



**HAL**  
open science

# Patrons de biodiversité à l'échelle globale chez les dinoflagellés planctoniques marins

Noan Le Bescot

► **To cite this version:**

Noan Le Bescot. Patrons de biodiversité à l'échelle globale chez les dinoflagellés planctoniques marins. Sciences du Vivant [q-bio]. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : . tel-00974287v1

**HAL Id: tel-00974287**

**<https://theses.hal.science/tel-00974287v1>**

Submitted on 6 Apr 2014 (v1), last revised 9 Jul 2014 (v2)

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

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



**THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité : Biologie marine – École Doctorale Diversité du Vivant

Présentée par

**M. Noan Le Bescot**

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

Sujet de la thèse :

**Patrons de biodiversité à l'échelle globale chez les  
dinoflagellés planctoniques marins**

Soutenue le : 10 MARS 2014

Devant un jury composé de :

**Dr. Rowena Stern**, Sir Alister Hardy Foundation for Ocean Science (SAHFOS), Plymouth — Royaume-Uni

**Prof. Jan Pawlowski**, Université de Genève — Suisse

**Dr. Marina Montesor**, Stazione Zoologica Anton Dohrn of Naples — Italie

**Prof. Lars Stemmann**, UPMC/CNRS Station Zoologique de Villefranche-sur-Mer — France

**Dr. Raffaele Siano**, DYNECO/Pelagos, IFREMER Centre de Brest, Plouzané — France

**Dr. Colomban de Vargas**, UPMC/CNRS Station Biologique de Roscoff — France

Rapporteur

Rapporteur

Examineur

Examineur

Co-directeur de thèse

Directeur de thèse









---

## Remerciements

A l'issue de ce travail de recherche et de sa rédaction, j'ai la preuve que la thèse est loin d'être un travail solitaire. En effet, je n'aurais jamais pu réaliser ce travail doctoral sans le soutien d'un grand nombre de personnes dont l'amitié, la générosité, la bonne humeur et l'intérêt manifestés à l'égard de ma recherche m'ont permis de progresser dans cette phase délicate de « l'apprenti-chercheur ».

En premier lieu, je tiens à remercier mes directeurs de thèse, M. Colombaro de Vargas, pour la confiance qu'il m'a accordée en me proposant une thèse au sein de son laboratoire et M. Raffaele Siano qui a accepté d'encadrer ce travail doctoral. Merci pour vos multiples conseils et pour le temps que vous m'avez consacré. Merci Colombaro de m'avoir laissé parcourir les océans, et pour ta très grande ouverture d'esprit. Tu m'as offert un très beau sujet de recherche, et beaucoup de liberté. J'aimerais vous dire à tous les deux à quel point j'ai apprécié les importants moyens que vous avez su mettre à ma disposition pour que mon travail soit le plus riche et le plus ambitieux possible. J'ai conscience de la chance que vous m'avez donnée. Enfin, j'ai été extrêmement sensible à vos qualités humaines d'écoute et de compréhension tout au long de ce travail doctoral. Merci.

Je souhaiterais exprimer ma gratitude à messieurs Etienne Bourgois, Philippe Clais et Romain Troublé pour m'avoir donné l'envie de poursuivre, de réaliser une thèse, et fourni l'opportunité de réaliser un petit bout de mes rêves d'enfant sur le bateau Tara. Je remercie également toute l'équipe de Tara expéditions dans le bureau à Paris pour son accueil chaleureux à chaque fois que j'ai sollicité son aide, ainsi que pour ses multiples encouragements. Merci mes amis. Je vous donne rendez-vous très vite pour de nouveaux projets toujours plus ambitieux autour du monde.

Mes remerciements vont également à M. Lars Stemmann et Mme Marina Montresor d'avoir accepté de participer à ce jury de thèse. Je suis infiniment gré à Mme. Rowena Stern d'avoir accepté la fonction de rapporteur. De même, je suis particulièrement reconnaissant à M. Jan Pawlowski de l'intérêt qu'il a manifesté à l'égard de cette recherche en s'engageant à être rapporteur. M. Alban Ramette, M. Rodolphe Lemée et Mme Laure Guillou je vous remercie d'avoir accepté d'évaluer mon travail tout au long de ma thèse et d'avoir su me remettre sur le bon chemin quand il le fallait. Je tiens à remercier Bernard Kloreg directeur de la station biologique, et François Lallier, directeur de l'UMR 7144 pour leur accueil.

Je remercie la région Bretagne et l'université Paris 6 Pierre et Marie Curie, au travers de son président M. Jean-Charles Pomerol, qui ont financé cette thèse.

Mon travail de recherche sur le terrain a été particulièrement facilité par le soutien financier de l'entreprise agnès b. (financier majoritaire de l'expédition Tara Oceans) et de ses partenaires. Je tiens à exprimer tout particulièrement ma reconnaissance aux marins avec qui j'ai parcouru plus de 11 000 kilomètres. Hervé Bourmaud (Capitaine) pour m'avoir toujours fait confiance et initié à la navigation hauturière, merci pour tous ces merveilleux souvenirs. Loïc Valette (Capitaine) pour m'avoir accueilli très chaleureusement à bord pour un long périple dans le Pacifique Nord ; ainsi que pour ta confiance et l'intérêt que tu as manifesté, vis-à-vis de mon travail. Julien Girardot (Cuistographe) pour ta bonne humeur, ton énergie débordante, tes bons petits plats et tes précieux conseils photo. Daniel Cron, mon Dani-Danou (Chef mécanicien), le mécano le plus farceur et drôle que je connaisse, tes déhanchés me manquent déjà ! François Noël, Le Vieux (Chef mécanicien), véritable McGyver de tous les temps qui a toujours su trouver une solution et a sauvé bon nombre de nos échantillonnages. Vincent Le Pennec (Second Capitaine) pour nos escapades dans les hauteurs marquisiennes, c'est vrai tu es le digne fils de Hulk ! Martin Herteau (Marin) pour nos fous-rires en cabine à San Diego ; Baptiste Bernard (Mousse), dit Le Jeune, pour tes 18 ans inoubliables au milieu du Pacifique, ta bonne humeur à bord ; et bien entendu François Aurat (Marin) pour nos virées en Zodiac et le plaisir partagé de faire de belles images. Alain Giese (Second Capitaine), merci d'avoir partagé un coin de ton bureau climatisé avec moi, de m'avoir fait rêver devant tes récits et tes images de *Tuiga*. Céline Blanchard (Cuisinière), merci pour tes petits plats et ta gentillesse, et encore bravo pour avoir su garder des semaines durant des fruits et des légumes dans nos assiettes. Sibylle d'Orgeval (Journaliste), je nous revois aux Gambier photographier les enfants, les interviewer, plonger sur les récifs parmi

## Remerciements

---

les requins ou encore parcourir les montagnes marquisiennes, de grands et beaux moments. Yann Chavance (Journaliste), je me souviens de notre départ impatient de San Diego, parmi les baleines à bosse et les loutres de mer, ou encore de notre virée en zodiac au large de Clipperton, et enfin de notre admiration face aux portes d'écluse du Canal de Panama. Je garderai longtemps le souvenir de ces heures passées à contempler l'océan à vos côtés. Merci à tous. Bon vent !

Bien entendu je n'oublie pas mes nombreux collègues scientifiques et amis qui ont foulé des jours durant le pont de Tara à mes côtés, partagé ma vie, les joies et les galères de la science en mer. Francesca Benzoni, Emmanuel Reynaud, Fabrice Not, Gabriele Procaccini (Chefs de mission) je vous remercie pour votre confiance, toutes les heures de discussion et d'échanges très enrichissants que nous avons eu autour de la stratégie d'échantillonnage. Francesca, Gabriele, ce petit air d'Italie me manque ! Manu, on repart quant en mission ? D'autres cultures, d'autres lieux, sont sans aucun doute à découvrir, à préserver et à imager ! Sarah Searson (ingénieur de pont), ma néo-zélandaise préférée, que de bons souvenirs et de bon temps passés à tes côtés sur la plage arrière de Tara ou à terre. Mon bracelet jaune fluo est toujours bien accroché à mon poignet. Merci beaucoup pour ta confiance, ta patience et ton immense gentillesse. Éric Beiro, je n'oublierai pas ton humour et ton port du slip (de bain) de si tôt ! Claudio Stalder, je nous revois parcourant les Gambier à la recherche de foraminifères, en compagnie de Cornelia Maier tous deux toujours volontaires et souriants. Éric Roettinger, mon mentor en photographie animalière, ton talent est égal à ta pédagogie, à ta patience et à ta gentillesse. Grâce à toi je sais aujourd'hui comment tailler le portrait du petit peuple des océans. Que de bons souvenirs en plongée à tes côtés, à chasser ces petits êtres si fascinants ou encore pendant nos longues nuits à les photographier sur le pont arrière. Francesca, merci de m'avoir pris sous ton aile et fait confiance pour cette inoubliable mission. Mathieu Oriot, notre responsable plongée mais aussi marin, nous avons passé de sacrés bons moments à parler science ou à naviguer. Nous avons imaginé pas mal de choses pendant nos quarts, en zodiac ou encore dans la cabine et tu en as concrétisé certains, bravo et bonne chance pour la suite !

Fabrizio d'Ortenzo, Pierre Testor les deux amis passionnés qui manient comme personne l'art de planer sous l'eau ou plus exactement de faire planer un petit sous-marin jaune au milieu de l'océan. Merci pour nos riches échanges scientifiques sur la structure des océans et l'adrénaline de la chasse pour aller repêcher votre petit protégé. Hervé Le Goff dit Astérix, sacré bidouilleur électronicien qui m'a bien aidé dans nos galères informatiques mais aussi et surtout sur le pont pour échantillonner. Manu encore une fois merci pour ton soutien et tes belles images de mes filets. Sophie Nicaud, ta bonne humeur et tes précieux coups de main pendant les stations furent déterminants. Bravo pour le petit récit que tu en as fait à ton retour. Steffi Kandels-Lewis, je nous revois au coucher du soleil à Fakarava jouant avec les reflets sur l'eau ou encore sur le roof de Tara à disserter du futur de la science et de la beauté des lieux. Céline Dimier, ma camarade shaddock, celle avec qui j'ai pompé, pompé, pompé tant de litres d'eau de mer que nous aurions pu couler une bonne dizaine de fois le bateau. Merci pour ta bonne humeur et ton aide précieuse, on formait une bonne équipe je crois !

Je n'oublie pas bien entendu Francisco Cornejo Castillo Miguel, mon autre compère shaddock avec qui j'ai partagé les galères mais aussi et surtout de mémorables fous-rires. Gabriele, j'ai adoré travailler à tes côtés, j'espère que nous nous reverrons bientôt. Denis de la Broise, ton aide aux côtés de Sarah et ta gentillesse a fait de cette longue mission un succès. Dorothée Désobry (artiste), ton énergie, tes coups de gueule mais aussi ta joie de vivre ont égaillé notre périple. Luis Gutierrez, mon ami mexicain, l'île de Clipperton nous a résisté, mais la isla de Coco nous a ravis. Je suis prêt à repartir quand tu veux faire de belles images avec toi ! Et enfin Bianca Silva, un très gros bisou pour te remercier de tes sourires, de ta joie de vivre chaque matin dans la cabine, de ces petits instants de douceur que tu as su si bien et si discrètement entretenir tout au long de notre voyage. Tous, je pense bien fort à vous.

Je souhaite bien entendu remercier tous les membres de mon équipe à Roscoff pour ces 3 années très enrichissantes passées à vos côtés. Raphaël Morard pour nos discussions scientifiques, nos délires tardifs à Kerjeff et nos playlists à décorner les bœufs. Sébastien Colin pour m'avoir supporté dans le bureau et beaucoup aidé dans mes analyses en imagerie. Jean Pennors pour ta chaleureuse compagnie jour après jour. Marie-José Garet pour ta sympathie et ton soutien tout au long de mes travaux, notamment lors de la relecture de ce manuscrit. Sarah Romac, la blouse blanche toujours impeccable et les gants bleus chaussés, toujours prête pour

une PCR. Merci pour ton aide dans mes travaux de laboratoire. Johan Decelle, Mahdi Bendif, Nicolas Henry, Daniel Richter merci pour l'ambiance toujours conviviale que vous entretenez au sein de l'équipe. Je n'oublie bien évidemment pas tous les autres membres de l'équipe actuelle et passée, Damien Guiffant, Frédéric Mahé, Angélique Gobet, Lucie Bittner, Margaux Carmichaël, les stagiaires et tous les autres. Cet esprit de travail bon enfant c'est vous tous qui le faites !

