filamentous marine diazotrophic cyanobacteria grown in culture. *Front Mar Sci* **2**, (2015).
264. Barnard, R., Leadley, P. W. & Hungate, B. A. Global change, nitrification, and denitrification: A review. *Global Biogeochemical Cycles* vol. 19 Preprint at <https://doi.org/10.1029/2004GB002282> (2005).
265. Dore, J. E. & Karl, D. M. Nitrification in the euphotic zone as a source for nitrite, nitrate, and nitrous oxide at Station ALOHA. *Limnol Oceanogr* **41**, (1996).
266. Martin, A. P. & Pondaven, P. New primary production and nitrification in the western subtropical North Atlantic: A modeling study. *Global Biogeochem Cycles* **20**, (2006).
267. Zhang, X., Ward, B. B. & Sigman, D. M. Global Nitrogen Cycle: Critical Enzymes, Organisms, and Processes for Nitrogen Budgets and Dynamics. *Chem Rev* **120**, 5308–5351 (2020).
268. Beman, J. M. *et al.* Global declines in oceanic nitrification rates as a consequence of ocean acidification. *Proc Natl Acad Sci U S A* **108**, (2011).
269. Buchanan, P. J., Aumont, O., Bopp, L., Mahaffey, C. & Tagliabue, A. Impact of intensifying nitrogen limitation on ocean net primary production is fingerprinted by nitrogen isotopes. *Nat Commun* **12**, (2021).
270. Laws, E. A., D'Sa, E. & Naik, P. Simple equations to estimate ratios of new or export production to total production from satellite-derived estimates of sea surface temperature and primary production. *Limnol Oceanogr Methods* **9**, (2011).
271. Wrightson, L. & Tagliabue, A. Quantifying the Impact of Climate Change on Marine Diazotrophy: Insights From Earth System Models. *Frontiers in Marine Science* vol. 7, Preprint at <https://www.frontiersin.org/articles/10.3389/fmars.2020.00635/full> 635 (2020).
272. Chakraborty, S. *et al.* Quantifying nitrogen fixation by heterotrophic bacteria in sinking marine particles. *Nat Commun* **12**, (2021).

273. Redfield, A. C. On the Proportions of Organic Derivatives in Sea Water and Their Relation to the Composition of Plankton. In *James Johnstone Memorial Volume 1767–192* (Liverpool Univ. Press, Liverpool, U.K., 1934).
274. Geider, R. J. & La Roche, J. Redfield revisited: Variability of C:N:P in marine microalgae and its biochemical basis. *European Journal of Phycology* vol. 37 Preprint at <https://doi.org/10.1017/S0967026201003456> (2002).
275. Bunge, M. *Causality and Modern Science*. (2009). doi:<https://doi.org/10.4324/9781315081656>.
276. Zhang, D. D. *et al.* The causality analysis of climate change and large-scale human crisis. *Proc Natl Acad Sci U S A* **108**, (2011).
277. The Nobel Prize in Physics 2021. NobelPrize.org. Nobel Prize Outreach AB 2022. <https://www.nobelprize.org/prizes/physics/2021/summary/>.
278. Marti, O. *et al.* Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. *Clim Dyn* **34**, (2010).
279. Timmermann, R. *et al.* On the representation of high latitude processes in the ORCA-LIM global coupled sea ice-ocean model. *Ocean Model (Oxf)* **8**, (2005).
280. Dutkiewicz, S. *et al.* Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences* **12**, (2015).
281. Tagliabue, A. & Völker, C. Towards accounting for dissolved iron speciation in global ocean models. *Biogeosciences* **8**, (2011).
282. Steffen, W. *et al.* The emergence and evolution of Earth System Science. *Nature Reviews Earth and Environment* vol. 1 Preprint at <https://doi.org/10.1038/s43017-019-0005-6> (2020).
283. Tagliabue, A. *et al.* How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Global Biogeochem Cycles* **30**, (2016).
284. Braun, R. & Burnham, A. Chemical reaction model for oil and gas generation from type 1 and type 2 kerogen. (1993) doi:10.2172/10169154.
285. Kvenvolden, K. A. Organic geochemistry - A retrospective of its first 70 years. In *Organic Geochemistry* vol. 37 (2006).
286. Schobert, H. *Chemistry of fossil fuels and biofuels. Chemistry of Fossil Fuels and Biofuels* vol. 9780521114004 (2010).
287. Peak oil: Does it really matter? Oil & Gas Middle East. <https://www.oilandgasmiddleeast.com/news/article-8400-peak-oil-does-it->

- really-matter.
288. Energy Alternatives and the Future of Oil and Gas in the Gulf. Al Jazeera Center for Studies. <https://studies.aljazeera.net/en/dossiers/2015/03/201533183514675179.html>.
 289. How long will world's oil reserves last? 53 years, says BP. Christian Science Monitor. <https://www.csmonitor.com/Environment/Energy-Voices/2014/0714/How-long-will-world-s-oil-reserves-last-53-years-says-BP>.
 290. Masson-Delmotte, V., P. *et al.* *The Physical Science Basis. Contribution of Working Group I to the 552 Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change*, vol. 2021 (2021).
 291. Hauck, J. *et al.* Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget. *Front Mar Sci* **7**, (2020).
 292. Foote, E. Circumstances Affecting the Heat of Sun's Rays. *American Journal of Art and Science* **LXVI**, 382–383 (1856).
 293. Weekly Evening Meeting, Friday, June 10, 1859. The Prince Consort, Vice-Patron, in the Chair. John Tyndall, Esq. F.R.S. 'On the Transmission of Heat of different qualities through Gases of different kinds'. In *Notices of the Proceedings at the Meetings of the Members of the Royal Institution of Great Britain: With Abstracts of the Discourses Delivered at the Evening Meetings* 155–158 (Royal Institution of Great Britain, 1862).
 294. Arrhenius, S. XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **41**, 237–276 (1896).
 295. Garcia-Soto, C. *et al.* An Overview of Ocean Climate Change Indicators: Sea Surface Temperature, Ocean Heat Content, Ocean pH, Dissolved Oxygen Concentration, Arctic Sea Ice Extent, Thickness and Volume, Sea Level and Strength of the AMOC (Atlantic Meridional Overturning Circulation). *Frontiers in Marine Science* vol. 8 Preprint at <https://doi.org/10.3389/fmars.2021.642372> (2021).
 296. Figuerola, B. *et al.* A Review and Meta-Analysis of Potential Impacts of Ocean Acidification on Marine Calcifiers From the Southern Ocean. *Frontiers in Marine Science* vol. 8 Preprint at <https://doi.org/10.3389/fmars.2021.584445> (2021).
 297. Lachkar, Z., Lévy, M. & Smith, S. Intensification and deepening of the Arabian

- Sea oxygen minimum zone in response to increase in Indian monsoon wind intensity. *Biogeosciences* **15**, (2018).
298. Frölicher, T. L., Fischer, E. M. & Gruber, N. Marine heatwaves under global warming. *Nature* **560**, (2018).
 299. Williams, L. & Author, C. Impact of climate change on fisheries and aquaculture in the developing world and opportunities for adaptation Key words. *Fisheries Thematic paper* (2011).
 300. Stenevik, E. K. & Sundby, S. Impacts of climate change on commercial fish stocks in Norwegian waters. *Mar Policy* **31**, (2007).
 301. <https://www.worldwildlife.org/industries/sustainable-seafood#:~:text=Approximately%203%20billion%20people%20in,to%20billions%20of%20people%20worldwide>.
 302. M., R. C. *et al.* Marine reserves can mitigate and promote adaptation to climate change. *Proceedings of the National Academy of Sciences* **114**, 6167–6175 (2017).
 303. Creighton, C., Hobday, A. J., Lockwood, M. & Pecl, G. T. Adapting Management of Marine Environments to a Changing Climate: A Checklist to Guide Reform and Assess Progress. *Ecosystems* **19**, (2016).
 304. Blythe, J. L., Armitage, D., Bennett, N. J., Silver, J. J. & Song, A. M. The Politics of Ocean Governance Transformations. *Front Mar Sci* **8**, (2021).
 305. Hartmann, D. L. *et al.* Observations: Atmosphere and surface. In *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (2013). doi:10.1017/CBO9781107415324.008.
 306. <https://www.eea.europa.eu/data-and-maps/indicators/atmospheric-greenhouse-gas-concentrations/atmospheric-greenhouse-gas-concentrations-assessment-1#:~:text=The%20present%20concentrations%20of%20CO2,the%20past%201%20000%20years>.
 307. <https://www.eea.europa.eu/ims/atmospheric-greenhouse-gas-concentrations>.
 308. Bopp, L. *et al.* Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences* (2013) doi:10.5194/bg-10-6225-2013.
 309. Xu, C., Kohler, T. A., Lenton, T. M., Svenning, J. C. & Scheffer, M. Future of the

- human climate niche. *Proc Natl Acad Sci U S A* **117**, (2020).
310. Burrows, M. T. *et al.* Geographical limits to species-range shifts are suggested by climate velocity. *Nature* **507**, 492–495 (2014).
 311. Irwin, A. J., Finkel, Z. v., Müller-Karger, F. E. & Troccoli Ghinaglia, L. Phytoplankton adapt to changing ocean environments. *Proceedings of the National Academy of Sciences* (2015) doi:10.1073/pnas.1414752112.
 312. Mishra, S. K. Acclimatization BT - Encyclopedia of Animal Cognition and Behavior. in (eds. Vonk, J. & Shackelford, T.) 1–2 (Springer International Publishing, 2019). doi:10.1007/978-3-319-47829-6_1035-1.
 313. Darwin, C. The Origin of Species, by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life. *The Crayon* **7**, (1860).
 314. Mendel, G. EXPERIMENTS IN PLANT HYBRIDIZATION (1865). *Verhandlungen des naturforschenden Ver-eines in Brünn* **4**, (1866).
 315. Morgan, T. H. *Heredity and sex / by Thomas Hunt Morgan. Heredity and sex / by Thomas Hunt Morgan.* (2011). doi:10.5962/bhl.title.6238.
 316. Morgan, T. H. *The physical basis of heredity / by Thomas Hunt Morgan. The physical basis of heredity / by Thomas Hunt Morgan.* (2011). doi:10.5962/bhl.title.46225.
 317. Huxley, J. & others. *Evolution. The modern synthesis. Evolution. The Modern Synthesis.* (London: George Alien & Unwin Ltd., 1942).
 318. Frías L, D. Omissions in the synthetic theory of evolution. *Biol Res* **43**, (2010).
 319. Vaidyanathan, G. The world's species are playing musical chairs: how will it end? *Nature* vol. 596 Preprint at <https://doi.org/10.1038/d41586-021-02088-3> (2021).
 320. Cowie, R. H., Bouchet, P. & Fontaine, B. The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews* (2022).
 321. Frémont, P. *et al.* Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat Clim Chang* (2022) doi:10.1038/s41558-022-01314-8.
 322. Demmig-Adams, B., Dumlao, M. R., Herzenach, M. K. & Adams, W. W. Acclimation. In *Encyclopedia of Ecology* (eds. Jørgensen, S. E. & Fath, B. D.) 15–23 (Academic Press, 2008). doi:<https://doi.org/10.1016/B978-008045405-4.00001-X>.

323. Hegde, P. S., White, I. R. & Debouck, C. Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol* **14**, 647–651 (2003).
324. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
325. Wassmann, P. *et al.* The contiguous domains of Arctic Ocean advection: Trails of life and death. *Prog Oceanogr* **139**, (2015).
326. Polyakov, I. v. *et al.* Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. *Science (1979)* **356**, (2017).
327. Polyakov, I. v. *et al.* Arctic ocean warming contributes to reduced polar ice cap. *J Phys Oceanogr* **40**, (2010).
328. Polyakov, I. v. *et al.* Weakening of cold halocline layer exposes sea ice to oceanic heat in the eastern arctic ocean. *J Clim* **33**, (2020).
329. Neukermans, G., Oziel, L. & Babin, M. Increased intrusion of warming Atlantic water leads to rapid expansion of temperate phytoplankton in the Arctic. *Glob Chang Biol* **24**, (2018).
330. Oziel, L. *et al.* Faster Atlantic currents drive poleward expansion of temperate phytoplankton in the Arctic Ocean. *Nat Commun* **11**, (2020).
331. Duarte, C. M. *et al.* Tipping elements in the Arctic marine ecosystem. *Ambio* **41**, (2012).
332. Wang, Y. *et al.* Identification of tRNA nucleoside modification genes critical for stress response and development in rice and Arabidopsis. *BMC Plant Biol* **17**, (2017).
333. VanBogelen, R. A. & Neidhardt, F. C. Ribosomes as sensors of heat and cold shock in *Escherichia coli*. *Proc Natl Acad Sci U S A* **87**, (1990).
334. Bayles, D. O., Tunick, M. H., Foglia, T. A. & Miller, A. J. Cold shock and its effect on ribosomes and thermal tolerance in *Listeria monocytogenes*. *Appl Environ Microbiol* **66**, (2000).
335. Huby, E., Napier, J. A., Baillieul, F., Michaelson, L. v. & Dhondt-Cordelier, S. Sphingolipids: towards an integrated view of metabolism during the plant stress response. *New Phytologist* vol. 225 Preprint at <https://doi.org/10.1111/nph.15997> (2020).
336. Fabri, J. H. T. M., de Sá, N. P., Malavazi, I. & del Poeta, M. The dynamics and role of sphingolipids in eukaryotic organisms upon thermal adaptation. *Progress in*

- Lipid Research* vol. 80 Preprint at <https://doi.org/10.1016/j.plipres.2020.101063> (2020).
337. Kong, F., Romero, I. T., Warakanont, J. & Li-Beisson, Y. Lipid catabolism in microalgae. *New Phytologist* **218**, (2018).
 338. Davies, C. H. *et al.* A database of marine phytoplankton abundance, biomass and species composition in Australian waters. *Sci Data* **3**, 160043 (2016).
 339. Hallegraeff, G. *et al.* *Algae of Australia: Phytoplankton of temperate coastal waters*. (2010).
 340. Beckmann, A., Schaum, C. E. & Hense, I. Phytoplankton adaptation in ecosystem models. *J Theor Biol* **468**, (2019).
 341. Ward, B. A., Dutkiewicz, S., Jahn, O. & Follows, M. J. A size-structured food-web model for the global ocean. *Limnol Oceanogr* **57**, 1877–1891 (2012).
 342. Hairston, N. G., Ellner, S. P., Geber, M. A., Yoshida, T. & Fox, J. A. Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters* vol. 8 Preprint at <https://doi.org/10.1111/j.1461-0248.2005.00812.x> (2005).
 343. Jin, P. & Agustí, S. Fast adaptation of tropical diatoms to increased warming with trade-offs. *Sci Rep* **8**, (2018).
 344. Schiffers, K., Bourne, E. C., Lavergne, S., Thuiller, W. & Travis, J. M. J. Limited evolutionary rescue of locally adapted populations facing climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, (2013).
 345. Listmann, L., LeRoch, M., Schlüter, L., Thomas, M. K. & Reusch, T. B. H. Swift thermal reaction norm evolution in a key marine phytoplankton species. *Evol Appl* **9**, (2016).
 346. Bell, G. Evolutionary rescue and the limits of adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 368 Preprint at <https://doi.org/10.1098/rstb.2012.0080> (2013).
 347. Finlay, B. J. & Clarke, K. J. Ubiquitous dispersal of microbial species. *Nature* **400**, 828 (1999).
 348. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
 349. Polovina, J. J., Dunne, J. P., Woodworth, P. A. & Howell, E. A. Projected expansion of the subtropical biome and contraction of the temperate and

- equatorial upwelling biomes in the North Pacific under global warming. *ICES Journal of Marine Science* (2011) doi:10.1093/icesjms/fsq198.
350. Laufkotter, C. *et al.* Drivers and uncertainties of future global marine primary production in marine ecosystem models. *Biogeosciences* (2015) doi:10.5194/bg-12-6955-2015.
 351. Armstrong McKay, D. I. *et al.* Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science* (1979) **377**, eabn7950 (2022).
 352. Zhang, C. & Jia, G. Reversible RNA Modification N1-methyladenosine (m1A) in mRNA and tRNA. *Genomics, Proteomics and Bioinformatics* vol. 16 Preprint at <https://doi.org/10.1016/j.gpb.2018.03.003> (2018).
 353. Valadon, C. & Namy, O. The importance of the Epi-transcriptome in translation fidelity. *Non-coding RNA* vol. 7 Preprint at <https://doi.org/10.3390/ncrna7030051> (2021).
 354. Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A. & Edwards, M. Reorganization of North Atlantic marine copepod biodiversity and climate. *Science* (1979) (2002) doi:10.1126/science.1071329.
 355. Rahmstorf, S. *et al.* Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat Clim Chang* (2015) doi:10.1038/nclimate2554.
 356. Hutchinson, G. E. The Paradox of the Plankton. *Am Nat* **95**, (1961).
 357. Roy, S. & Chattopadhyay, J. Towards a resolution of 'the paradox of the plankton': A brief overview of the proposed mechanisms. *Ecological Complexity* vol. 4 Preprint at <https://doi.org/10.1016/j.ecocom.2007.02.016> (2007).
 358. Weitz, J. S. & Wilhelm, S. W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* **4**, (2012).
 359. Demory, D. *et al.* Linking Light-Dependent Life History Traits with Population Dynamics for Prochlorococcus and Cyanophage. *mSystems* **5**, (2020).
 360. Demory, D. *et al.* A thermal trade-off between viral production and degradation drives virus-phytoplankton population dynamics. *Ecol Lett* **24**, (2021).
 361. Menge, D. N. L. & Weitz, J. S. Dangerous nutrients: evolution of phytoplankton resource uptake subject to virus attack. *J Theor Biol* **257**, 104–115 (2009).
 362. Dufresne, J. L. *et al.* Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Clim Dyn* (2013) doi:10.1007/s00382-

012-1636-1.

363. Hellweger, F. L. *et al.* The role of ocean currents in the temperature selection of plankton: Insights from an individual-based model. *PLoS One* **11**, (2016).




ANNEXE

I. Contributions diverses

1. Article 1: Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean.

Pour cet article, j'ai fourni mes codes de modélisation par machine learning à Greta Busseni alors à la Stazione Zoologica Anton Dohrn, l'auteure principale de l'article. J'ai aussi fourni les données de climatologies (utilisées dans le Chapitre I). Nous avons discuté ensemble de l'adaptation des codes à l'étude et elle les a modifiés.

Large scale patterns of marine diatom richness: Drivers and trends in a changing ocean

Greta Busseni¹  | Luigi Caputi¹  | Roberta Piredda¹ | Paul Fremont^{1,2} |
Bruno Hay Mele¹  | Lucia Campese¹ | Eleonora Scalco¹ | Colombar de Vargas^{3,4} |
Chris Bowler⁵ | Francesco d'Ovidio⁶ | Adriana Zingone¹ | Maurizio Ribera d'Alcalà¹ |
Daniele Iudicone¹

¹Stazione Zoologica Anton Dohrn, Villa Comunale, Naples, Italy

²Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France

³Sorbonne Université, CNRS, Station Biologique de Roscoff, Roscoff, France

⁴Research Federation for the study of Global Ocean Systems Ecology and Evolution, Tara Oceans GOSEE, Paris, France

⁵Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Université Paris, Paris, France

⁶Sorbonne Université, CNRS, Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN-IPSL), Paris, France

Correspondence

Daniele Iudicone, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy.
Email: daniele.iudicone@szn.it

Editor: Derek Tittensor

Abstract

Aim: Plankton diversity is a pivotal element of marine ecosystem stability and functioning. A major obstacle in the assessment of diversity is the lack of consistency between patterns assessed by molecular and morphological data. This work aims to reconcile the two in a single richness measure, to investigate the environmental drivers affecting this measure, and finally to predict its spatio-temporal patterns.

Location and time period: This is a global scale study, based on data collected within the 2009–2013 interval during the *Tara* Oceans expedition.

Major taxa studied: The focus of this study is diatoms. They play an important role in several biogeochemical cycles and within marine food webs, and display high taxonomic and functional richness.

Methods: We integrate measures of diatom richness across the global ocean using molecular and morphological approaches, giving particular attention to 'the rare biosphere'. We then perform a machine-learning-based analysis of these reconciled patterns to extrapolate diatom richness at the global scale and to identify the main environmental processes governing it. Finally, we model the response of diatom richness to climate change.

Results: By filtering out 0.3% of the rarest operational taxonomic units, molecular-based richness patterns show the best possible match with the morphological approach. Temperature, phosphate, chlorophyll *a* and the Lyapunov exponent are the major explainers of these reconciled patterns. Global scale predictions provide a first approximation of the global geography of diatom richness and of the possible impacts of climate change.

Main conclusions: Our models suggest that diatom richness is controlled by different processes characteristic of distinct environmental scenarios: lateral mixing in highly dynamic regions, and both nutrient availability and temperature elsewhere. We present herein the effects of these processes on richness and how these same effects












This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Global Ecology and Biogeography published by John Wiley & Sons Ltd

2. Article 2: Genomic adaptation of the picoeukaryote *Pelagomonas calceolata* to iron-poor oceans revealed by a chromosome-scale genome sequence.