Je souhaite aussi exprimer ma gratitude à tous mes autres collègues chercheurs pour l'intérêt dont ils ont fait preuve envers ma recherche, les moyens qu'ils ont mis en œuvre pour me donner accès à leur laboratoire, pour leur accueil toujours enthousiaste à chaque fois que je les ai rencontrés. Je tiens à témoigner toute ma reconnaissance à M. Philippe Grosjean (Belgique), pour sa bienveillance et son accueil chez lui puis dans son laboratoire. Kevin Denis (Belgique), Jean-Baptiste Romagnan et Alina Tunin Ley (France), pour le temps que vous m'avez accordé et l'aide précieuse fournie pour mettre en place et réaliser mes analyses. Merci également à Maureen Williams (Irlande) pour la poursuite de mes travaux à Dublin. Je pense aussi à tous ceux avec qui j'ai étroitement collaboré, échangé et partagé mes recherches : Éric Karsenti, Stéphane Pesant, Uros Krzic (Allemagne), Silvia Acinas Gonzalez, Isabel Ferrara (Espagne), Nicole Poulton, Mike Sieraki, Matthew B Sullivan (USA), Mattias Ormestad (Suède), Jan Pawlowski (Suisse), Adriana Zingone, Eleonora Scalco, Daniele Ludicone (Italie), Gaby Gorsky, Marc Picheral, Lars Stemmann, Christian Rouvière, John R Dolan, Martina Ferraris, Lionel Guidi, Samuel Chaffron, Josephine Ras, Nigel Grimsley, Yves Desdevises, Hiroyuki Ogata, Émilie Villar, Defne Arselan, Chris Bowler, Atsuko Tanaka, Shruti Malviya, Alice Soccodato, Gipsi Lima-Mendez, Sabrina Speich, Florian Kokoszka (France) et tous les autres que je ne peux pas citer ici.

Ces remerciements seraient incomplets si je n'en adressais pas à l'ensemble des membres de la station biologique et du groupe plancton pour leurs nombreux conseils, leur soutien logistique et moral ainsi que pour la très bonne ambiance que j'ai toujours trouvée au sein de la station biologique. Je vous remercie et vous embrasse tous très très chaleureusement. Je voudrais aussi remercier le personnel administratif pour son soutien sans failles face au défi de la paperasse et l'équipe de restauration/hébergement pour la rapidité et la qualité de leurs services.

J'exprime ma sincère gratitude à tous les amis - chercheurs qui m'ont aidé de manière bénévole tout au long de ces trois années passées à Roscoff. Je souhaite tout particulièrement remercier mes deux mousquetaires, Docteurs et amis, Aliou Dia et Frédéric Lepelletier, pour les moments inoubliables passés à la station et surtout en dehors ! De Barcelone à Brest, nous en avons vécu des aventures ! Votre soutien et votre aide sans limite dans les derniers instants de ce travail me sont allés droit au cœur. Merci les gars ! Doriane Stagnol Pagès, je ne t'oublie pas bien évidemment, merci pour tous ces bons souvenirs en dehors de la station.

Ma reconnaissance va bien entendu à ceux qui ont plus particulièrement assuré le soutien affectif de ce travail doctoral : ma famille. Merci à mes parents qui m'ont toujours soutenu quel qu'en ait été le prix, mes oncles, mes tantes toujours là pour m'encourager et ma grand-mère qui m'a toujours aidé, motivé dans mes études. J'aurais aimé te présenter ce travail.

Enfin, j'adresse mille mercis à Delphine Rioual pour m'avoir souvent vu travailler très tard et toujours avoir fait preuve de tolérance, et avoir su exprimer ton soutien affectif sans failles. Tu as su me donner cette force incroyable pour aller au bout.

Cette thèse est dédiée à mes parents, Anny et Ange, pour leur soutien moral, leur affection et leur patience.



## Sommaire

|   |           |
|---|-----------|
| <b>Sommaire</b>   | <b>1</b>  |
| <b>Résumé</b>   | <b>5</b>  |
| <b>Publications</b>   | <b>6</b>  |
| <b>Abstract</b>   | <b>7</b>  |
| <br>  |           |
| <b>Préambule</b>  | <b>11</b> |
| <br>  |           |
| <b>Introduction</b>   | <b>17</b> |
| <b>1. Le plancton</b>                                       | <b>17</b> |
| 1.1. <i>La classification phylogénétique des espèces</i>    | 19        |
| 1.2. <i>Les Alvéolés</i>                                    | 21        |
| <b>2. La diversité des dinoflagellés</b>                    | <b>21</b> |
| 2.1. <i>Biologie : description générale</i>                 | 21        |
| 2.2. <i>Reproduction</i>                                    | 24        |
| 2.3. <i>Classification</i>                                  | 26        |
| 2.4. <i>Biogéographie</i>                                   | 28        |
| 2.5. <i>Écologie</i>  | 28        |
| 2.6. <i>Comment étudier la micro-biodiversité marine ?</i>  | 29        |
| <b>3. Les code-barres génétiques</b>                        | <b>30</b> |
| 3.1. <i>Le concept</i>                                      | 30        |
| 3.2. <i>Développement du barcoding</i>                      | 31        |
| 3.3. <i>Le barcoding en milieu marin</i>                    | 32        |
| 3.4. <i>Le barcoding chez les protistes</i>                 | 33        |
| <b>4. Les code-barres génétiques chez les dinoflagellés</b> | <b>34</b> |
| <br>  |           |
| <b>Objectifs de la thèse et structure du document</b>       | <b>37</b> |
| <br>  |           |
| <i>Références</i>   | 39        |

## Partie 1

|   |           |
|---|-----------|
| <b>Chapitre 1 : L'expédition Tara Océans : Un échantillonnage complet du plancton marin dans son contexte environnemental</b> | <b>47</b> |
| <br>  |           |
| <b>Tara Océans</b>  | <b>49</b> |
| <b>1. Contexte du projet et objectifs</b>   | <b>49</b> |
| <b>2. Les principales problématiques de l'échantillonnage</b>   | <b>50</b> |
| 2.1. <i>Les contraintes techniques</i>  | 50        |
| 2.2. <i>Les contraintes biologiques</i>   | 51        |
| 2.3. <i>La stratégie d'échantillonnage</i>  | 51        |
| <b>The Tara Oceans Expedition: A comprehensive sampling of marine plankton biota in their environmental context</b>           | <b>53</b> |

|  |            |
|--|------------|
| <b>Chapitre 2 : Analyse morphologique par imagerie à haut débit des dinoflagellés marins</b>                 | <b>69</b>  |
| <b>1. Introduction</b>   | <b>71</b>  |
| 1.1. <i>Histoire de l'étude des communautés planctoniques : l'enjeu de l'imagerie</i>                        | 71         |
| 1.2. <i>Les contraintes de l'analyse d'image à haut débit</i>  | 73         |
| 1.3. <i>Objectifs</i>  | 74         |
| <b>2. Matériel et méthode</b>  | <b>75</b>  |
| 2.1. <i>Présentation et caractéristiques du dispositif FlowCAM</i>   | 75         |
| 2.2. <i>Présentation et caractéristiques du logiciel Zoo/PhytoImage</i>                                      | 76         |
| 2.3. <i>Méthodologie à bord de Tara</i>  | 78         |
| 2.4. <i>Traitement des données en laboratoire</i>  | 78         |
| 2.5. <i>L'analyse du genre Neoceratium</i>   | 80         |
| <b>3. Résultats et discussion</b>  | <b>81</b>  |
| 3.1. <i>Limitation technique de l'analyse</i>  | 81         |
| 3.2. <i>Analyse des dinoflagellés identifiables par FlowCAM/ZooPhytoImage</i>                                | 82         |
| <b>4. Conclusion</b>   | <b>83</b>  |
| <i>Références</i>  | 85         |
| <br>   |            |
| <b>Chapitre 3 : Création et validation d'une base de référence pour le traitement des données génétiques</b> | <b>87</b>  |
| <br>   |            |
| <b>Introduction</b>  | <b>89</b>  |
| <b>Matériel et méthode</b>   | <b>91</b>  |
| 1. Échantillonnage   | 91         |
| 2. Traitement des données génétiques   | 92         |
| 3. Constitution de la base de référence  | 92         |
| 4. Évaluation de la qualité des assignations   | 93         |
| 5. Analyse des données morphologiques  | 93         |
| <b>Résultats</b>   | <b>95</b>  |
| 1. Base de référence   | 95         |
| 2. Phylogénie  | 95         |
| 3. Évaluation des séquences de référence   | 95         |
| 4. Comparaison metabarcoding V9 / morphologie  | 98         |
| 5. Comparaison metabarcoding V9 / observations FlowCAM pour le genre Neoceratium                             | 99         |
| <b>Discussion</b>  | <b>100</b> |
| 1. La base de référence V9   | 100        |
| 2. Phylogénie  | 100        |
| 3. Comparaison metabarcoding V9 / morphologie  | 101        |
| <b>Conclusion</b>  | <b>102</b> |
| <i>Références</i>  | 103        |

## Partie 2

|   |            |
|---|------------|
| <b>Chapitre 4 : Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding</b>                       | <b>107</b> |
| <b>Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding</b>                                    | <b>109</b> |
| Abstract  | 109        |
| <b>Introduction</b>   | <b>110</b> |
| <b>Experimental procedures</b>  | <b>112</b> |
| Sampling  | 112        |
| Genomic DNA extraction, purification and sequencing   | 113        |
| Database of reference dinoflagellate 18S rDNA and metabarcode annotation  | 115        |
| Assessment of molecular biodiversity  | 115        |
| <b>Results</b>  | <b>116</b> |
| Dinoflagellates versus protists   | 116        |
| Dinoflagellate community characterization   | 117        |
| Pico-dinoflagellates  | 120        |
| Nano-dinoflagellates  | 120        |
| Micro-dinoflagellates   | 121        |
| Dinoflagellate diversity in meso-plankton   | 121        |
| <b>Discussion</b>   | <b>125</b> |
| <b>Concluding remarks</b>   | <b>132</b> |
| <i>References</i>   | 133        |
| <br>  |            |
| <b>Chapitre 5 : Évaluation du rôle éco-fonctionnel des communautés de dinoflagellés à partir de metabarcoding massif dans l'ensemble des océans</b> | <b>137</b> |
| <b>Eco-functional assessment of the world's oceans dinoflagellate communities based on massive metabarcoding</b>                                    | <b>139</b> |
| <b>Introduction</b>   | <b>140</b> |
| <b>Matériel et méthode</b>  | <b>144</b> |
| Échantillonnage   | 144        |
| Traitement statistique  | 144        |
| <b>Résultat et discussion</b>   | <b>145</b> |
| <i>Références</i>   | 150        |
| <br>  |            |
| <b>Conclusion et perspectives</b>   | <b>151</b> |
| <br>  |            |
| <b>Références générales</b>   | <b>159</b> |
| <br>  |            |
| <b>Annexes</b>  |            |
| Annexe — Préambule  | 173        |
| Annexe — Chapitre 1   | 175        |
| Annexe — Chapitre 3   | 203        |
| Annexe — Chapitre 4   | 243        |





## Résumé

Les dinoflagellés (Alveolata, Dinophyceae) forment un groupe de protistes apparu il y a plus de 400 millions d'années, comportant  $\pm 2400$  espèces vivantes et  $\pm 2500$  espèces fossiles. Environ la moitié des espèces actuelles est photosynthétique et contribue significativement à la production primaire des écosystèmes aquatiques. Les autres espèces acquièrent leur carbone grâce à une diversité spectaculaire de relations écologiques : symbioses *sensu lato*, phagotrophie, mixotrophie. Les dinoflagellés présentent une diversité phénoménale de morphologies, physiologies, et comportements complexes, leur conférant une capacité supérieure d'adaptation à l'ensemble des milieux aquatiques marins et dulçaquicoles. Or cette même bio-complexité a rendu très difficile toute approche intégrative du groupe permettant l'établissement de stratégies adaptatives ou écologiques générales. L'objectif principal de cette thèse a été la recherche de patrons globaux de biodiversité et de structuration des communautés de dinoflagellés planctoniques à l'échelle des océans planétaires. Nous avons tout d'abord participé à l'élaboration d'un protocole d'échantillonnage morphogénétique de l'ensemble des dinoflagellés planctoniques, couvrant la totalité de leur spectre de taille cellulaire, de 3 à 500 $\mu\text{m}$ , et une partie importante de leurs niches spatio-temporelles dans la zone photique de l'océan mondial (Not et al., soumis). Ce protocole homogène et robuste, conçu pour fonctionner à bord d'un voilier, fut appliqué à chacun des 154 points d'échantillonnage de l'expédition circum-globale *Tara-Oceans* (2009-2012) (Karsenti et al., 2011). Nous avons ensuite exploré divers outils d'acquisition automatique à haut-débit de données permettant de quantifier la biodiversité des dinoflagellés marins pélagiques issue des échantillons Tara Océans. Une première approche basée sur la technologie d'imagerie automatique *FlowCAM*, a permis d'évaluer la diversité, l'abondance relative et la distribution géographique des morpho-espèces au sein du genre de dinoflagellés *Neoceratium* en mer Méditerranée (Le Bescot et al., In prep). Le métabarcoding du fragment V9 de l'ADN ribosomal 18S a permis, dans un deuxième temps, d'évaluer la diversité totale des dinoflagellés du plancton de la zone océanique photique. La construction d'une base de séquences ADNr de dinoflagellés de référence (DinR2 — Dinoflagellate Ribosomal Reference database), intégrée à une base de référence plus large concernant l'ensemble des protistes marins (PR2 — Protist Ribosomal Reference database) (Guillou et al., 2013) a permis l'assignation taxonomique des >170 millions de métabarcodes environnementaux. Des analyses de l'abondance et de la diversité des métabarcodes assignés ont permis de mettre en évidence des patrons généraux de structuration de la biodiversité des dinoflagellés à travers les classes de tailles organismiques et les niches écologiques explorées (Le Bescot et al., Submitted). L'abondance des différents ordres dépend essentiellement de la taille (pico-, nano-, micro-, et méso-plancton) indépendamment de l'écosystème étudié et de la période d'échantillonnage. Une diversité remarquable et insoupçonnée de pico-dinoflagellés (<5 $\mu\text{m}$ ) a été identifiée dans l'ordre des Gymnodiniales, avec un nombre très important de métabarcodes restant indéterminés. La fraction méso-planctonique (180-2000 $\mu\text{m}$ ), quant à elle, contient une forte diversité et abondance de genres de dinoflagellés parasites et photo-symbiotiques jouant un rôle fonctionnel potentiellement clef pour le maintien des communautés planctoniques (Le Bescot et al., Submitted). Finalement, une étude préliminaire de l'impact des conditions environnementales sur la méta-communauté des dinoflagellés a permis d'identifier certains facteurs environnementaux influant la structuration compositionnelle des communautés susmentionnées, ainsi que leur structuration fonctionnelle entre la zone photique et la zone méso pélagique (Le Bescot et al., In prep).