Pour cette étude j'ai collaboré avec deux de mes collègues du Laboratoire d'Analyse Génomique des Eucaryotes (LAGE) du Genoscope (le laboratoire où j'ai réalisé ma thèse), Nina Guérin et Quentin Carradec. J'ai adapté mes codes de machine learning pour l'étude et réalisé l'ensemble des figures et de la partie sur la modélisation de l'abondance relative de *Pelagomonas calceolata* dans l'océan global, en discussion avec Quentin et Nina.

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

Nina Guérin^{1,2}, Marta Ciccarella ¹, Elisa Flamant^{1,2}, Paul Frémont ^{1,2}, Sophie Mangenot^{1,2}, Benjamin Istace¹, Benjamin Noel ¹, Caroline Belser ¹, Laurie Bertrand^{1,2}, Karine Labadie ^{2,3}, Corinne Cruaud^{2,3}, Sarah Romac ⁴, Charles Bachy ^{4,5}, Martin Gachenot⁵, Eric Pelletier^{1,2}, Adriana Alberti^{1,2,6}, Olivier Jaillon ^{1,2}, Patrick Wincker ^{1,2}, Jean-Marc Aury ¹ & Quentin Carradec ^{1,2} ✉

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

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

II. Suppléments

1. Figures de l'article "Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems"

J'ai réalisé les figures suivantes: Figure 1-figure supplement 2, Figure 1-figure supplement 4 (g-h-i), Figure 2a, Figure 2-figure supplement 1, Figure 3, Figure 3-figure supplement 1.



Figures and figure supplements

Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems

Daniel J Richter et al

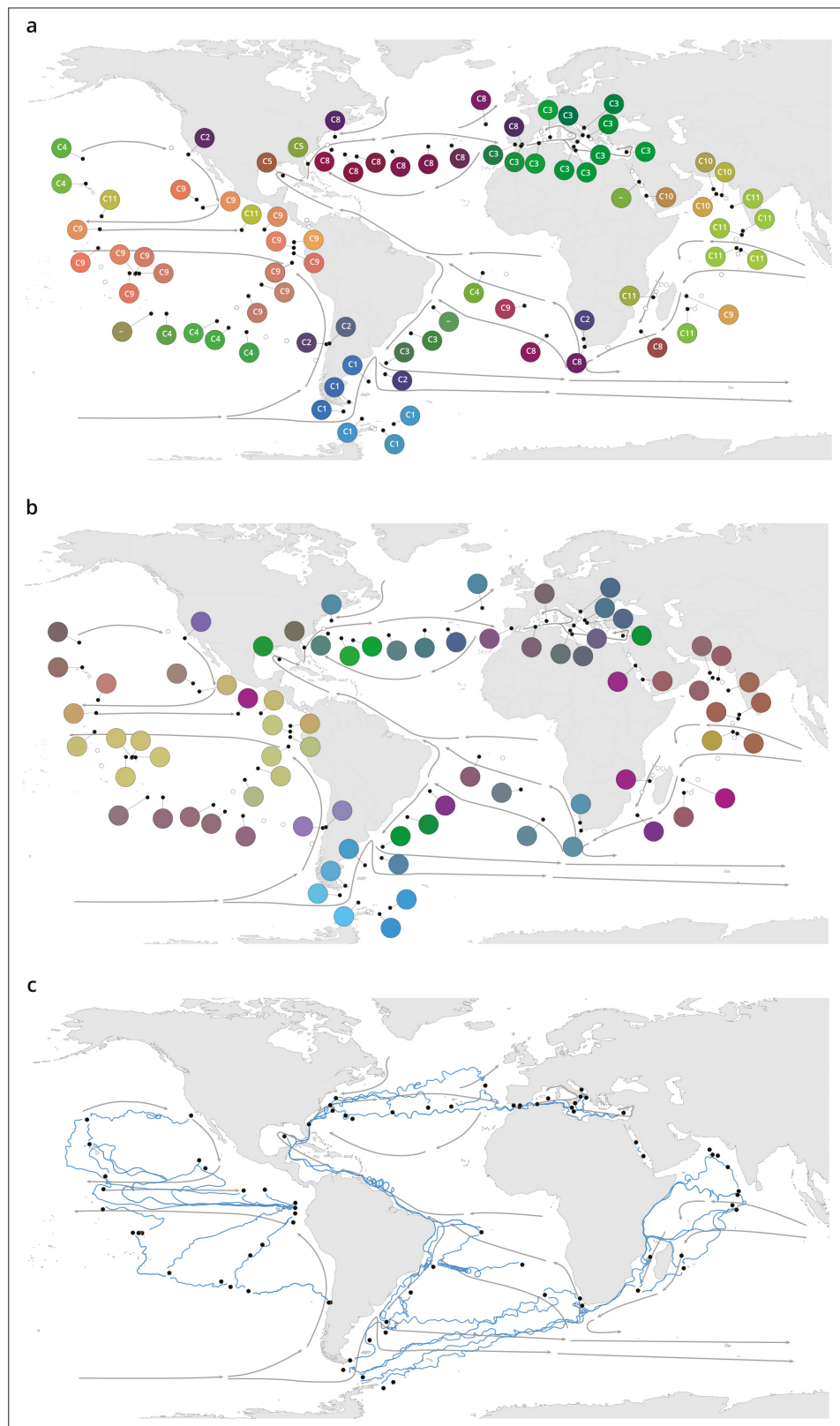


Figure 1. Plankton biogeography, environmental variation, and ocean transport among *Tara* Oceans stations. Major currents are represented by solid arrows. **(a)** Genomic provinces of *Tara* Oceans surface samples for the 0.8–5 μm size fraction, each labeled with a letter prefix ('C' represents the 0.8–5 μm size fraction) and a number; samples not assigned to a genomic province are labeled with '-'. Maps of all six size fractions and including deep
Figure 1 continued on next page

Figure 1 continued

chlorophyll maximum samples are available in **Figure 1—figure supplement 4**. Station colors are derived from an ordination of metagenomic dissimilarities; more dissimilar colors indicate more dissimilar communities (see Materials and methods). **(b)** Stations colored based on an ordination of temperature and the ratio of $\text{NO}_3 + \text{NO}_2$ to PO_4 (replaced by 10^{-6} for three stations where the measurement of PO_4 was 0) and of $\text{NO}_3 + \text{NO}_2$ to Fe. Colors do not correspond directly between maps; however, the geographical partitioning among stations is similar between the two maps. **(c)** Simulated trajectories corresponding to the minimum travel time (T_{\min}) for pairs of stations (black dots) connected by $T_{\min} < 1.5$ years. Directionality of trajectories is not represented.

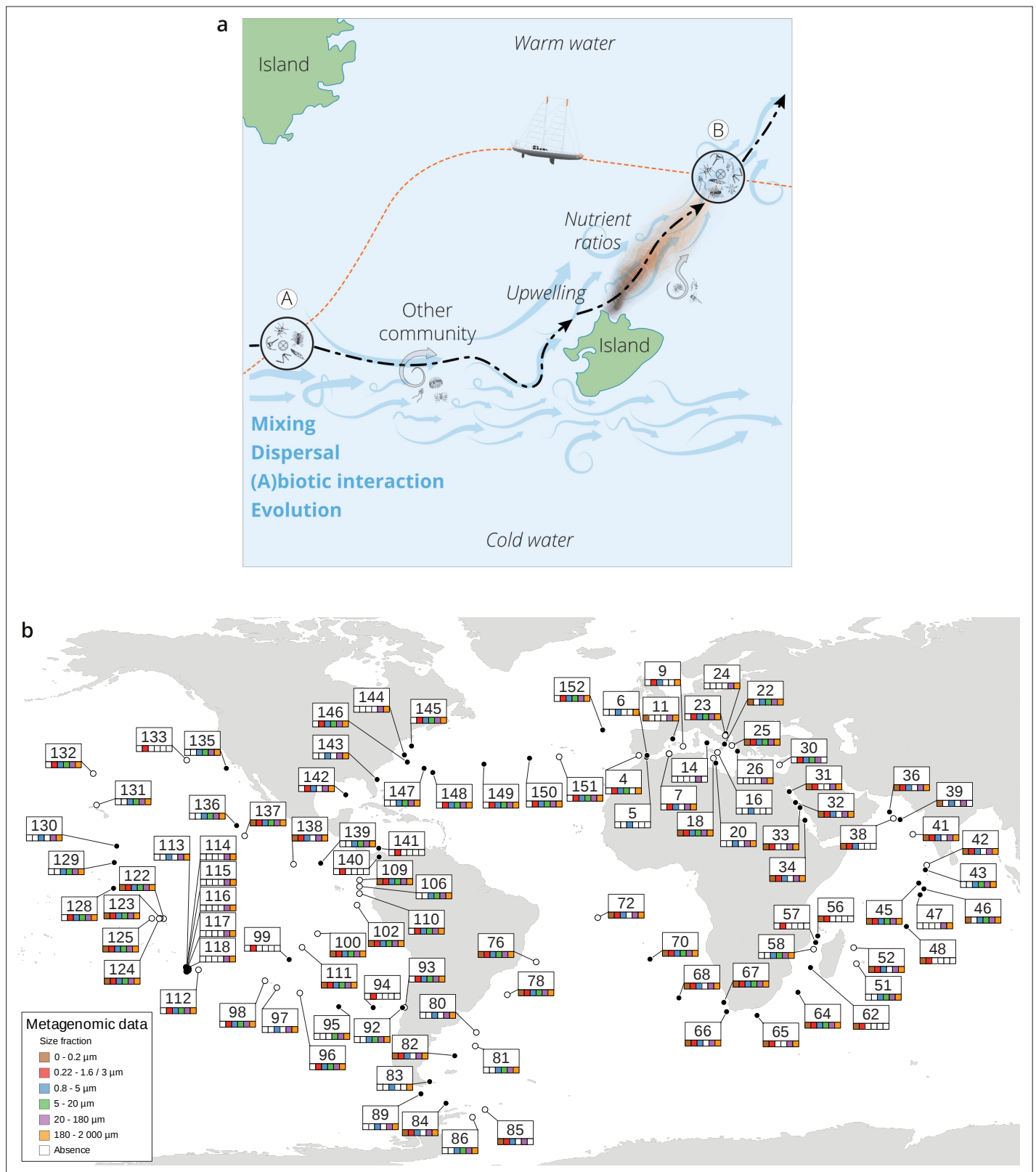


Figure 1—figure supplement 1. The seascape, plankton transport, and community metagenomic samples of Tara Oceans stations. (a) A community sampled at a given location A changes over time as it travels along ocean currents (dashed bold line) to a second location B. It is affected by numerous external processes, including mixing with water containing other communities and changes in local nutrient concentration, and by internal processes, such as biotic interactions. In this study, the Tara schooner followed a sampling route (orange dashed line) leading to an elapsed time between the two Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

sampling sites A and B that was independent of plankton travel time. **(b)** Location, station number, and sequenced surface metagenomic samples. For the stations indicated with a hollow circle, there was a greater than 0.5°C difference in temperature between surface and deep chlorophyll maximum (DCM); those indicated with a full circle had a difference in temperature less than or equal to 0.5°C or did not have samples sequenced for both the surface and DCM.