**Mots clefs :** plancton, dinoflagellé, métabarcode, biodiversité, écologie moléculaire, océan global, Tara-Océans

## Publications

- Le Bescot N.** March 10th 2014. Patrons de biodiversité à l'échelle globale chez les dinoflagellés planctoniques marins. Thèse de Doctorat de l'Université Pierre et Marie Curie.
- Le Bescot N.**, Audic S., Poulain J., Chaffron S., Guidi L., Siano R., de Vargas C. (in prep 2014). Eco-functional assessment of world oceans dinoflagellate communities based on massive metabarcoding. *Journal of Phycology*.
- Le Bescot N.**, Mahé F., Audic S., Katinka M., Wincker P., Céline Dimier, Gareth MJ, de Vargas C., Siano R., (submitted). Global patterns of dinoflagellate biodiversity and community structuring across protist size classes revealed by metabarcoding. *Environmental Microbiology*.
- Not F., Pesant S., Picheral M., Kandels-Lewis S., **Le Bescot N.**, et Al. (submitted). The Tara Oceans Expedition: A comprehensive sampling of marine plankton biota in their environmental context. *PlosOne*.
- Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., Boutte C., Burgaud G., de Vargas C., Decelle J., del Campo J., Dolan J., Dunthorn M., Bente E., Holzmann M., Kooistra W. H. C. F., Lara E., **Le Bescot N.**, Logares R., Mahé F., Massana R., Montresor M., Morard R., Not F., Pawlowski J., Probert I., Sauvadet A.-L., Siano R., Stoeck T., Vaulot D., Zimmermann P. & Christen R. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small SubUnit rRNA sequences with curated taxonomy. *Nucleic Acids Research*. 41:D597-D604
- Karsenti E., Acinas S.G., Bork P., Bowler C., de Vargas C., Raes J., Sullivan M., Arendt D., Benzoni F., Claverie J.-M., Follows M., Gorsky G., Hingamp P., Iudicone D., Jaillon O., Kandels-Lewis S., Krzic U., Not F., Ogata H., Pesant S., Reynaud E.G., Sardet C., Sieracki M.E., Speich S., Velayoudon D., Weissenbach J., Wincker P., the **Tara Oceans Consortium**, 2011. A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biol* 9, e1001177.

## Abstract

Dinoflagellates (Alveolata, Dinophyceae) are a group of protists that first appeared over 400 million years ago, and contain  $\pm 2,400$  living species and  $\pm 2,500$  fossil species. About half of the current dinoflagellate species are photosynthetic and contribute significantly to the primary production of aquatic ecosystems, whereas other species acquire their carbon through a spectacular diversity of ecological relationships: symbiosis *sensu lato*, phagotrophy, mixotrophy). Dinoflagellates have a unique and phenomenal diversity of morphology, physiology, and complex behaviors that allow them to adapt to aquatic marine and freshwater environments. This bio-complexity has made it very difficult to perform integrative studies across the entire group and to establish general ecological or adaptive strategies. The main objective of this thesis was to search for global biodiversity patterns and structure of planktonic dinoflagellate communities across the world's oceans. We first participated in the development of a morphogenetic sampling protocol for all planktonic dinoflagellates, covering their entire spectrum of cell size from 3 to 500 $\mu\text{m}$ , and a significant portion of their spatio-temporal niches in the photic zone of global oceans (Not et al., submitted). This homogeneous and robust protocol was designed to operate on a sailboat, and was applied to each of the 154 sampling points of the circum-global Tara-Oceans planktonic ecosystems sampling project (2009-2012) (Karsenti et al., 2011). The first part of this work was to not only implement the new protocol directly on the boat, but also to explore various high-throughput analyses of the biodiversity of pelagic marine dinoflagellates within Tara-Oceans samples. A first approach, based on the automatic FlowCAM imaging technology carried aboard Tara, allowed assessment of diversity, relative abundance, and geographical distribution of dinoflagellate morpho-species within the genus *Neoceratium* from the Mediterranean Sea (Le Bescot et al., in prep). A second approach was the use of environmental ribosomal DNA fragment ("metabarcodes"), amplified from nucleic acids extracted from seawater, as markers of dinoflagellate biodiversity. The building of a reference database of dinoflagellate DNA barcodes (*DinR2* — Dinoflagellate Ribosomal Reference database), incorporated into a broader database for all marine protists (*PR2* — Protist Ribosomal Reference database) (Guillou et al, 2013), allowed us to assign a traditional taxonomy to meta-barcodes, thus linking classical knowledge of dinoflagellate biodiversity to our analysis based on high-throughput sequencing. We then explored whether general patterns in the structure of dinoflagellate biodiversity emerged across size classes and organismal and ecological niches (Le Bescot et al., submitted). The abundance of different orders of dinoflagellates varies primarily between size classes (pico-, nano-, micro- and meso-plankton), independently of the ecosystem studied and the sampling period. A remarkable and unexpected diversity of pico-dinoflagellates ( $< 5\mu\text{m}$ ) was identified in the order Gymnodiniales, with a large number of barcodes remaining to be identified. The meso-plankton fraction (180-2,000 $\mu\text{m}$ ), meanwhile, contained a high diversity and abundance of parasitic and photo-symbiotic genera of dinoflagellates, which play a key functional role in the maintenance of plankton communities (Le Bescot et al., submitted). Finally, a preliminary study of the impact of environmental conditions on the meta-community of dinoflagellates identified factors affecting the composition of dinoflagellate communities and their functional structure between the photic zone and the meso-pelagic zone (Le Bescot et al., in prep).

**Keywords:** plankton, dinoflagellate, meta-barcode, biodiversity, molecular ecology, global ocean, Tara-Oceans



*Something hidden, go and find it  
Go and look behind the ranges  
Something lost behind the ranges  
Lost and waiting for you....Go!*

*Rudyard Kipling, "The Explorer" (1898)*



## Préambule

L'océan nous semble immense, pourtant lorsque l'on parle des mers et de leurs habitants notre imagination s'arrête à des baleines, des dauphins, des phoques et des poissons multicolores, des espèces pour beaucoup commerciales dont la présence est visible et l'exploitation parfois préoccupante. Mais arrêtons-nous un instant et imaginons une autre dimension, une dimension plus vaste, masquée aux premiers regards, qui comme l'univers serait beaucoup plus complexe et riche, avec ses galaxies, ses astéroïdes, ses planètes. Imaginons un monde invisible, un monde de l'infiniment petit. On a tous entendu parler des « poussières d'étoiles », et bien il existe aussi des « poussières des mers ». Des « poussières » qui, vues au microscope, prennent des allures de créatures fantastiques toutes droites sorties des meilleurs romans de science-fiction. Des milliards d'entre elles errent dans les océans, à la fois essentielles et éphémères, translucides et frêles, d'une incroyable richesse de tailles et de formes, elles sont comme une ébauche de la vie. Sans ces fragiles « poussières », dont l'existence ne tient qu'à quelques dixièmes de degrés, à une variation de sels minéraux ou de lumière, la vie marine n'existerait pas. Qu'elles soient végétales ou animales, elles constituent ce que l'on appelle le Plancton.

Ce petit peuple de la mer s'agite, frémit, palpète, bat des cils, remue du flagelle, dessine des arabesques : le plancton est beau, il vit. Avec lui, deux petites gouttes d'eau ne se ressemblent jamais. Chacune d'elle contient un concentré de vie, une « société » organisée et structurée de manière unique. Les divers planctons s'allient ou se livrent des guerres incroyables dans un univers immense et invisible.

Ces microorganismes, bien que représentant une biomasse quasi invisible, sont des acteurs clefs dans les réseaux trophiques des océans et les cycles biogéochimiques globaux. Présents depuis des millions d'années dans le plus vaste, mais aussi probablement le moins bien exploré des espaces de notre planète, nous n'en connaissons presque rien. La Terre a environ 4,6 milliards d'années, et est depuis 4,3 milliards d'années recouverte majoritairement d'un film persistant d'eau liquide (Watson & Harrison 2005). La Terre dispose de six éléments chimiques majeurs, l'hydrogène, le carbone, l'oxygène, l'azote, le phosphore, le soufre qui forment les ingrédients nécessaires à la vie. Ces éléments sont les constituants de base des biopolymères majeurs comme les protéines, les lipides, les polysaccharides ou les acides nucléiques, qui servent d'architecture à la Vie. Ces biopolymères se forment par le transfert d'électrons ou de protons (redox), des processus de réduction/oxydation qui forment l'épine dorsale de la chimie biologique (Falkowski & Knoll 2007). Pour réduire le carbone inorganique en matière organique, la photosynthèse est le moyen le plus efficace. Les organismes capables de cet exploit métabolique sont capables de fournir de la matière organique à tous les autres



organismes de l'écosystème, on les appelle les «producteurs primaires». La distribution de la photosynthèse sur l'arbre de la vie indique une histoire complexe et une origine dans les océans (les enregistrements géologiques le confirment). La structure actuelle, à la fois écologique et phylogénétique de l'autotrophie dans les écosystèmes marins, est née il y a environ 200 millions d'années (Falkowski & Knoll 2007), pour progressivement se complexifier au cours du temps (annexe I) faisant place aujourd'hui à une structure très diversifiée (hétérotrophie, mixotrophie) dérivée des processus évolutifs.

Si les grands processus physico-chimiques marins sont assez largement étudiés, l'approche biologique des océans n'est encore qu'à ses débuts. La taille microscopique de ces micro-organismes marins, leur distribution, en ont fait le parent pauvre de l'étude de la biodiversité marine. Pourtant, notre compréhension de leur évolution, de leur fonctionnement est sur le point d'être aujourd'hui considérablement enrichie par l'apparition de nouvelles techniques de biologie moléculaire à très haut débit et des méthodes d'imagerie optique toujours plus précises, rapides et bon marché.

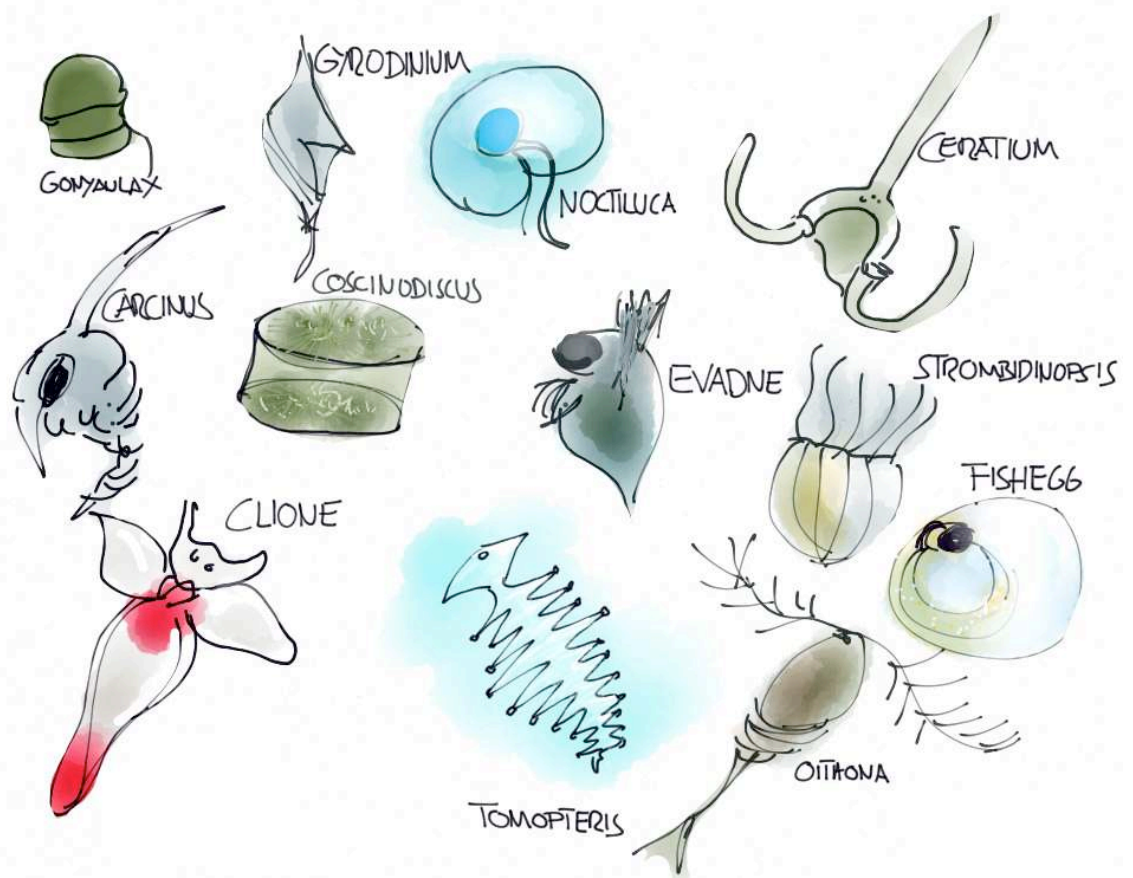
Ce travail de thèse aborde l'étude de cet environnement de l'infiniment petit marin à travers les techniques d'aujourd'hui. Depuis l'échantillonnage, jusqu'à l'analyse d'une population au rôle majeur dans l'équilibre océanique je vous propose d'entamer une exploration biologique de l'Océan mondial.