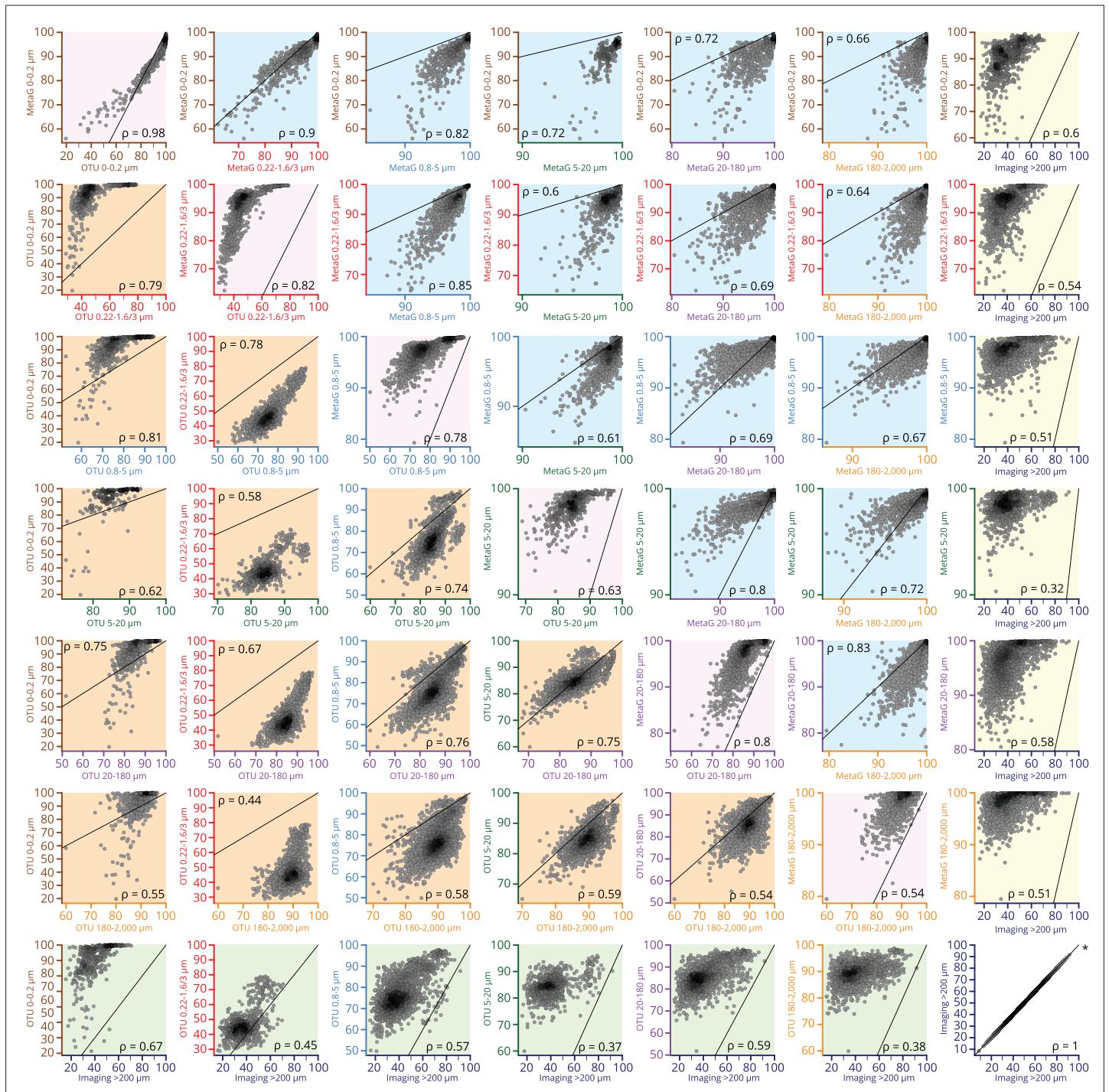


Figure 1—figure supplement 2. Scatter plots comparing β -diversity estimates from metagenomic, operational taxonomic unit (OTU)-based, and imaging-based dissimilarity. Source data for comparisons are indicated on the axes of each plot (axis colors correspond to size fractions or imaging data as in other figures, e.g., **Figure 3—figure supplement 1**). Axes are not necessarily drawn on the same scales; the identity line is indicated on each plot to help interpret the relationship between axes. Plots with a pink background are comparisons of metagenomic versus OTU-based dissimilarity within the same size fraction. Plots with a blue background are comparisons of metagenomic dissimilarity among size fractions, and those with an orange background compare OTU-based dissimilarity among size fractions. Plots with a yellow or green background compare imaging-based dissimilarity to either metagenomic or OTU-based dissimilarity, respectively. Each point within a plot represents a pairwise comparison of β -diversity estimates between two Tara Oceans samples. Rank-based correlations (Spearman, $p \leq 10^{-4}$) are indicated in each plot.

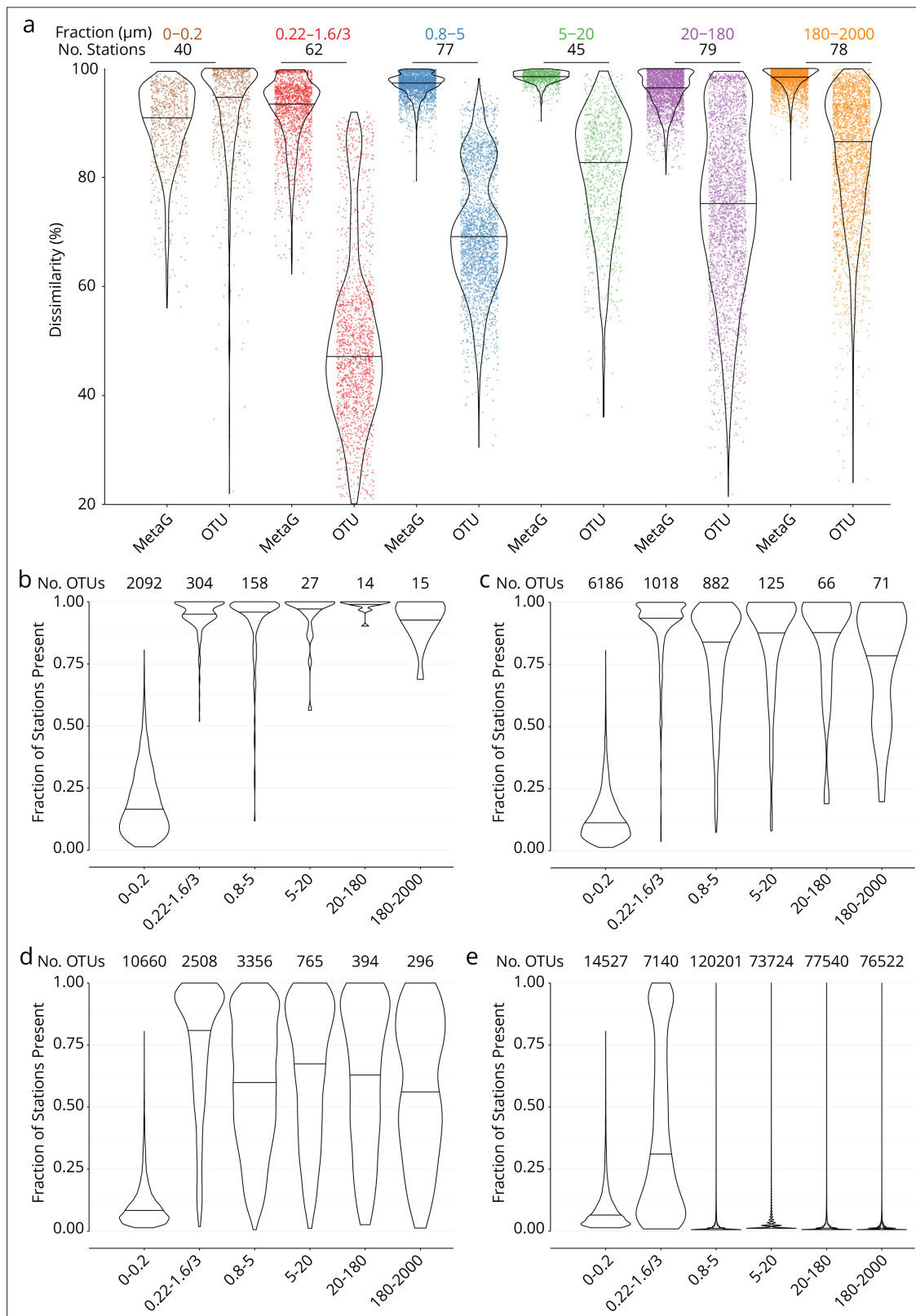


Figure 1—figure supplement 3. Global dissimilarity and operational taxonomic unit (OTU) occupancy. **(a)** Distributions of dissimilarity for six organismal size fractions (measured either as metagenomic or OTU dissimilarity; see Appendix 1). One colored point represents one pair of stations. Violin plots (horizontal line: median) summarize each distribution. The number of stations in common between the metagenomic/OTU data sets within each size fraction is indicated above. **(b–e)** OTU occupancy for different proportions of total abundance. Fraction of stations present (occupancy) for the

Figure 1—figure supplement 3 continued on next page

Figure 1—figure supplement 3 continued

minimum number of OTUs (indicated above) necessary to represent different proportions of the total abundance within each organismal size fraction. A relatively small number of abundant and cosmopolitan taxa represents the majority of the abundance within each size fraction; this effect is more pronounced with increasing organismal size. **(b)** OTUs representing 50% of the total abundance within each size fraction. **(c)** 80%. **(d)** 95%. **(e)** 100% (all OTUs).

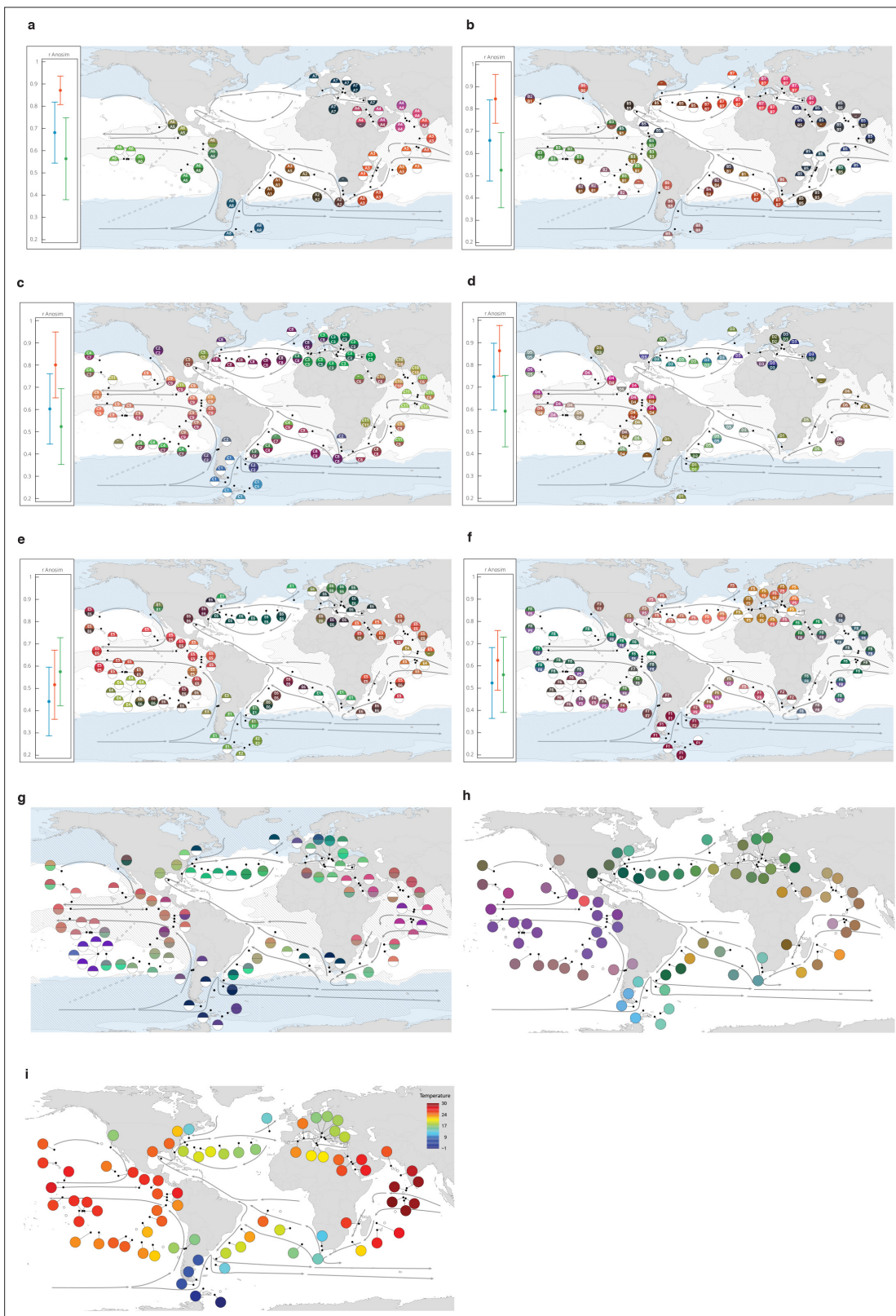


Figure 1—figure supplement 4. Genomic provinces in comparison to previous ocean divisions and to metagenome-assembled genome (MAG) abundance variation, and ordination maps of environmental parameters. Colors are based on principal coordinates analysis-RGB (Methods) and do not correspond directly among maps. (a–f) Geographical maps of genomic provinces by organismal size fraction (see Appendix 2). Circles denote stations with data available for the size fraction and contain the corresponding genomic province identifiers (one letter prefix per size fraction [A–F]); stations
 Figure 1—figure supplement 4 continued on next page

Figure 1—figure supplement 4 continued

not assigned to genomic provinces are shown as '-'). The top portion of each circle represents samples collected at the surface and the bottom portion represents the deep chlorophyll maximum (stations missing metagenomic data for one of the two depths are drawn as half circles). Major currents are shown with solid black arrows, wind transport with dashed gray arrows. Blue zones indicate temperature <14°C. Hashed zones indicate phosphate concentration >0.4 mmol. Hierarchical dendrograms that were used to build genomic provinces are shown in **Figure 1—figure supplement 6**. Maps with colors based on operational taxonomic unit dissimilarity are shown in **Figure 1—figure supplement 5**. (a) 'A' prefix, 0–0.2 µm size fraction. (b) 'B' prefix, 0.22–1.6/3 µm. (c) 'C' prefix, 0.8–5 µm. (d) 'D' prefix, 5–20 µm. (e) 'E' prefix, 20–180 µm. (f) 'F' prefix, 180–2000 µm. **Insets**, Results of ANOSIM to determine, independently for each size fraction, the ability of three nested levels of ocean partitioning to explain metagenomic dissimilarities among stations (blue, Longhurst biomes; red, Longhurst biogeochemical provinces; green, Oliver and Irwin objective provinces; see Materials and methods and Appendix 3). (g) Geographical map for the 20–180 µm size fraction, for comparison with panel (e) generated from MAG dissimilarity among stations. (h) The distribution of temperature and nutrient variations matches the biogeography of small plankton (<20 µm). Stations are colored based on an ordination of Euclidean distances in temperature, NO₃ + NO₂, PO₄, and Fe. (i) The distribution of temperature matches the biogeography of large plankton (>20 µm). Stations are colored following a Box-Cox transformation (Methods).

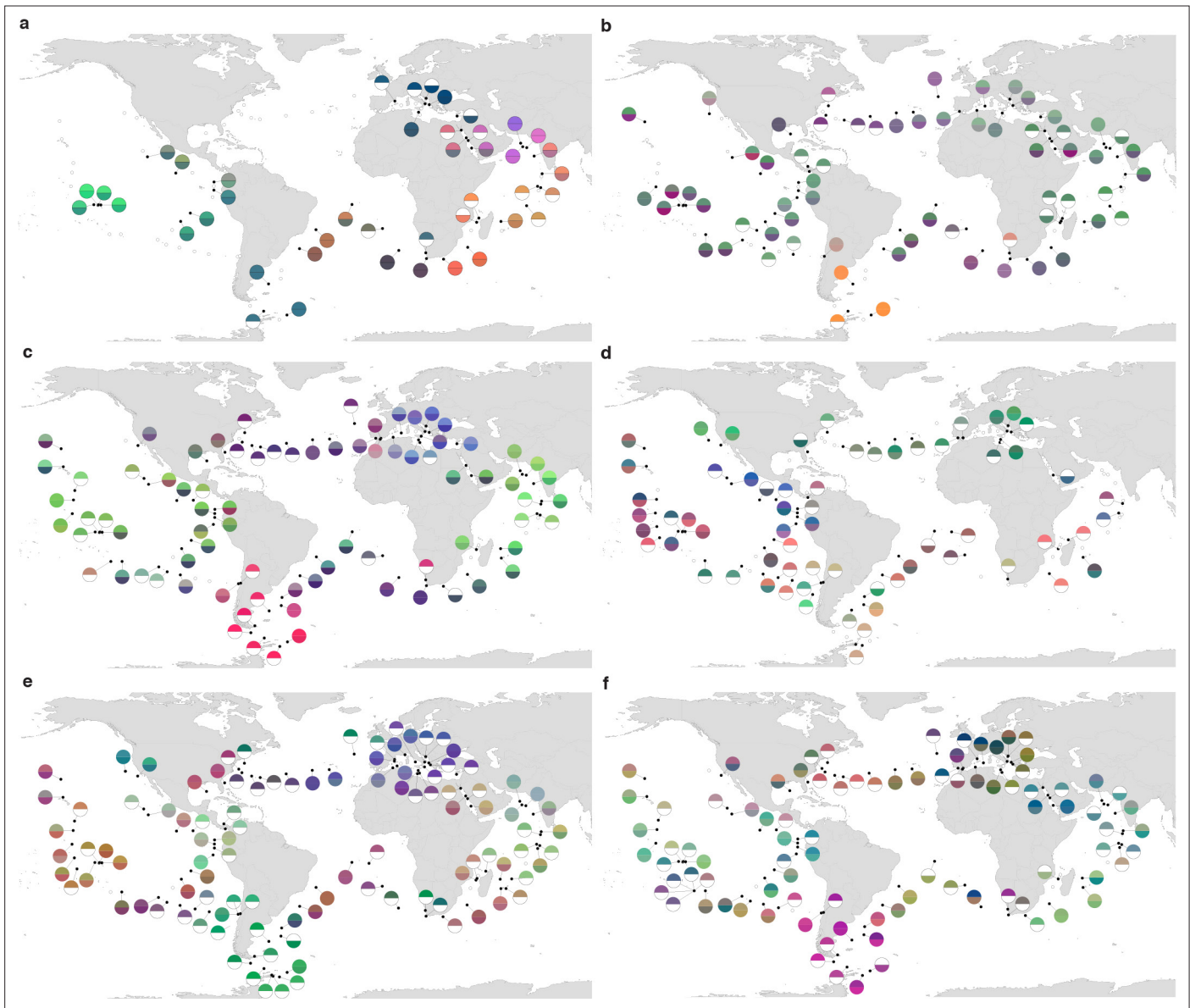


Figure 1—figure supplement 5. Biogeography based on an ordination of operational taxonomic unit (OTU) dissimilarity. (a–f) Principal coordinates analysis-RGB color maps for OTUs (see Materials and methods). The top of each half circle represents samples collected at the surface and the bottom portion represents the deep chlorophyll maximum (stations missing OTU data for one of the two depths are drawn as half circles). Station colors do not correspond among size fractions. (a) 0–0.2 μm . (b) 0.22–1.6/3 μm . (c) 0.8–5 μm . (d) 5–20 μm . (e) 20–180 μm . (f) 180–2000 μm .