*« Né de comparaisons très anciennes faites entre la marche des planètes, apparemment errantes, à travers le dessin des constellations célestes et la flottaison des créatures vivantes à travers la mer, le terme plancton est entré de nos jours dans le vocabulaire courant. Il ne définit pas les êtres dont il suggère la présence, mais il en caractérise le comportement puisqu'il évoque un état physique, l'errance, que nous plaçons dans un milieu connu, l'océan. Ainsi, toute créature qui vit et flotte sans avoir une vigueur de nage volontairement et énergiquement dirigée fait partie de ce plancton. »*

*Anita Conti, océanographe*



# Introduction



# Sommaire

|  |           |
|--|-----------|
| <b>Introduction</b>  | <b>17</b> |
| <b>1. <i>Le plancton</i></b>                                       | <b>17</b> |
| 1.1. La classification phylogénétique des espèces                  | 19        |
| 1.2. Les Alvéolés  | 21        |
| <b>2. <i>La diversité des dinoflagellés</i></b>                    | <b>21</b> |
| 2.1. Biologie : description générale                               | 21        |
| 2.2. Reproduction  | 24        |
| 2.3. Classification  | 26        |
| 2.4. Biogéographie   | 28        |
| 2.5. Écologie  | 28        |
| 2.6. Comment étudier la micro-biodiversité marine ?                | 29        |
| <b>3. <i>Les code-barres génétiques</i></b>                        | <b>30</b> |
| 3.1. Le concept  | 30        |
| 3.2. Développement du barcoding                                    | 31        |
| 3.3. Le barcoding en milieu marin                                  | 32        |
| 3.4. Le barcoding chez les protistes                               | 33        |
| <b>4. <i>Les code-barres génétiques chez les dinoflagellés</i></b> | <b>34</b> |
| <b>Objectifs de la thèse et structure du document</b>              | <b>37</b> |
| <b>Références</b>  | <b>39</b> |

## Introduction

Le littoral s'enfonce sous les flots pour créer un paysage en relief, sorte de miroir de l'espace terrestre, jusqu'aux abysses. Dans cet univers en trois dimensions de l'océan, la vie est intense, les organismes sont d'une grande diversité, parfois étonnants, et se livrent des luttes incessantes. Les écologistes l'ont comparé à un immense écosystème où les espèces aquatiques sont classées en fonction de leur mode de vie, distinguant le benthos du pelagos. Le benthos regroupe l'ensemble des organismes aquatiques qui colonisent le plancher océanique, des côtes jusqu'aux abysses. Le pelagos désigne lui l'ensemble des organismes qui occupent la tranche d'eau supérieure, depuis le fond jusqu'à la surface. Il est constitué d'organismes pouvant vaincre les courants marins, appelés necton, et d'autres trop peu vigoureux pour s'y opposer, qui constituent le plancton. Le necton correspond aux animaux nageurs, qu'ils soient de l'immense variété des poissons, des cétacés et autres mammifères marins, des tortues ou encore de certains mollusques comme le calmar.

### 1. Le plancton

Le mot « plancton » vient du grec ancien *planktós* signifiant « errant ». Il désignait chez Homère les animaux errants à la surface des flots. Identifiés dès la première moitié du XIXe siècle, d'un côté en Irlande par Charles Wyville Thomson et d'un autre en Allemagne par Johannes Müller, ce n'est qu'en 1887 qu'apparut pour la première fois dans la littérature le terme de plancton, créé par l'océanographe allemand Viktor Hensen pour désigner « tout ce qui flotte dans les eaux », imaginant que les plus petits nourrissent les plus grands. Ce sens primitif s'est, depuis, considérablement restreint. On entend aujourd'hui par plancton « ce qui est vivant et qui flotte passivement dans les eaux », réservant le nom de seston au « matériel flottant non vivant ». Le terme de plancton couvre malgré cela encore une réalité complexe du point de vue taxonomique, dimensionnel, et de sa distribution bathymétrique et géographique.

Si les mouvements propres (quand ils existent) de ces petits organismes n'ont pas une ampleur suffisante pour leur permettre de surmonter ceux des eaux qui les portent, ils sont toutefois capables dans certains cas de déplacements verticaux. On observe ainsi une ascension du plancton à la nuit tombée et un mouvement inverse à l'aube. Ces oscillations pouvant atteindre plusieurs centaines de mètres d'amplitude (Ochoa *et al.* 2013). De natures très diverses, ces organismes ont des tailles variant du dixième de micromètre à plusieurs centimètres, voire plusieurs mètres pour les plus grands (comme les méduses). Ces différences de taille permettent à chaque espèce planctophage de trouver de la nourriture à la taille de sa « bouche ». Ce monde planctonique peut ainsi être classé selon cinq grands groupes d'organismes : les virus, les bactéries, les archées, les protistes et les métazoaires. Chacune de ces catégories représentant une taille, des modes de vie, une fonction qui leurs sont propres (Fig. 1).

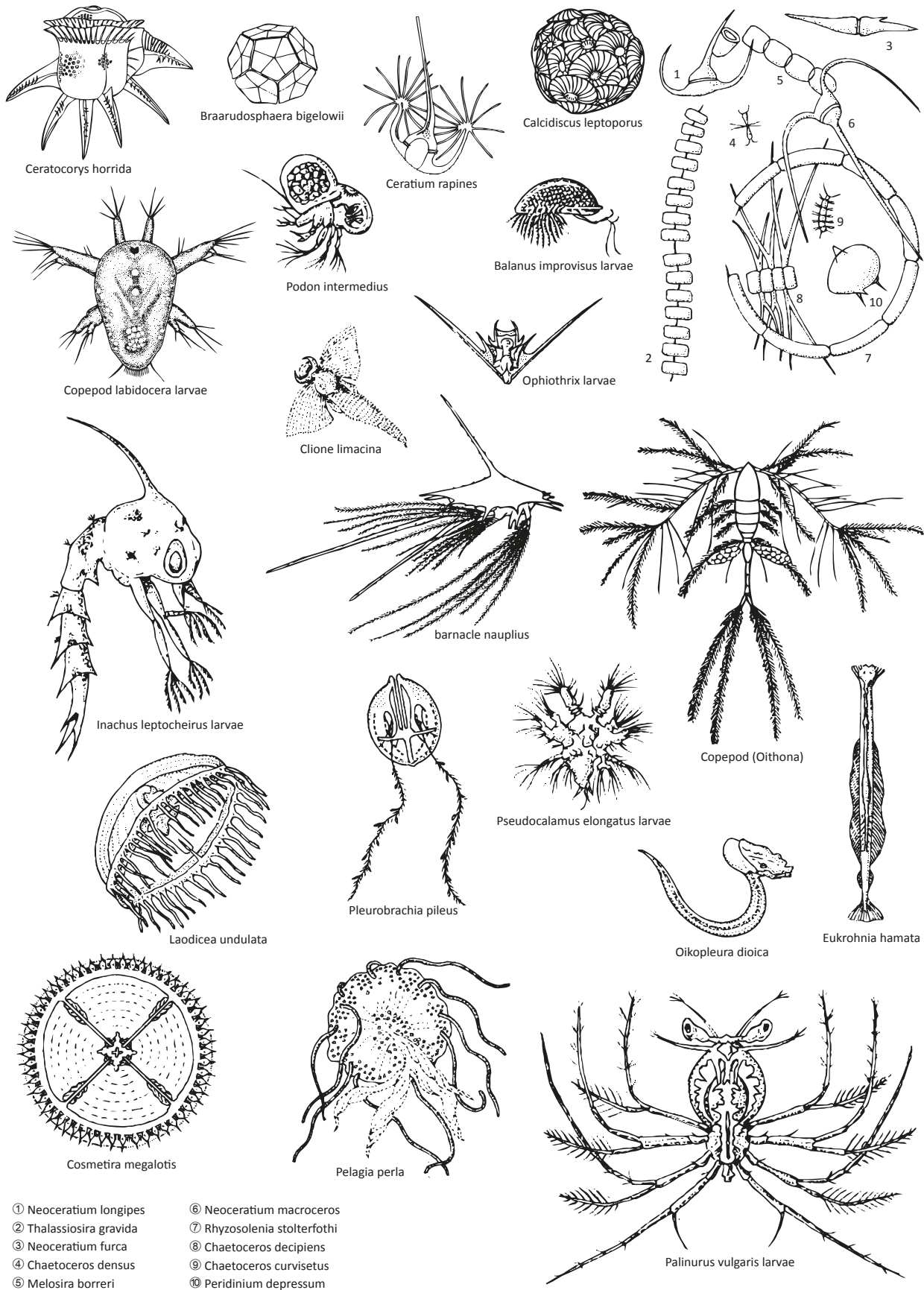


Figure 1 : le plancton.

Omniprésent dans les océans, le plancton est un acteur majeur des cycles biogéochimiques de notre planète : le carburant (Arrigo 2004; Martin & Quigg 2013). La photosynthèse et le cycle du carbone, deux des principaux cycles biogéochimiques connus, sont assurés dans les océans par ces microbes océaniques produisant à eux seuls presque autant d'oxygène que les plantes terrestres (Field *et al.* 1998). Les virus modulent les processus photosynthétiques en provoquant la mortalité de populations entières, ou en agissant comme régulateurs sur les gènes exprimés par leurs hôtes (Lindell *et al.* 2005). Les protistes photosynthétiques et leurs congénères brouteurs, les herbivores des océans (autres protistes et métazoaires), modulent également la photosynthèse globale de multiples façons.

Tous restent pourtant largement inconnus et leurs fonctions largement inexplorées.

La principale difficulté dans l'étude de la biodiversité en milieu marin c'est l'accès limité à certains habitats (vaste étendue océanique), à certaines communautés (micro-organismes) et de la répartition inégale des espèces dans le milieu (oasis et déserts). L'identification de ces espèces est pourtant un point de départ critique pour tout type de recherche en biologie marine. Jusqu'à il y a peu, la description de nouvelles espèces en microbiologie reposait en notre capacité à les isoler et à les mettre en culture. Cependant, aujourd'hui, le développement de nouvelles techniques moléculaires, non dépendantes de cette mise en culture, permet une profonde réévaluation de la diversité microbienne au sein des écosystèmes naturels (Venter *et al.* 2004). Plusieurs publications ont par exemple décrit la puissance des séquences ribosomales dans l'identification des micro-organismes procaryotes et eucaryotes (Amann *et al.* 1990; Simon *et al.* 2000; Groben *et al.* 2004). Le clonage direct et le séquençage de la petite sous-unité 18 de l'ADN ribosomal (ADNr 18S) à partir d'échantillons naturels a, entre autres, permis une vision plus large de la structure et de la composition des communautés de pico-plancton (Giovannoni *et al.* 1990; Lopez-Garcia *et al.* 2001; Metfies *et al.* 2006).

### *1.1. La classification phylogénétique des espèces*

Pendant plus de la moitié de l'histoire de la Vie sur Terre, tous les organismes étaient de taille microscopique, c'était des bactéries unicellulaires. Aussi appelées procaryotes, ces cellules ne possédaient pas de membranes internes (à l'exception de l'invagination de la membrane externe chez les bactéries photosynthétiques). Par la suite des cellules plus complexes contenant des membranes ont fait leur apparition. Cette étape qui sépare les cellules procaryotes, relativement simples, des cellules très complexes est une des plus importantes de l'évolution de la Vie, c'est l'apparition d'un nouveau type d'organisme : les eucaryotes.