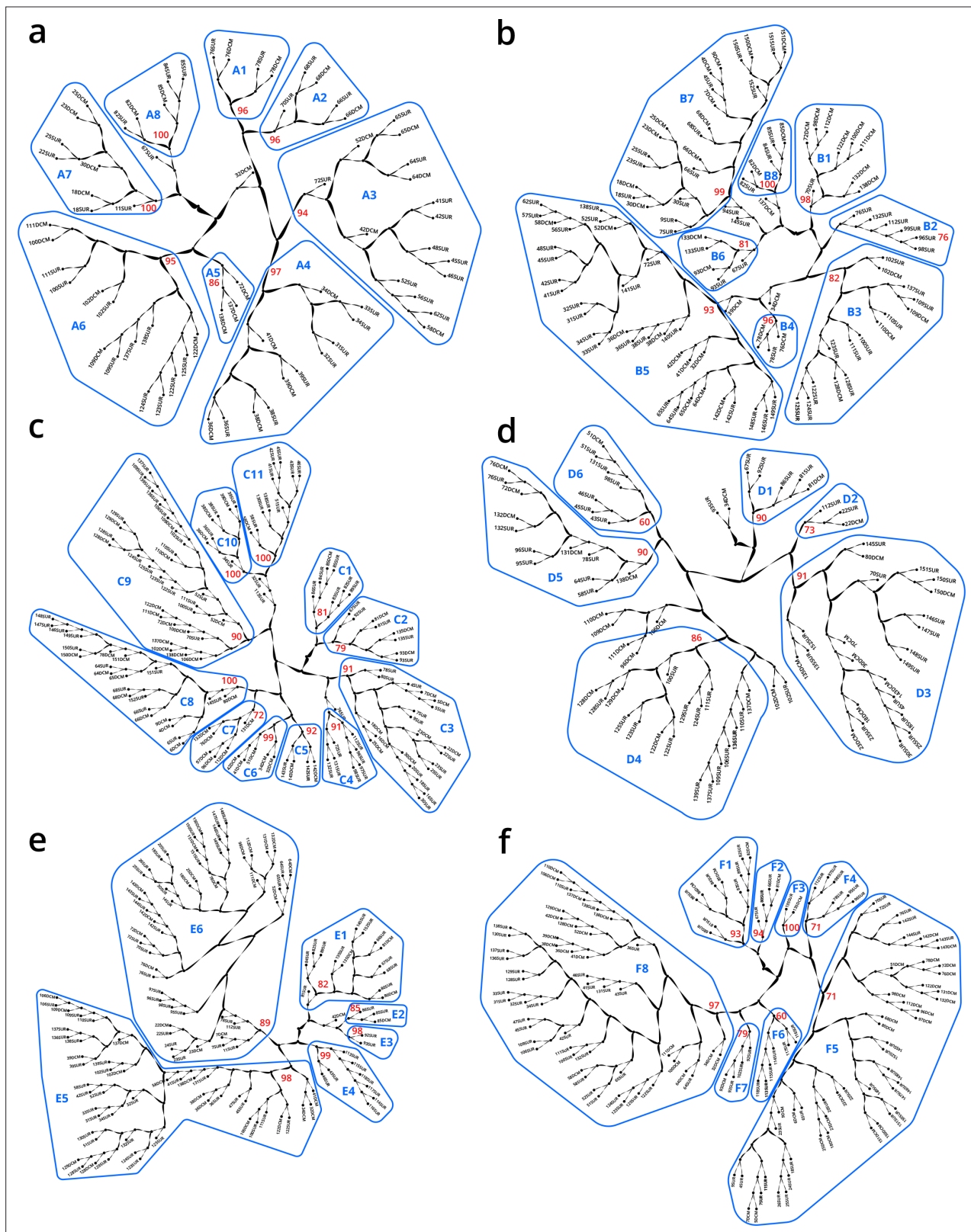


Figure 1—figure supplement 6. Hierarchical trees illustrating how samples were partitioned into genomic provinces. Dendrograms resulted from unweighted pair group method with arithmetic mean clustering. Each sample (SUR: surface, DCM: deep chlorophyll maximum) is shown as a leaf. Genomic provinces are shown with their identifiers in blue polygons; identifiers are composed of one letter prefix per size fraction (A–F) and a number. Bootstrap values in red show the support at the key nodes that separate genomic provinces from one another. See also Appendix 2 on the robustness *Figure 1—figure supplement 6 continued on next page*

Figure 1—figure supplement 6 continued

of genomic provinces. **(a)** 'A' prefix, 0–0.2 μm size fraction. **(b)** 'B' prefix, 0.22–1.6/3 μm . **(c)** 'C' prefix, 0.8–5 μm . **(d)** 'D' prefix, 5–20 μm . **(e)** 'E' prefix, 20–180 μm . **(f)** 'F' prefix, 180–2000 μm .

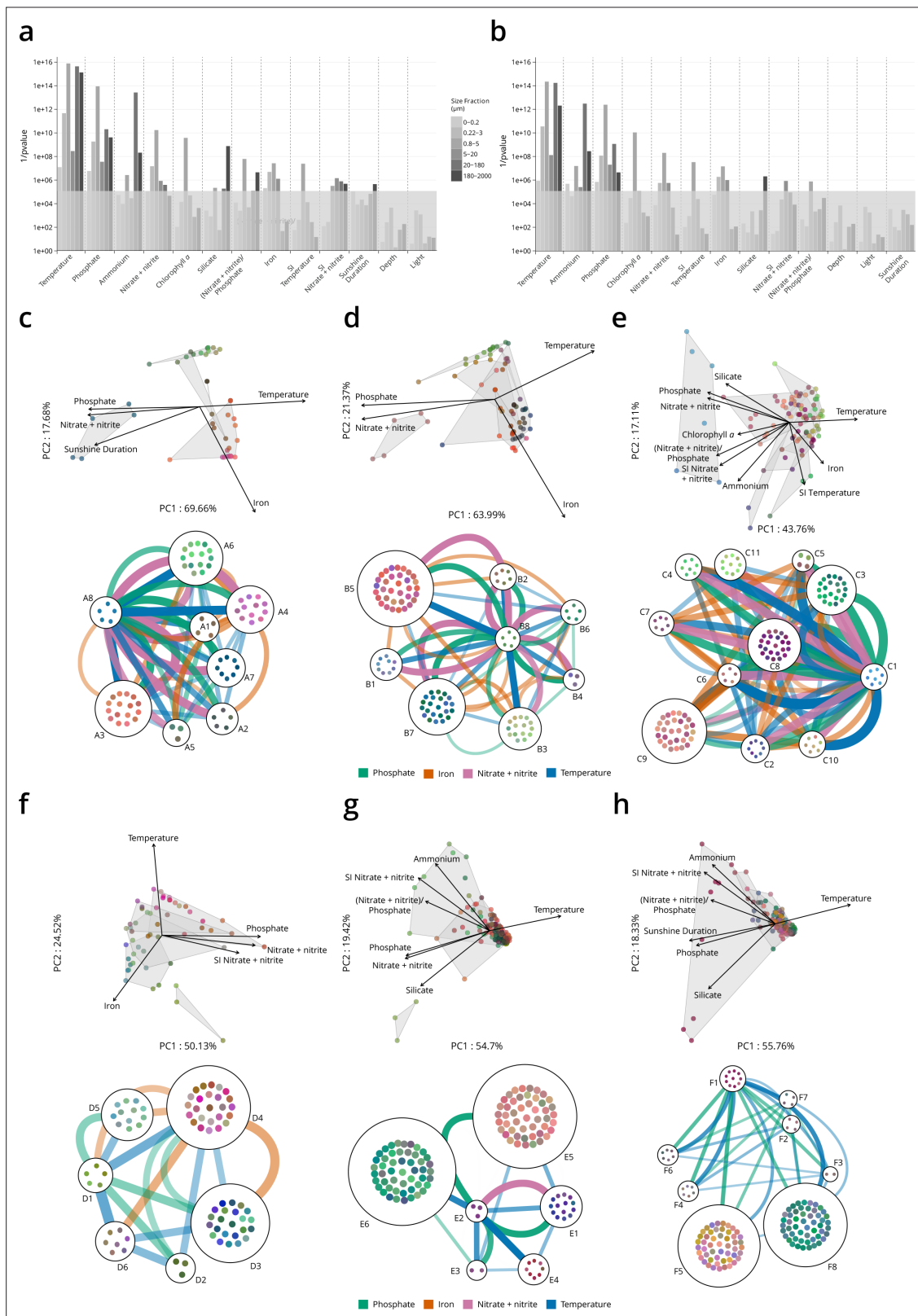


Figure 1—figure supplement 7. Environmental parameters that distinguish genomic provinces. **(a–b)** Environmental parameters that significantly differentiate among genomic provinces (Kruskal-Wallis test, gray box indicates p values $>10^{-5}$). SI: seasonality index. **(a)** All stations. **(b)** Antarctic stations removed (see Materials and methods). Eliminating Antarctic stations does not result in a large change in the parameters that significantly differentiate among provinces. **(c–h)** Two types of visualizations of the relationships between genomic provinces and environmental parameters. Sample colors

Figure 1—figure supplement 7 continued on next page

Figure 1—figure supplement 7 continued

are those from **Figure 1—figure supplement 4a-f**. Top plots within panels **(c-h)**: principal components analysis-based visualization. Samples and environmental parameters differing significantly ($p \leq 10^{-5}$) among genomic provinces, are projected onto the first two axes of variation. Gray polygons enclose different genomic provinces. Bottom plots within panels **(c-h)**: network-based visualization. Each genomic province is represented as a node, with the individual samples composing the province within the node. Edges between nodes represent differences in temperature, nitrate + nitrite, phosphate, and iron that significantly differentiate ($p \leq 10^{-5}$) among genomic provinces, that are statistically significantly different between individual pairs of genomic provinces (post hoc Tukey test, $p < 0.01$) and whose difference in median parameter values is ≥ 1 SD (calculated from the parameter values of all samples in the size fraction). Thicker edges represent larger differences. **(c)** 0–0.2 μm size fraction. **(d)** 0.22–1.6/3 μm . **(e)** 0.8–5 μm . **(f)** 5–20 μm . **(g)** 20–180 μm . **(h)** 180–2000 μm .