Les eucaryotes sont des organismes possédant un noyau et d'autres organites délimités par des membranes. On pense que le réticulum endoplasmique et la membrane nucléaire ont évolué à partir de l'invagination de la membrane externe. Selon la théorie endosymbiotique (Delwiche 1999),



les mitochondries et les chloroplastes ont évolué à partir de bactéries ingérées qui sont restées intactes. Les mitochondries et les chloroplastes ont ainsi leur propre ADN qui ressemble à celui des procaryotes.

Cinq grands groupes sont caractérisés au sein des eucaryotes (Fig. 2) : les opisthocontes (Opisthokonta) qui regroupent des organismes apparemment très divers comme les champignons et les métazoaires ; les amibozoaires (Amoebozoa) qui constituent un grand groupe de protozoaires simples ; les Excavés (Excavata) qui rassemblent une grande variété de cellules libres et symbiotiques dont certaines sont des parasites pour l'Homme ; les Archaeplastida (ou lignée verte) qui regroupent des êtres vivants capables de faire de la photosynthèse (ils possèdent des chloroplastes) ; enfin le SAR qui est un groupe qui contient les Stramenopiles, les Alveolata et les Rhizaria.

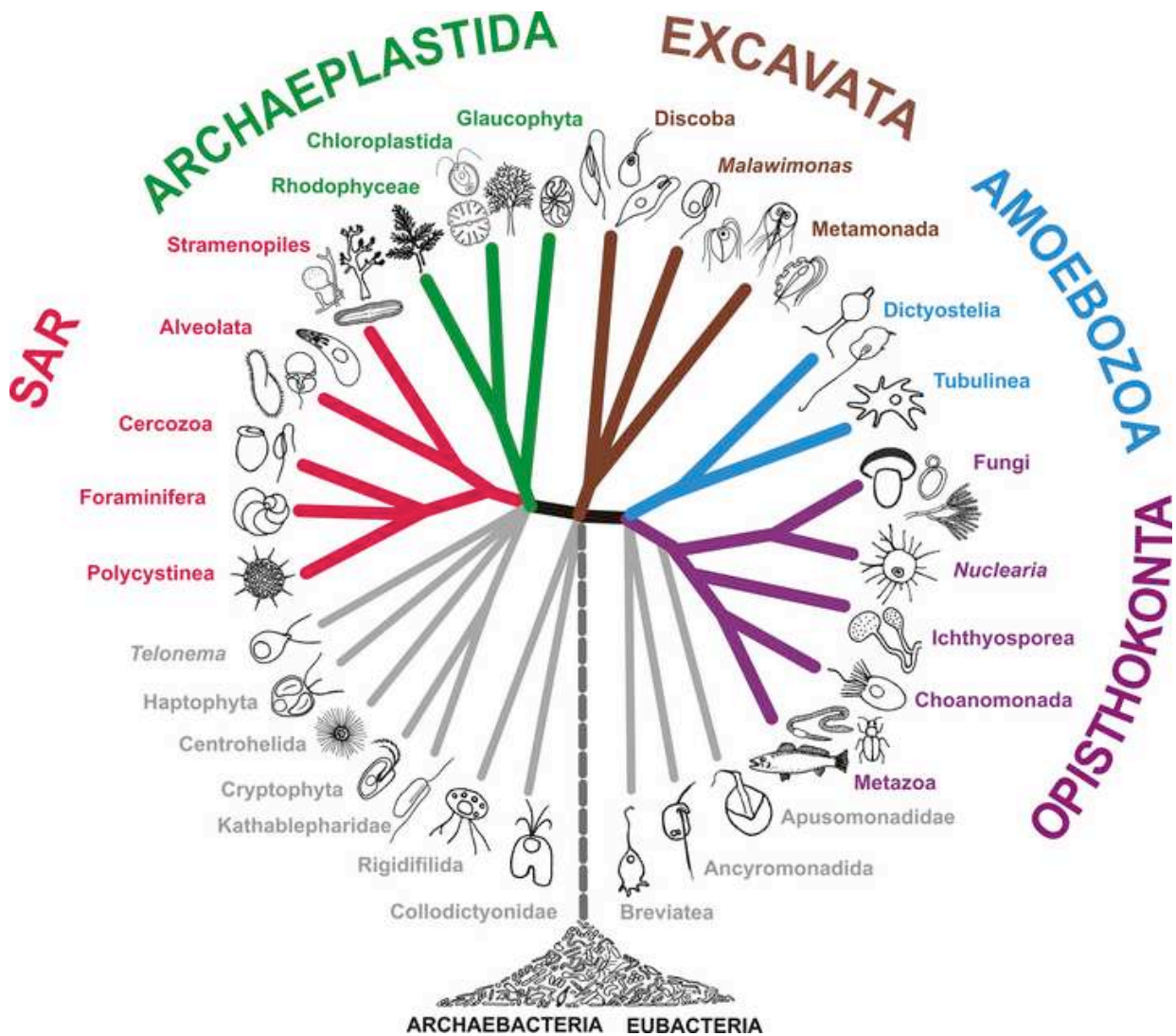


Figure 2 : Phylogénie des eucaryotes. D'après (Adl *et al.* 2012)

## 1.2. Les Alvéolés

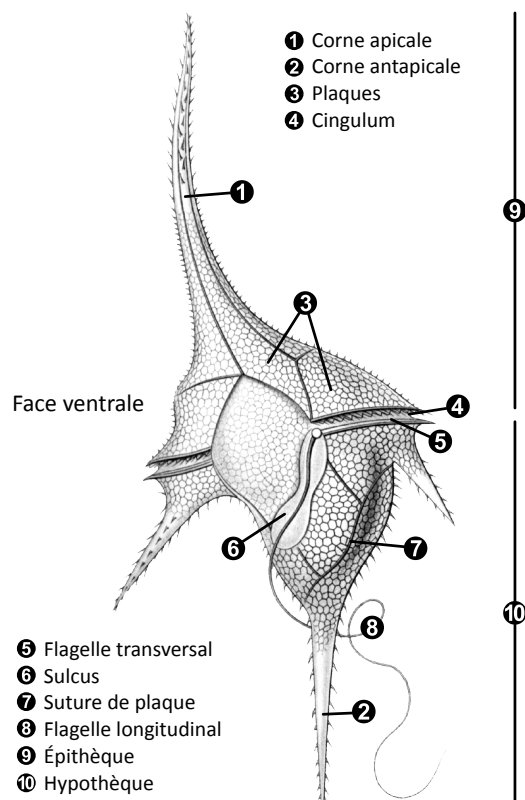
Les alvéolés (Alveolata) constituent un groupe hétérogène des eucaryotes unicellulaires (ou protistes). Présents dans les écosystèmes marins et terrestres leur principale caractéristique morphologique commune est la présence d'alvéoles corticales. Ces vésicules aplaties, emballées par une membrane en une couche continue, forment une pellicule ou « alvéole » flexible (Cavalier-Smith & Chao 2004). Les alvéolés présentent des stratégies trophiques très diverses, comme la prédation, la photo-autotrophie ou la symbiose (le parasitisme, le mutualisme, le commensalisme). La plupart d'entre eux se regroupent en trois phylums principaux : les ciliés, les dinoflagellés et les apicomplexa, partageant tous un ancêtre commun (Leander & Keeling 2003). D'autres lignées d'alvéolés marins, MALV (Marine Alveolates), ont été identifiées en culture et associées à la classe des Syndiniales (Syndinea). Ces derniers sont divisés en deux groupes principaux, MALV-I et MALV-II (Massana & Pedrós-Alió 2008). Les apicomplexa sont des parasites des cellules animales (métazoaires), par exemple *Plasmodium* qui cause la malaria. Les ciliés sont principalement des prédateurs aquatiques, même si certains taxons sont parasitaires. Les dinoflagellés sont des photo-autotrophes ou des hétérotrophes (des prédateurs vivant en liberté), ou les deux, soit simultanément, soit alternativement (mixotrophes), ou encore des parasites ou symbiotes (Taylor *et al.* 2007). Les dinoflagellés et les apicomplexa sont les seuls groupes d'alvéolés à posséder des plastes. Les plastes sont activement impliqués dans les processus de photosynthèse chez les dinoflagellés alors qu'ils sont résiduels chez les apicomplexa (McFadden *et al.* 1996). Ce caractère commun entre les deux groupes suggère que leur ancêtre commun était photosynthétique (Leander & Keeling 2003). Contrairement aux autres alvéolés, certains dinoflagellés sont également capables de fabriquer de la cellulose présente sous la forme de plaques de polysaccharides. L'agencement et la décoration de ces plaques fournissent une étonnante variété de formes ce qui constitue le fondement traditionnel de la classification des espèces de dinoflagellés.

## 2. La diversité des dinoflagellés

### 2.1. Biologie : description générale

Les dinoflagellés sont des micro-organismes planctoniques, aux trois quarts marins, le reste vivant en eau douce (Foissner & Hawksworth 2009). Le nombre d'espèces de dinoflagellés actuellement vivants est estimé à environ 2 400 (Gómez 2012; 2013), avec 2 500 espèces fossiles nommées (Taylor *et al.* 2007). Récemment, Gómez (Gómez 2012) a reconnu 1 555 espèces de dinoflagellés marins libres. Environ la moitié des espèces de dinoflagellés modernes est photosynthétique, l'autre étant hétérotrophe, ou présentant les deux modes de nutrition (Gómez 2012). La plupart des espèces photosynthétiques contiennent de la chlorophylle *a* et *c*<sub>2</sub>, le caroténoïde

bêta-carotène, et un groupe de xanthophylles qui leur semble caractéristique, la péricidine, la dinoxanthine et la diadinoxanthine (Lee 2007). Certains dinoflagellés ont également acquis d'autres pigments à travers des processus endosymbiotiques tertiaires, comme la fucoxanthine. Chez ces espèces photosynthétiques, les plastes ont la particularité d'avoir une enveloppe composée de 3 membranes. Les mitochondries présentes dans chaque cellule se caractérisent par de courtes crêtes tubulaires.

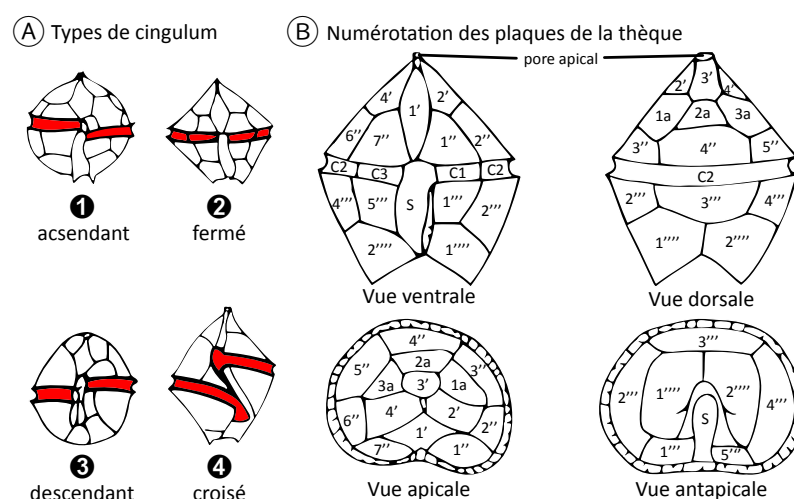


**Figure 3 :** Principaux caractères morphologiques du stade mobile d'un dinoflagellé type.  
Annoté depuis un dessin d'Ernst Haeckel.

Les dinoflagellés sont unicellulaires, bien que certains puissent former des chaînes ou des pseudo-colonies. Leur noyau, appelé dinocaryon, est très caractéristique, car les chromosomes qu'il contient sont généralement nombreux et restent condensés à tous les stades cellulaires. Le stade mobile de leur cycle de vie est haploïde, et la cellule possède alors en général deux flagelles dissemblables. Un flagelle en forme de ruban, situé dans une rainure transversale (cingulum), et un flagelle plus conventionnel émergeant d'un sillon ventral (sulcus) qui bat à l'arrière de la cellule (Fensome *et al.* 1998). La cellule est entourée par une membrane extérieure continue (amphiesma) dans laquelle se trouvent les vésicules amphiesmales. Certains dinoflagellés sont recouverts de plusieurs plaques polysaccharidiques formant une thèque plus ou moins épaisse, alors que d'autres ont des parois fragiles et sont appelés dinoflagellés nus, ou sans thèque. La cellule est divisée en deux

parties, l'épithèque et l'hypothèque, séparées par le sillon équatorial (cingulum). L'extrémité antérieure de la cellule est appelée apex alors que l'extrémité postérieure est appelée antapex (Fig. 3).