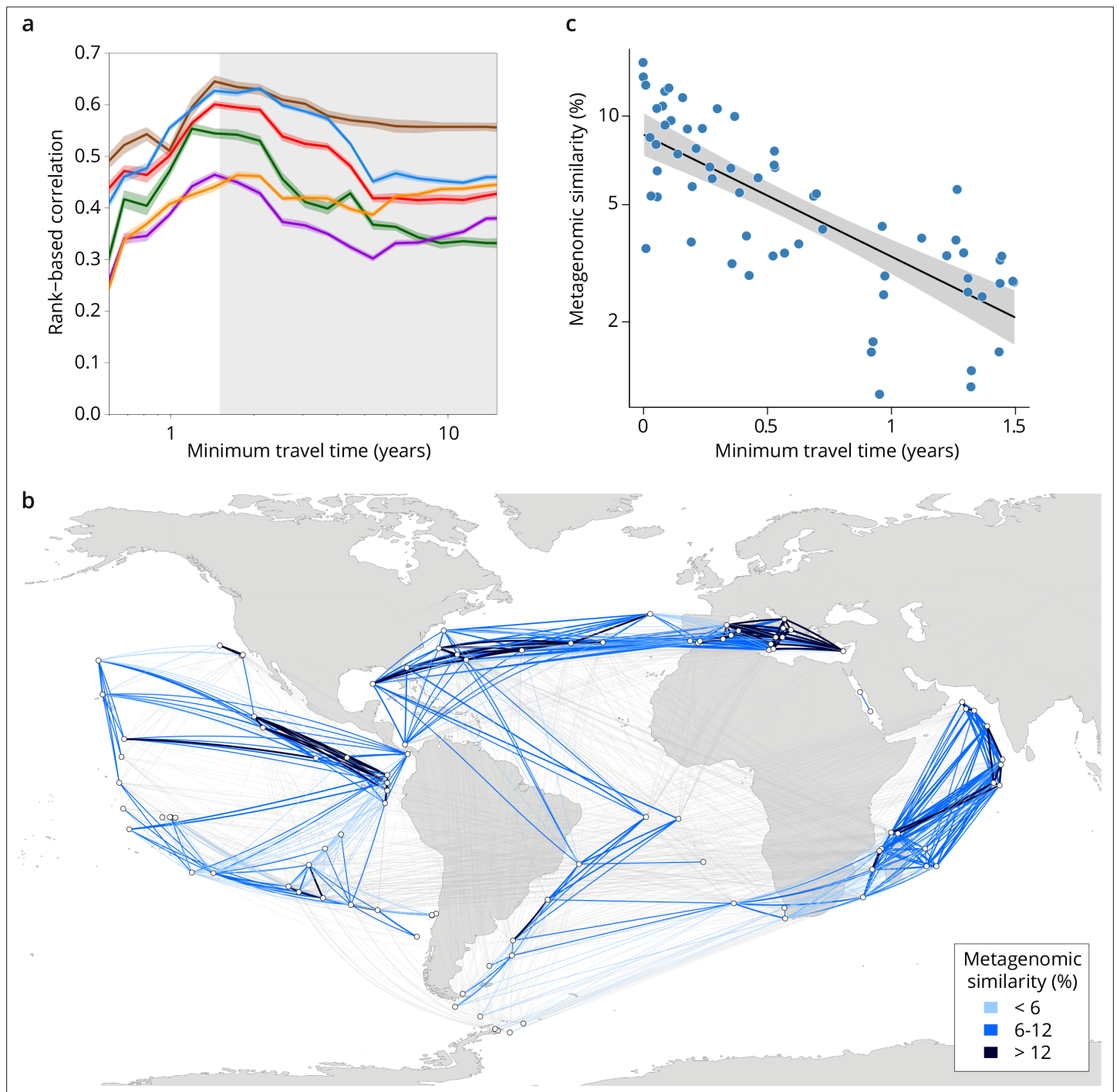


Figure 2. Metagenomic dissimilarity and travel time of plankton are maximally correlated up to ~1.5 years. **(a)** Spearman rank-based correlation by size fraction between metagenomic dissimilarity and minimum travel time along ocean currents (T_{\min}) for pairs of *Tara* Oceans samples separated by a minimum travel time less than the value of T_{\min} on the x-axis. Brown line: 0–0.2 μm size fraction, red: 0.22–1.6/3 μm , blue: 0.8–5 μm , green: 5–20 μm , purple: 20–180 μm , orange: 180–2000 μm . Shaded colored areas represent 95% CI. $T_{\min} > 1.5$ years is shaded in gray. See plots for operational taxonomic unit (OTU) dissimilarity in **Figure 3—figure supplement 1**. **(b)** Pairs of *Tara* stations connected by $T_{\min} < 1.5$ years in blue/black and > 1.5 years in gray. Shading reflects metagenomic similarity from the 0.8–5 μm size fraction. **(c)** The relationship of metagenomic similarity to T_{\min} with an exponential fit (black line, gray 95% CI), for pairs of surface samples in the 0.8–5 μm size fraction within the North Atlantic and Mediterranean current system (see map and plots for other size fractions and OTUs in **Figure 2—figure supplement 2**, and Appendix 1 for a discussion of metagenomic similarity).

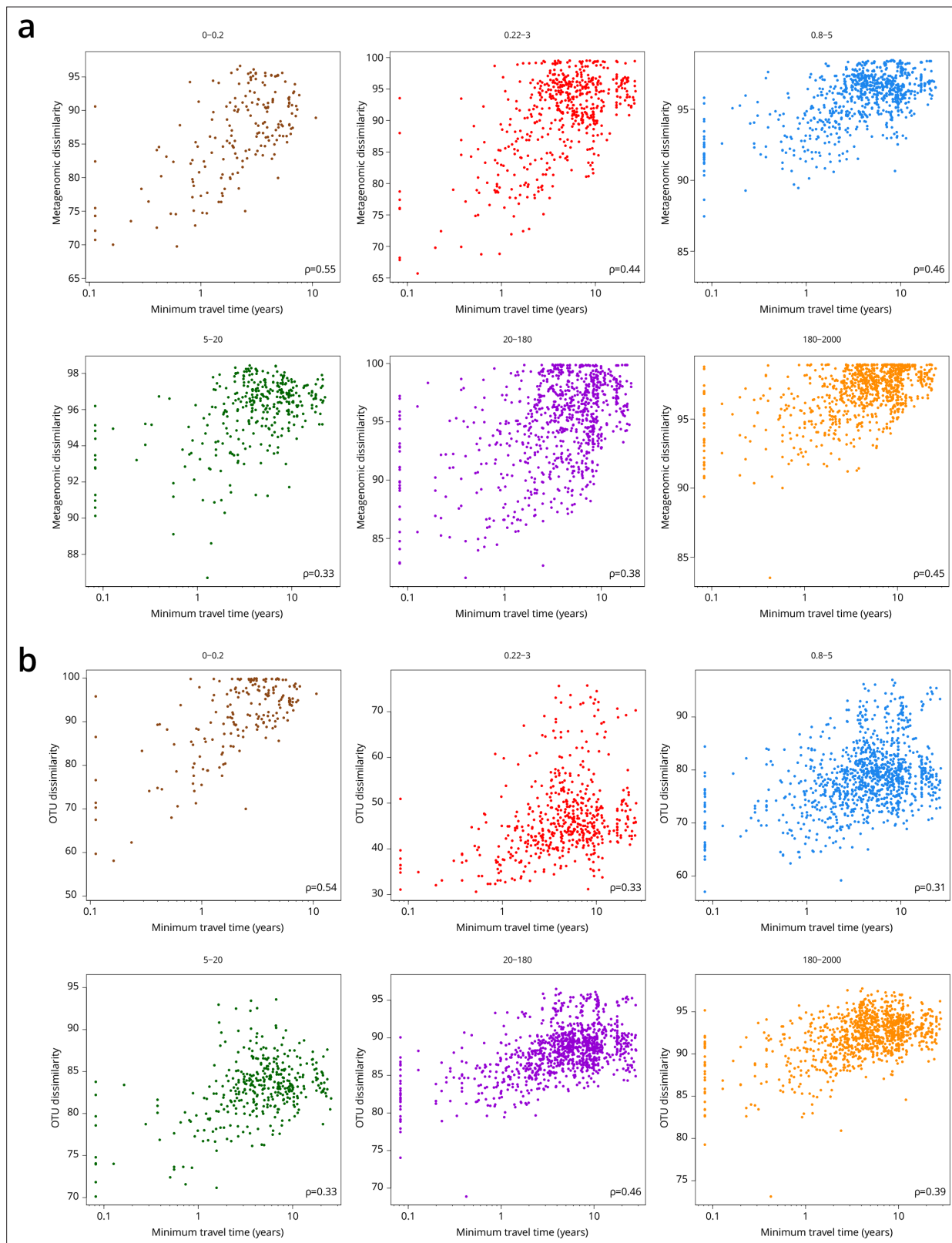


Figure 2—figure supplement 1. Global correlations of dissimilarity with minimum travel time (T_{min}). Scatter plots of dissimilarity versus T_{min} . One point represents a pair of samples. (a) Metagenomic dissimilarity. (b) Operational taxonomic unit (OTU) dissimilarity. Global Spearman correlation values are indicated within each panel.

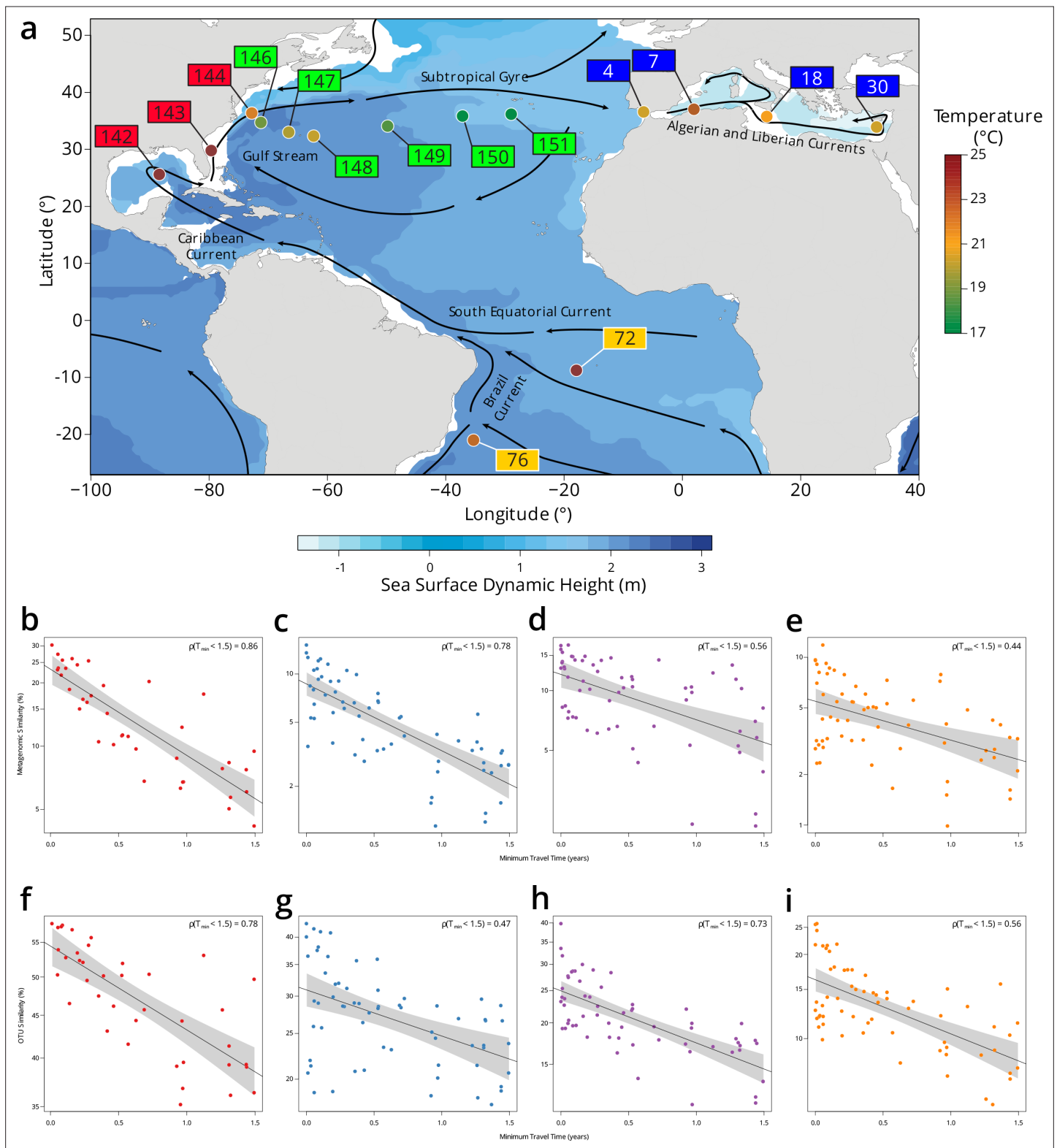


Figure 2—figure supplement 2. Plankton community composition turnover through the North Atlantic. (a) Map of Tara Oceans stations, currents (solid lines), temperature by station (colored circles), and sea surface climatological dynamic height from CARS2009 (<http://www.cmar.csiro.au/cars>). Each station label has a color corresponding to a subregion: South Atlantic in orange, Gulf Stream in red, Recirculation/Gyre in green, and Mediterranean Sea in blue. (b–e) Scatter plots of metagenomic similarity versus minimum travel time (T_{min}) for these stations in the (b) 0.22–3 μm ; (c) 0.8–5 μm ; (d) 20–180 μm ; and (e) 180–2000 μm size fractions. (f–i) Scatter plots of operational taxonomic unit (OTU) community similarity for the (f) 0.22–3 μm ; (g) 0.8–5 μm ; (h) 20–180 μm ; and (i) 180–2000 μm size fractions. Figure 2—figure supplement 2 continued on next page

Figure 2—figure supplement 2 continued

(h) 20–180 μm ; and (i) 180–2000 μm size fractions. The black line represents an exponential fit, with a light gray shaded 95% CI. The resulting turnover times using metagenomic similarity are $\tau = 0.91$ years for 0.22–3 μm , $\tau = 0.91$ years for 0.8–5 μm , $\tau = 2.22$ years for 20–180 μm , and $\tau = 1.99$ years for 180–2000 μm . Turnover times using the OTU community similarity are $\tau = 4.23$ years for 0.22–3 μm , $\tau = 4.08$ years for 0.8–5 μm , $\tau = 2.6$ years for 20–180 μm , and $\tau = 2.1$ years for 180–2000 μm . The viral-enriched 0–0.2 μm and the nanoplanktonic 5–20 μm size fractions are not shown due to insufficient sampling of these stations.

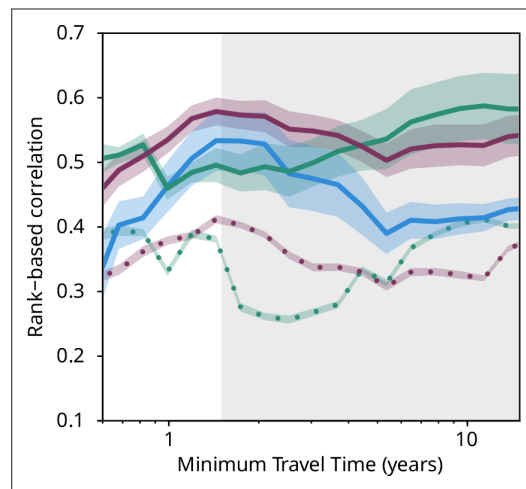


Figure 3. Plankton travel time, metagenomic dissimilarity, and environmental differences show different temporal patterns of pairwise correlation. Spearman rank-based correlations between metagenomic dissimilarity and minimum travel time (T_{\min} , blue), metagenomic dissimilarity and differences in $\text{NO}_3 + \text{NO}_2$, PO_4 , and Fe (purple), metagenomic dissimilarity and differences in temperature (turquoise), T_{\min} and differences in $\text{NO}_3 + \text{NO}_2$, PO_4 , and Fe (purple, dashed), and T_{\min} and differences in temperature (turquoise, dashed) for pairs of Tara Oceans samples separated by a minimum travel time less than the value of T_{\min} on the x-axis. Shaded regions represent SEM. Correlations represent averages across four of six size fractions represented in **Figure 2a**; the 0–0.2 μm and 5–20 μm size fractions are excluded due to a lack of samples at the global level. Individual size fractions, partial correlations, and correlations with operational taxonomic unit data are in **Figure 3—figure supplement 1**. Color palette is from microshades (<https://github.com/KarstensLab/microshades>) (Dahl et al., 2022).

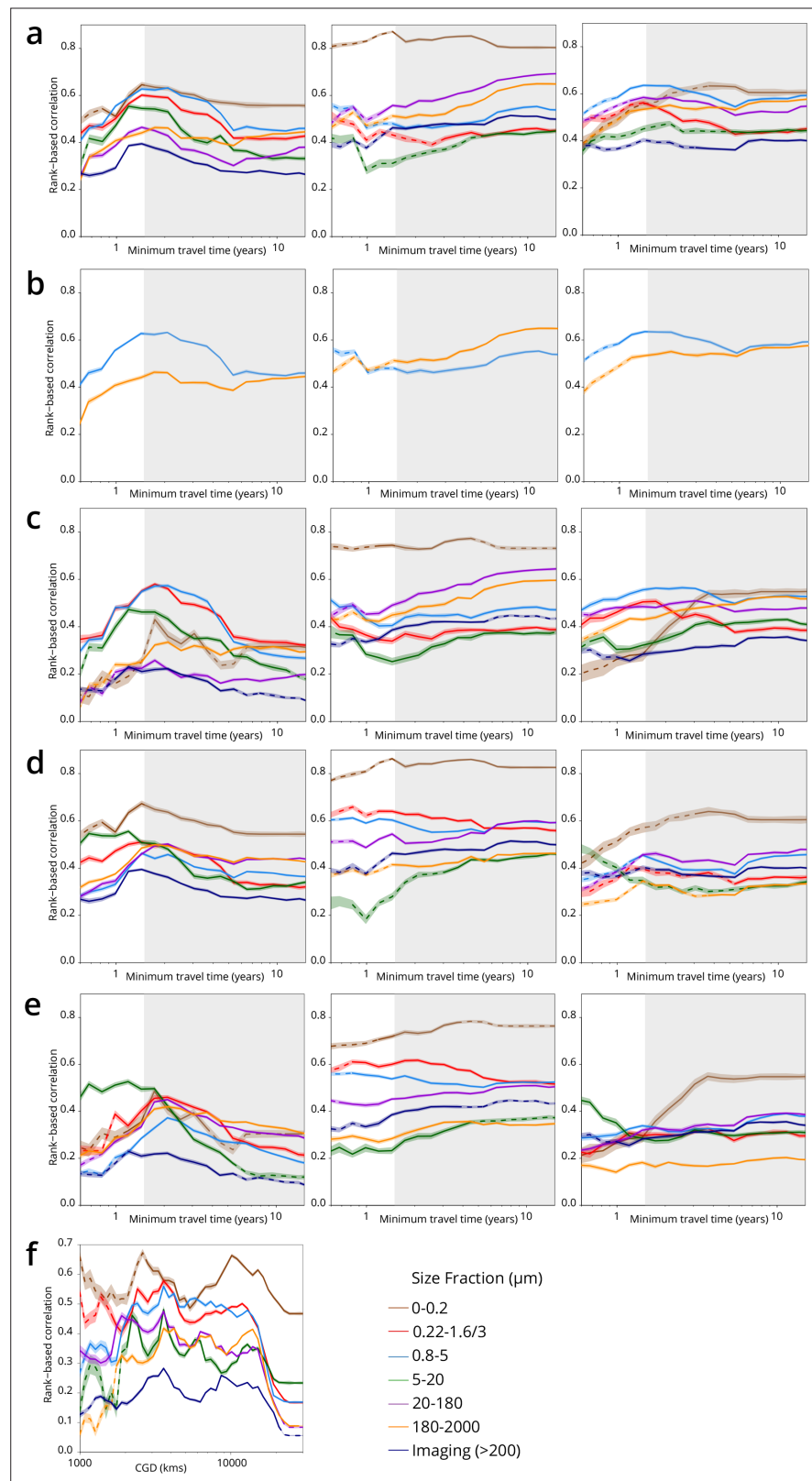


Figure 3—figure supplement 1. Plankton travel time, dissimilarity, environmental distance, and geographic distance show different temporal patterns of pairwise correlation. Spearman correlation values are shown separately by organismal size fraction. Non-significant correlations ($p > 0.01$) are shown with dashed lines. (a–e) Correlations for pairs of Tara Oceans samples separated by a minimum travel time less than the value of T_{min} .
 Figure 3—figure supplement 1 continued on next page

Figure 3—figure supplement 1 continued

on the x-axis. $T_{\min} > 1.5$ years is shaded in gray. Left panels: correlation of dissimilarity with T_{\min} , middle panels: dissimilarity with temperature, right panels: dissimilarity with differences in $\text{NO}_3 + \text{NO}_2$, PO_4 , and Fe. **(a–c)** Metagenomic dissimilarity. **(d–e)** Operational taxonomic unit dissimilarity. Correlations for imaging dissimilarity are superimposed on plots in **(a)** and **(c–e)** for comparison. There is a maximum correlation of dissimilarity with T_{\min} (and, for most size fractions, of dissimilarity with nutrients) for $T_{\min} < \sim 1.5$ years, but the correlation between dissimilarity and temperature does not display a similar maximum. **(b)** Displays only the 0.8–5 μm (blue) and 180–2000 μm (orange) size fractions from **(a)** to highlight that for smaller plankton, correlations with differences in nutrient concentrations were stronger for T_{\min} up to ~ 1.5 years, but for larger plankton, correlations were stronger with temperature variations for T_{\min} beyond ~ 1.5 years. **(c)** and **(e)** Partial correlations to estimate the independent effects of T_{\min} and environmental distances on β -diversity. Left panels: controlling for differences in temperature and for differences in $\text{NO}_3 + \text{NO}_2$, PO_4 , and Fe; middle and right panels: controlling for T_{\min} . Partial correlations do not affect the maximum correlation of dissimilarity with T_{\min} for $T_{\min} < \sim 1.5$ years. **(f)** Correlation of geographic distance (without traversing land; CGD) with metagenomic dissimilarity or imaging dissimilarity for pairs of Tara Oceans samples separated by a geographic distance less than the value on the x-axis.