Plusieurs conventions régissent l'étude de la morphologie chez les dinoflagellés dont la numérotation du côté gauche et droit de la cellule (Fig. 4B). La vue ventrale de la cellule est déterminée par la présence du sulcus et de la jonction de la ceinture à ce dernier; la vue dorsale lui est opposée. Le cingulum est fermé si ses deux extrémités se rejoignent à la même hauteur, ou ouvert si ses extrémités sont décalées. Le cingulum peut être descendant si son extrémité est plus postérieure que son origine, et ascendant si elle est plus antérieure. Le cingulum peut aussi être croisé lorsque son extrémité est très déplacée et se retrouve à la gauche de son origine (Fig. 4A). Il se présente enfin soit au niveau avec le reste de la thèque (planizone) ou bien creusé en gouttière (cavizone), et bordé de crêtes ou d'ailettes. La classification repose donc surtout sur la morphologie externe, et pour les cellules avec thèque, sur l'étude de la tabulation, c'est-à-dire le nombre et l'arrangement des plaques (Fig. 4B). Les plaques sont numérotées comme suit : apicales ('), intercalaires apicales (a), pré-cingulaires (''), cingulaires (c), post-cingulaires (''''), intercalaires postérieures (p), antapicales (''''). La région du pore apical ou complexe du pore apical («APC») comprend 3 petites plaques, la plaque du pore (Po), la plaque qui recouvre en partie cette ouverture (cp «closing plate») et la plaquette qui rejoint la plaque du pore à la première plaque apicale (X, «canal plate»). Les plaques sulcales (s) qui bordent le sulcus sont très difficiles à observer et comprennent principalement une plaque sulcale antérieure et une postérieure, une gauche et une droite. La numérotation se fait à partir de la plaque la plus ventrale, dans le sens contraire des aiguilles d'une montre, de gauche à droite en suivant le flagelle transversal.



**Figure 4** : Classification Kofoid de la phase mobile d'un dinoflagellé : cellule Peridinal en vue ventrale et dorsale, annotée selon la méthode Kofoidienne. Adapté depuis Dodge et al. 1982. (A) Le cingulum débute à gauche et se termine à droite de la cellule. (B) Les plaques sont numérotées comme suit : apicales ('), intercalaires apicales (a), cingulaires (C), pré-cingulaires (''), post-cingulaires (''''), antapicales ('''').

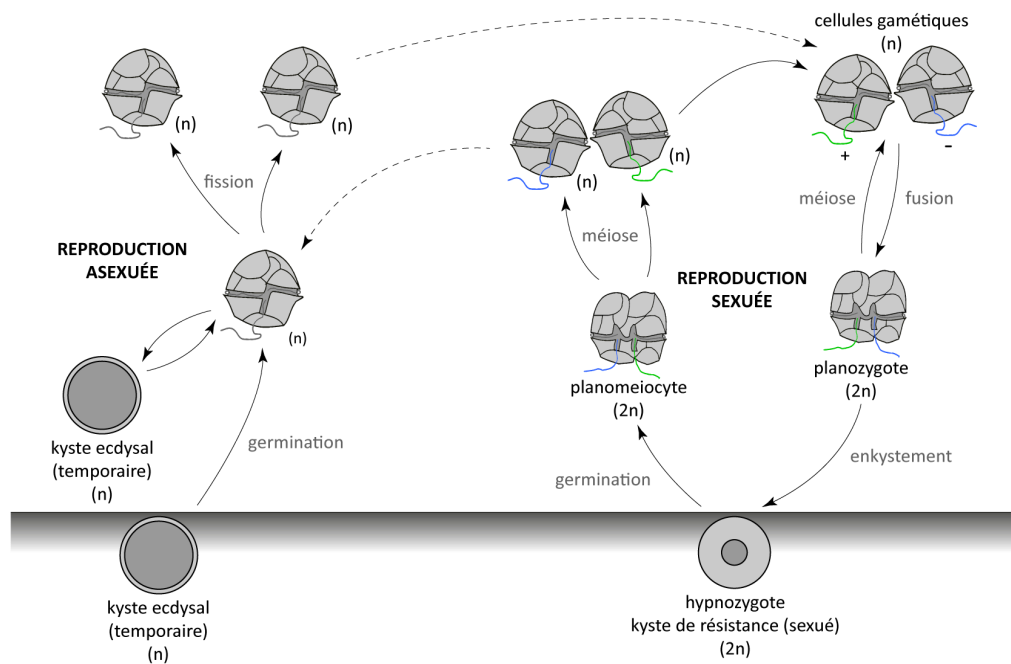
De façon générale il est possible de trouver de petites cellules sphériques, alors que d'autres ont des structures complexes qui ressemblent à des cornes, des ailes ou des colliers.

Enfin, certains dinoflagellés sont toxiques pour l'homme et les autres organismes marins. Certaines espèces de dinoflagellés, dont celles produisant des toxines, peuvent former de grandes efflorescences ou blooms (les «marées colorées») qui peuvent fortement impacter l'activité économique dans les zones côtières (Hallegraeff 2003).

### 2.2. Reproduction

Après une recombinaison sexuée, les dinoflagellés produisent des stades (diploïdes) benthiques résistants (kystes), ce qui leur permet d'échapper à la prédation ou à des conditions environnementales défavorables. C'est également un excellent mode de colonisation des écosystèmes. Les dinoflagellés sont des espèces dites « haplontiques » ce qui signifie que la plus grande partie de leur vie se fait au stade haploïde (Litaker *et al.* 2002; Figueroa *et al.* 2006). Cependant, même s'ils se reproduisent habituellement par voie végétative (notamment pendant les efflorescences), les deux modes de reproduction, asexuée et sexuée, peuvent être observés (Fig. 5).

Dans le cas d'une reproduction asexuée, une cellule se divise par fission binaire pour donner deux cellules filles identiques à la cellule mère (c'est le cas lors d'efflorescences). En cas de stress, ils peuvent former des kystes temporaires (ou kyste ecdysal) (Fistarol *et al.* 2004). Ces kystes temporaires apparaissent par exemple lors d'une carence en nutriments (Doucette *et al.* 1989), d'un changement de température (Grzebyk & Berland 1996) ou encore lors d'interactions allélopathiques (Fistarol *et al.* 2004). Ces kystes à membrane fine peuvent survivre pendant un temps limité, par exemple dans le cas de *Scrippsiella trochoidea*, la période de dormance de ces kystes temporaires est de 6 à 13 jours (Wang *et al.* 2007). La paroi de ces kystes semble être faiblement perméable (Garcés *et al.* 1998) et peut résister à la digestion lors de leur consommation par des bivalves (Laabir *et al.* 2007). Une modification du contenu en toxines de ces kystes par rapport aux cellules végétatives a aussi été observée (Persson *et al.* 2006). Lorsque les conditions redeviennent favorables, ces kystes peuvent très rapidement germer et retourner vers une multiplication végétative active.



**Figure 5 :** Cycle de reproduction d'un dinoflagellé type.

La reproduction sexuée fait intervenir plusieurs stades cellulaires intermédiaires. Elle est initiée par la fusion de deux cellules gamétiques. Il ne s'agit pas d'une reproduction sexuée au sens strict (dans le sens où il n'y a pas émission de « gamètes »), mais de la fusion de deux cellules appelées + et -. Certaines espèces de dinoflagellés sont homothalliques, c'est-à-dire que toutes les combinaisons de cellules gamétiques sont équivalentes et peuvent former un zygote diploïde. Chez les espèces hétérothalliques, la fusion ne peut se faire que si les deux cellules gamétiques sont de types sexuels différents (Pfiester & Anderson 1987). Dans d'autres cas, il s'agit même d'hétérothallisme complexe, avec la présence de plus de deux types sexuels. La fusion des gamètes forme un zygote diploïde mobile appelé planozygote qui possède deux flagelles. Chez la majorité des dinoflagellés, ce planozygote perd ses flagelles après plusieurs heures voire plusieurs semaines de vie et s'arrondit pour former un zygote diploïde non mobile appelé hypnozygote (Pfiester & Anderson 1987). Cet hypnozygote va former un kyste de résistance, dit de long terme, capable de rester viable plusieurs mois ou années avant de germer (Lee 2007). Après une période de dormance, l'hypnozygote germe et produit un planoméioocyte qui initie une méiose. Chez certaines espèces de dinoflagellés, le planozygote peut directement produire des cellules végétatives sans passer par la formation d'un hypnozygote (Figueroa *et al.* 2010).

### 2.3. Classification

Les dinoflagellés sont classés selon neuf grands ordres : les Gonyaulacales, les Peridinales, les Gymnodinales, les Suessiales, les Prorocentrales, les Dinophysiales, les Blastodinales, les Phytodinales et les Noctilucales (Fensome *et al.* 1993). Chaque ordre se définit par les principaux caractères morphologiques et les caractéristiques du cycle de vie de ses membres (Fig. 6).

Les Gonyaulacales et les Peridinales se caractérisent par la présence de plaques thécales en cellulose situées dans les alvéoles corticales. Les thèques de ces deux groupes sont constituées de cinq séries de plaques latitudinales (apicale, intercalaire antérieure, précingulaire, post-cingulaire et antapicale), d'un cingulum et d'un sulcus. Les membres de l'ordre des Blastodinales ont un mode de vie parasitaire et un dinokaryon temporaire (Saldarriaga *et al.* 2004). Les Prorocentrales et les Dinophysiales partagent un caractère synapomorphe majeur et unique chez les dinoflagellés : une division de la thèque en deux moitiés latérales reliées par une suture sagittale. L'ordre des Phytodinales regroupe des espèces caractérisées par le passage d'une cellule coccoïde non calcaire à un stade végétatif. Cet ordre est encore aujourd'hui largement méconnu (avec des genres pour lesquels très peu de données moléculaires sont disponibles). L'ordre des Noctilucales rassemble des dinoflagellés caractérisés par une grande mobilité de leur flagelle ventral. Ce flagelle ne joue pas un rôle dans le maintien de la cellule en suspension, mais semble plutôt être lié à la capture de nourriture. Les Noctilucales ont la capacité d'intégrer, de remplacer ou de perdre des chloroplastes, un phénomène rare dans les autres groupes d'alvéolés (Not *et al.* 2012).

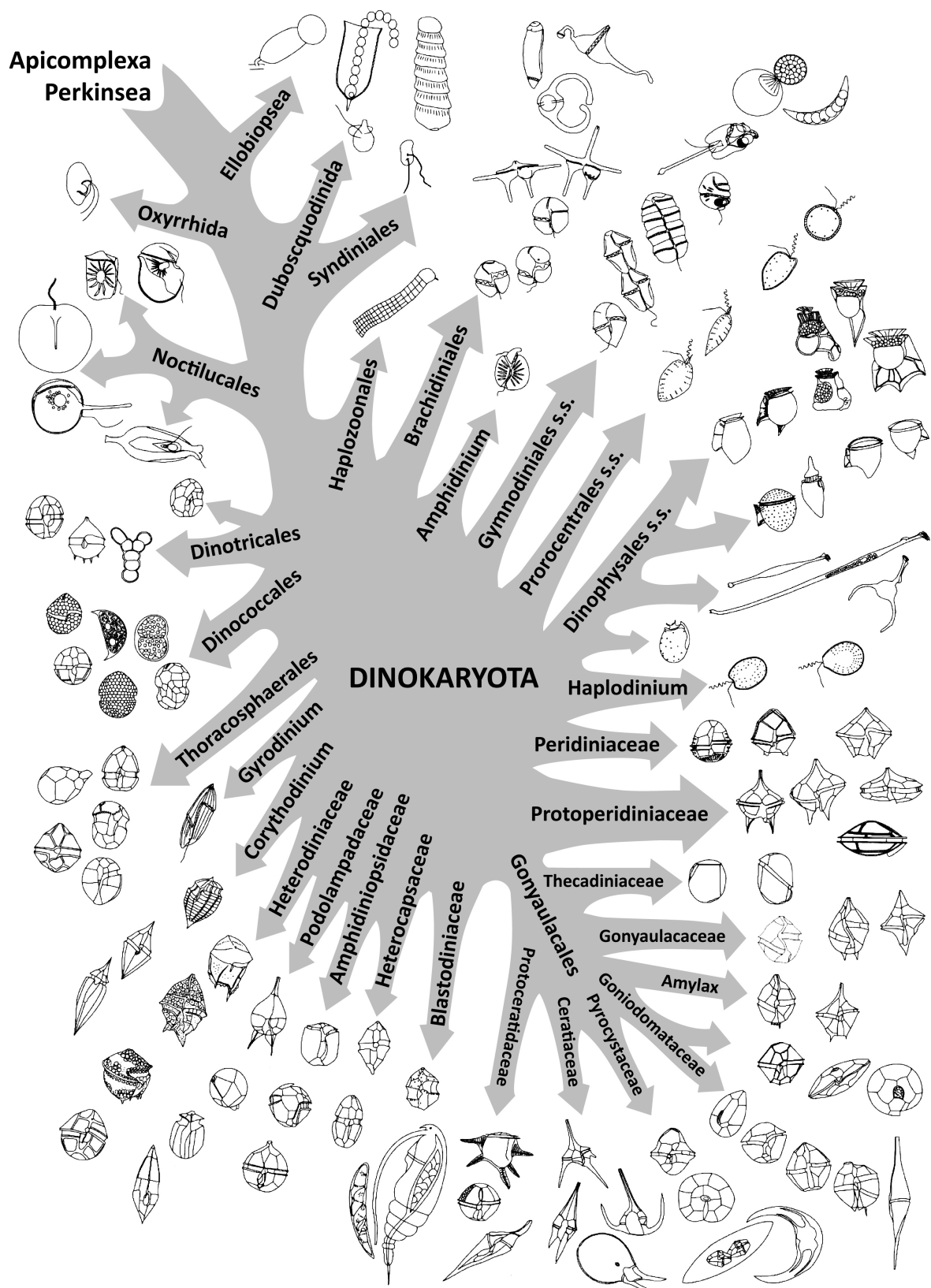


Figure 6 : La diversité des principales lignées de dinoflagellés. Adapté depuis Gomez 2012.



### 2.4. Biogéographie

Les dinoflagellés ont une distribution spatiale similaire à d'autres protistes aquatiques de taille similaire (Dolan 2005). Les morpho-espèces équivalentes sont présentes dans des zones climatiques similaires des deux hémisphères nord et sud. Il est ainsi admis que toute la communauté microplanctonique des eaux océaniques tempérées du Nord et du Sud sont pratiquement identiques (Taylor 2001; 2004). Dans les régions océaniques tropicales, les dinoflagellés forment le groupe dominant des protistes et présentent une très grande diversité (Taylor *et al.* 2007). Leur forte richesse spécifique ainsi que la diversité de leur morphologie et de leur physiologie témoignent de leur habilité à s'adapter et à coloniser une large gamme d'habitats. Cette plasticité rend impossible l'identification d'une stratégie adaptative ou d'une écologie commune à l'ensemble du groupe. C'est pourquoi cette diversité des dinoflagellés peut et doit être modérée par la définition des différents concepts d'espèces car ceux-ci ont une forte incidence sur la prise en compte de la biogéographie (Lundholm & Moestrup 2006).

### 2.5. Écologie

Une forte concentration en dinoflagellés correspond à de forts apports en nutriments issus des environnements terrestres et/ou à des remontées d'eau profonde. Les turbulences diminuent leur capacité de nage en provoquant la rétractation de leurs flagelles et entraîne une sédimentation, voire une agrégation des cellules (c'est le cas par exemple chez *Neoceratium tripos*) (Havskum *et al.* 2005). Leurs flagelles leur permettent cependant d'optimiser leur localisation et d'éviter de sédimenter lorsque l'environnement est très stable (non turbulent). Grâce à l'orientation perpendiculaire de ces derniers, qui présentent un mouvement dirigé, ils sont capables de répondre à des stimuli chimiques, ou à des variations physiques comme la gravité ou la lumière. Cette mobilité leur offre la capacité de trouver et de maintenir des conditions optimales à leur croissance et à leur survie en s'adaptant à une diminution de la lumière ou à une limitation en nutriments. Les dinoflagellés ont des préférences différentes en matière d'habitat, leurs stratégies de vie sont multiples et leurs modes nutritionnels polyvalents (Gómez 2012), ce qui en fait des organismes très compétitifs pour l'accès aux ressources vis-à-vis des autres groupes de protistes. Les dinoflagellés photosynthétiques sont principalement limités par le phosphore et l'azote (Not *et al.* 2012). Dans certains cas, certaines espèces de dinoflagellés ont développé la capacité à stocker ces éléments nutritifs ce qui leur permet de supplanter les autres espèces (Graham & Wilcox 2000; Labry *et al.* 2008).

En milieu côtier, les blooms observés pourraient être dus à la germination des cellules végétatives d'hypnozygotes. Les facteurs biologiques et/ou physiques qui déclenchent l'initiation de

ces blooms restent mal connus pour la plupart des dinoflagellés (Granéli & Turner 2007), mais leur taux de croissance peut aller jusqu'à 3,5 divisions par jour (Smayda 1997).

Même s'ils peuvent parfois être importants en terme de biomasse, les micros et nano-dinoflagellés sont aujourd'hui rarement considérés comme dominant en terme d'abondance au sein de la fraction phototrophe des communautés planctoniques (Siokou-Frangou *et al.* 2010). Cependant, il est probable que ce soit dû en partie au fait que l'information dont nous disposons est très fragmentaire et que les méthodologies d'identification utilisées jusqu'ici sont insuffisantes, notamment pour les espèces nues (Not *et al.* 2012).

## 2.6. Comment étudier la micro-biodiversité marine ?

Deux approches peuvent être envisagées pour explorer la diversité et l'organisation de la vie microscopique marine. La première, la plus ancienne, consiste à collecter et à concentrer dans des filets ou sur des filtres les micro-organismes pour les observer au microscope. Cette approche, coûteuse en temps, évolue progressivement vers des systèmes automatisés d'imagerie à haut-débit comme le FlowCAM (Sieracki *et al.* 1998) ou le ZooScan (Grosjean *et al.* 2004), embarqués directement sur des engins sous-marin ou des navires océanographiques (Wiebe & Benfield 2001). Ces technologies de pointe sont régulièrement associées à des systèmes de reconnaissance semi-automatiques basés sur des caractères géométriques qui accélèrent le travail de classification réalisé par les taxonomistes (Culverhouse *et al.* 2006). Cependant, cette approche reste liée à l'observation « visuelle » des échantillons et peut présenter des limitations pour la distinction et la caractérisation de certains groupes taxonomiques (Bell & Hopcroft 2008).

Une seconde approche, plus moderne, basée sur la biologie moléculaire permet de combler en partie ce biais. En 2003, Hebert *et al.* (Hebert *et al.* 2003b), ont introduit le concept du code-barres ADN (« DNA barcoding »), et proposé une nouvelle approche pour l'identification des espèces qui offre des possibilités intéressantes. De façon analogue aux code-barres utilisés dans les produits commerciaux, le code-barres génétique fournit un outil standardisé pour une identification rapide, robuste et précise des espèces.

### 3. Les code-barres génétiques

#### 3.1. Le concept

Depuis toujours, les espèces ont été décrites et caractérisées sur la base de critères morphologiques, qui trouvent notamment leurs limites dans les groupes où ils sont peu accessibles ou peu variables, comme c'est le cas pour bon nombre de micro-organismes.

#### **Les concepts d'espèce**

Le concept classique d'espèce se base sur la morphologie, c'est-à-dire qu'il se base sur une variation admise, qui n'est pas fixe, d'une espèce, et qui dépend clairement de l'interprétation des spécialistes qui travaillent sur le groupe en question. Il peut donc varier entre les groupes taxonomiques (Mann 1999).

Dans le concept d'espèce biologique, une espèce est définie comme une population qui est reproductivement isolée des autres populations. Cependant, l'isolement reproductif ne peut pas être étudié pour des espèces asexuées comme chez beaucoup de micro-organismes. Dans la plupart des cas, il y a si peu d'informations connues sur les capacités de reproduction de ces organismes que le concept d'espèce par défaut a été le concept d'espèce morphologique.

Dans le concept phylogénétique des espèces (Davis 1997), les espèces sont définies comme les plus petits regroupements d'organismes pouvant constituer une structure hiérarchique.

Habituellement, il y a un accord entre les différents concepts, mais certains conflits apparaissent, par exemple lorsque des espèces définies par le concept d'espèce phylogénétique ne sont pas isolées génétiquement. Le concept d'espèce phylogénétique met l'accent sur la différenciation (génétique, morphologique et écologique) plutôt que sur l'isolement reproductif. D'autres désaccords surviennent lorsque les spéciations moléculaires et morphologiques ne se produisent pas au même rythme. En d'autres termes, différents régimes environnementaux ou niches peuvent conduire à une spéciation (Weisse 2007). Le processus de spéciation peut entraîner différentes espèces morphologiques (ou espèces cryptiques), comme la pression de sélection peut ou non être accompagnée de changements morphologiques (Ryneronson & Armbrust 2004).

Le « barcoding moléculaire » ou « DNA barcoding » ou encore « code-barres génétique » est une technique récente de phylogénie moléculaire permettant la caractérisation génétique d'un individu ou d'un échantillon d'individus à partir d'un gène de son génome. Une région du gène mitochondrial codant pour le Cytochrome Oxydase 1 (COI), qui présente plusieurs propriétés intéressantes, a ainsi été définie en 2003 comme le fragment de référence pour la caractérisation des espèces animales (Hebert *et al.* 2003a). Si les taxonomistes utilisent ce concept de code-barres au sens strict, les écologistes en ont eue une vision plus large correspondant à l'utilisation de toute technique d'analyse de l'ADN pour l'identification de taxons dans des échantillons naturels (Valentini *et al.* 2009).

On parle alors de metabarcoding. L'objectif est de pouvoir identifier les espèces présentes dans un milieu alors même que les individus ne sont pas facilement caractérisables (trop petits, trop dispersés). Le principe consiste à extraire l'ADN d'un échantillon environnemental (eau, sol par exemple), puis à amplifier par une réaction en chaîne par polymérase (PCR — Polymerase Chain Reaction) le fragment cible correspondant au code-barres à l'aide d'un couple d'amorces prédéfini. Ces amorces peuvent être spécifiques d'une espèce ou amplifier un large spectre d'espèces.

Un bon code-barres ADN est une séquence variable entre espèces, mais très conservée au sein d'une même espèce, ce qui lui confère un fort pouvoir discriminant. Cette séquence doit être encadrée par deux zones très conservées d'une espèce à l'autre, pour permettre l'amplification du fragment par PCR chez un ensemble d'espèces le plus vaste possible, et assurer ainsi une bonne couverture taxonomique (Taberlet *et al.* 2007). Il est aussi important que les amplifications soient fiables, donc que les sites de fixation des amorces PCR soient également très conservés, et que le fragment amplifié soit assez court. Les fragments de plus de 250 paires de bases (pb) sont plus difficiles à amplifier (en illumina® Miseq). Enfin, pour travailler sur de faibles quantités d'ADN, on privilégie l'utilisation de fragments d'ADN mitochondriaux ou chloroplastiques, car leur nombre de copies par cellule est 100 à 1 000 fois supérieur à celui de l'ADN nucléaire.

Les code-barres ADN doivent être phylogénétiquement informatifs, c'est-à-dire que leur niveau de divergence de référence reflète au mieux le niveau de divergence entre les espèces qui les portent. Cette propriété permet d'assigner des espèces inconnues à un taxon d'ordre supérieur (genre, famille). De cette manière, sans identifier les espèces, on peut utiliser les différentes séquences obtenues pour définir des unités opérationnelles, ou OTUS (*Operational Taxonomic Units*), à partir desquelles il est possible de quantifier la biodiversité d'un échantillon.

### 3.2. Développement du barcoding

Hebert (Hebert *et al.* 2003b) fait le choix en 2003 de l'ADN mitochondrial (ADNmt) plutôt que de l'ADN nucléaire pour son héritage uni-parental (dans la plupart des phyla animaux), son taux d'évolution élevé, l'absence d'introns, un grand nombre de copies dans chaque cellule, et la présence limitée de recombinaisons (Galtier *et al.* 2009). Sa proposition d'utiliser le COI comme gène cible pour du barcoding repose sur des décennies de recherche qui ont montré que son signal phylogénétique est utile pour les niveaux taxonomiques supérieurs ou égaux à l'espèce et qu'il existe des amorces « universelles » capables de récupérer l'extrémité 5' du COI dans la plupart des phyla animaux. Selon cette approche par code-barres, les espèces peuvent être identifiées sur la base d'un « écart de code-

barres », soit une distance génétique, intra et interspécifique en utilisant une valeur de seuil de 2 à 3% (Hebert *et al.* 2003a) ou une valeur de 10 fois (Hebert *et al.* 2004) pour la délimitation des espèces.

Initialement proposé uniquement pour les taxons animaux, un système d'identification basé sur l'ADN a rapidement été imaginé et jugé efficace chez les plantes terrestres (Hollingsworth *et al.* 2009), les algues (Saunders 2005), puis les champignons (Seifert *et al.* 2007), utilisant d'autres régions de l'ADN (mitochondrial, plastidiale, nucléaire) pour une meilleure résolution. L'utilisation d'un outil standardisé (un même fragment d'ADN) permet des comparaisons entre des ensembles de données issues de plusieurs recherches, pouvant révéler des cas de synonymie, d'espèces cryptiques potentielles ou de populations génétiquement distinctes. Comme pour toute autre méthode, le décodage de l'ADN a ses limites : basse résolution dans certains cas (hybrides, espèces ayant récemment divergées), présence de pseudo-gènes (Song *et al.* 2008), amplification de contaminants par des amorces « universelles » (Siddall *et al.* 2009), taux de mutation variable selon les taxons (Galtier *et al.* 2009), etc...

Malgré ces limites, le DNA barcoding reste un outil attrayant pour l'étude de la biodiversité, car il permet l'identification des spécimens durant toutes les étapes de leur vie et s'abstient des cas de dimorphisme sexuel ou d'espèces cryptiques. Un consortium international, le Barcode of Life (CBOL, <http://www.barcodeoflife.org>), dédié à la mise en place d'un outil standard pour l'identification des espèces sur la base des code-barres génétiques a ainsi été créé en 2004. Ce dernier a lancé en 2010 le plus grand projet international de barcoding à ce jour, le Barcode of Life Project (iBOL, <http://www.ibol.org>), qui a pour but d'acquérir les code-barres de 500.000 espèces d'ici à 2015. Il regroupe plus de 120 organisations dans quarante-cinq pays. Cette structure est également proche du consortium « Census of Marine Life » qui a pour objectif d'évaluer la biodiversité, sa répartition et son abondance dans les océans : passé, présent et futur.

### 3.3. Le barcoding en milieu marin

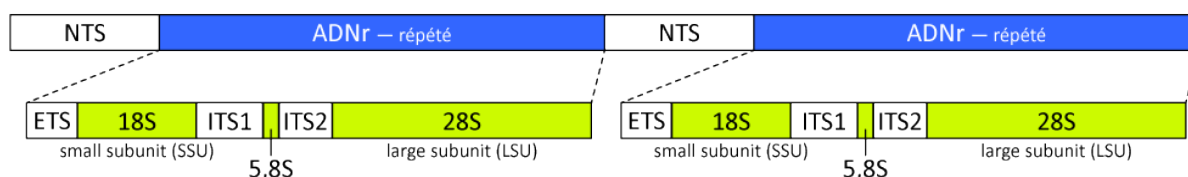
Les code-barres à ADN ont fait leur apparition dans le milieu marin lorsqu'il a été question d'identifier des organismes marins dont l'identification morphologique n'était plus possible, par exemple lorsqu'ils avaient été transformés par l'industrie. L'exemple le plus connu est le poisson, *Lutjanus campechanus*, vendu comme « red-snapper » aux États-Unis (vivaneau rouge en France) et qui en fait s'est avéré être une autre espèce dans 77% des cas (séquences CYT-b, (Marko *et al.* 2004)). Ce cas emblématique a été rapidement suivi par d'autres études qui ont prouvé que les substitutions de produits de la mer sont communes. La question de la validité de l'évaluation des stocks de poissons s'est alors posée.

La plupart des organismes marins ont des stades larvaires difficiles à identifier sur la base de caractères morphologiques, pourtant l'évaluation des stocks passe par l'identification et le

dénombrement de ces juvéniles. Mesurer leur abondance et leur diversité par l'utilisation de code-barres génétiques peut clairement changer les choses, à condition bien entendu qu'une bibliothèque de référence complète pour les adultes ait été développée (Pegg *et al.* 2006). Ce travail de référence a été réalisé ce qui a permis à la méthode d'être appliquée pour le suivi en routine des organismes marins, notamment ceux identifiés comme potentiellement invasifs (Saunders 2009; Manghisi *et al.* 2010).

### 3.4. Le barcoding chez les protistes

L'utilité du COI pour les code-barres à ADN est controversée chez les protistes, avec des tentatives plus ou moins documentées et un succès variable (Ehara *et al.* 2000; Evans *et al.* 2007). Il a cependant été montré qu'un système de code-barres reposant sur des marqueurs mitochondriaux tels que le COI ne fournit pas une solution universelle pour l'identification des protistes dans leur globalité car certaines espèces sont anaérobiques et ne possèdent donc pas de mitochondries (Henze & Martin 2003). En l'absence du COI, l'autre principal groupe de marqueurs pour l'analyse phylogénétique est l'ADN ribosomal. Ces gènes sont présents partout dans le monde vivant et ils évoluent relativement lentement, et leur agencement répétitif dans le génome fournit suffisamment d'ADN pour effectuer des PCR, et ce même chez les plus petits organismes (Hillis & Dixon 1991). Ces marqueurs sont l'ADNr 18S ou la petite sous-unité de l'ADNr (SSU ADNr), l'ADNr 28S ou grande sous-unité ADNr (LSU ADNr), les espaceurs transcrits internes 1 (ITS1) et 2 (ITS2) et l'ADNr 5,8S (Fig. 7).



**Figure 7** : Structure de l'ADN ribosomique chez les eucaryotes

D'autres marqueurs sont utilisés dans les phylogénies moléculaires, ils sont principalement basés sur les gènes codant pour des protéines :  $\alpha$  et  $\beta$ -tubuline, actine, cytochrome oxydase B, et autres (Pochon *et al.* 2012). Ces marqueurs alternatifs ne sont pas bien représentés dans les bases de données génétiques publiques (contrairement à leurs homologues de l'ADNr 18S). La grande majorité des séquences est issue de souches photosynthétiques disponibles dans les collections de cultures. Les phylogénies obtenues ne comprennent alors en général pas assez de taxons pour être utilisées à grande échelle (assez peu d'espèces sont aujourd'hui cultivables en laboratoire). Quoi qu'il en soit, le choix du code-barres spécifique à un groupe est souvent une question de tradition ou de facilité d'utilisation, et les études comparant systématiquement le pouvoir de résolution de différents code-

barres à ADN chez les protistes sont plutôt rares (Saunders & Kucera 2010; Nassonova *et al.* 2010; Heger *et al.* 2011; Hamsher *et al.* 2011). En raison de leur histoire évolutive longue, indépendante et souvent complexe, les protistes sont génétiquement très variables, ce qui réduit considérablement la probabilité de trouver un unique code-barres universel et bien adapté à chacun d'eux. Le consortium ProWG (l'antenne protiste du CBOL) recommande ainsi depuis 2012 (Pawlowski *et al.* 2012) une approche en deux étapes, comprenant une identification préliminaire à l'aide d'un code-barres eucaryote universel, appelé pré-code-barres, suivi d'une identification au niveau de l'espèce à l'aide d'un second code-barres spécifique au groupe ciblé.

#### **4. Les code-barres génétiques chez les dinoflagellés**

Les dinoflagellés possèdent des génomes de taille gigantesque, de 0,2 à 200 pg d'ADN par cellule (Veldhuis *et al.* 1997; LaJeunesse *et al.* 2005) (soit de 1 à 80 fois le génome haploïde humain (Lin 2011)) et de nombreux gènes avec un nombre élevé de copies (Rowan *et al.* 1996). Cette taille gigantesque de leurs génomes a, à ce jour, empêché leur séquençage complet, mais plusieurs gènes ont été séquencés.

La première séquence génétique produite chez les dinoflagellés fut de l'ADN ribosomal, l'ADNr 5S (Hinnebusch *et al.* 1981). Les marqueurs les plus importants utilisés par la suite font eux aussi partie de la gamme de l'ADN ribosomal : SSU ADNr, LSU ADNr et ITS. Les gènes ribosomiques des dinoflagellés ont des taux d'évolution très différents, en particulier dans la SSU ADNr et la LSU ADNr. Ces différences se traduisent en phylogénie par des structures d'arbres phylogénétiques caractéristiques : d'une part, un grand groupe de branches courtes et fortement ramifiées, regroupant les Gymnodiniales, les Prorocentrales et les Peridinales, et d'autre part un groupe avec des branches plus longues comprenant les Gonyaulacales et les autres taxons (Saunders *et al.* 1997; Saldarriaga *et al.* 2004). Les taux de divergence extrêmement faibles entre les marqueurs issus de ces deux régions ne résolvent donc pas les relations entre les principaux ordres. De manière générale, la SSU de l'ADNr permet une assez bonne discrimination au niveau de la famille ou d'un genre alors que le marqueur de la LSU de l'ADNr est plus efficace au niveau des espèces (Gómez 2013).

Dans le cas de l'ITS, le marqueur est présent en plusieurs copies distinctes, avec des variations intra- et inter- génomique relativement importantes associées à la présence de délétions. Ces éléments peuvent rendre difficiles le séquençage puis l'alignement des séquences. Il est donc vraiment difficile de concevoir une amorce spécifique aux dinoflagellés dans la région ITS pour la totalité ou même la plupart des taxons de dinoflagellés (Litaker *et al.* 2007; Pochon *et al.* 2012; Stern *et al.* 2012).

Un autre marqueur, le cytochrome oxydase (COI), est parfois utilisé chez les dinoflagellés. Ce dernier s'est démontré utile pour identifier des cellules individuelles ou des cultures mono-spécifiques. Il est utilisé, avec succès, par exemple pour distinguer les différents génotypes des symbiotes de

coraux (*Symbiodinium*) (Takabayashi *et al.* 2004). Cependant, son utilisation comme code-barres chez les dinoflagellés de façon générale reste très discutable car on observe régulièrement des échecs de fixation de ses amorces dans les échantillons naturels. Plusieurs travaux ont d'ailleurs montré que son pouvoir de résolution est en général plus faible que les autres gènes jusqu'ici étudiés et utilisés (Zhang *et al.* 2007; Lin *et al.* 2009; Stern *et al.* 2010).

Les gènes plastidiaux peuvent également servir de code-barres efficaces pour certaines espèces, mais ils sont à écarter si l'on souhaite s'intéresser à l'ensemble du groupe des dinoflagellés car seuls 50% des taxons sont susceptibles d'en posséder. Il faut aussi savoir que parmi les espèces qui possèdent des plastides, certaines ont pu les gagner ou les perdre à de multiples occasions : la raison est la kleptoplastidie (la capacité à asservir des chloroplastes d'algues ingérés et à effectuer de la photosynthèse de façon éphémère (Feinstein *et al.* 2002)) (Saldarriaga *et al.* 2001; Hackett *et al.* 2004).

Si les code-barres génétiques nous permettent une vision plus globale de l'environnement, il n'existe pas, chez les dinoflagellés, de marqueur universel nous permettant d'en faire une analyse *in situ* détaillée. Cependant, grâce aux différents travaux accomplis jusqu'à aujourd'hui nous pouvons nous faire une idée assez précise des limites techniques de chaque marqueur. Il nous est ainsi possible d'appréhender la diversité de ce groupe à l'échelle des océans pour déterminer la présence de patrons généraux.





## Objectifs de la thèse et structure du document

L'objectif principal de cette thèse est la recherche de patrons globaux de biodiversité et de structuration des communautés de dinoflagellés planctoniques à l'échelle de l'océan mondial. Pour atteindre cet objectif, il a fallu mettre en place un échantillonnage à l'échelle globale, identifier la technologie qui me permettrait d'obtenir un jeu de données conséquent puis analyser ces données de manière pertinente.

La première partie de ce travail de thèse a donc consisté en l'élaboration et en la mise en pratique sur le terrain d'un protocole d'échantillonnage des protistes marins, puis à l'exploration d'outils d'analyses à haut-débit (en optique et en génétique) et enfin, au développement et à la validation de protocoles pour l'analyse de la biodiversité des dinoflagellés.

J'ai participé à l'élaboration d'un protocole d'échantillonnage morphogénétique homogène, robuste et continu de l'ensemble des dinoflagellés marins, couvrant la totalité de leurs spectres de taille cellulaire. J'ai personnellement participé à la mise en pratique de ce protocole d'échantillonnage pendant l'expédition Tara Océans pour une période de presque 5 mois répartie en 3 missions successives dans l'océan Pacifique : 2 en Polynésie française et une entre San Diego et Panama. La première mission a porté sur les milieux récifaux, la seconde fut marquée par l'étude et le suivi de masses d'eau au large des îles Marquises et la dernière s'est déroulée dans l'océan Pacifique ouvert avec l'étude de la biodiversité dans un milieu ultra-oligotrophique. Cette partie de mon travail de thèse a fait l'objet d'un article méthodologique présenté dans ce manuscrit (Not *et al.*, submitted) (**chapitre 1**). Une première approche de l'analyse de la diversité a été basée sur la technologie d'imagerie automatique FlowCAM embarquée à bord de Tara, et dont les images acquises à chaque station ont été analysées en mode semi-automatique. La diversité, l'abondance relative et la distribution géographique des espèces au sein du genre de dinoflagellés *Neoceratium* ont ainsi été évaluées en mer Méditerranée (**chapitre 2**). Une deuxième approche a consisté à utiliser des metabarcodes comme marqueurs de la biodiversité des dinoflagellés, un fragment de l'ADN ribosomique (région V9 du 18S de l'ADNr). Cette région a été amplifiée par pyroséquençage à partir d'extraits d'eau de mer issue des échantillons de la mission Tara Océans. L'analyse par metabarcoding nécessite la construction d'une base de séquences génétiques de référence afin d'assigner une taxonomie aux fragments d'ADN environnementaux. J'ai donc constitué une base de référence « dinoflagellés » (DinR2 — Dinoflagellate Ribosomal Reference database) qui a été intégrée à une base de référence plus large concernant l'ensemble des protistes marins (PR2 — Protist Ribosomal Reference database) (Guillou *et al.*, 2013). Cette base de référence a été validée en comparant les données génétiques aux comptages cellulaires des échantillons de certaines stations du projet Tara

Océans. Les données morphologiques ont été analysées et mises à disposition par la Stazione Zoologica de Naples (Italie) (Dr. Adriana Zingone et Dr. Eleonora Scalco) partenaire dans le projet Tara (**chapitre 3**).

L'évaluation des résultats obtenus par les deux approches a montré que le metabarcoding permet une caractérisation plus complète de la biodiversité des dinoflagellés marins. Cette technologie est donc plus pertinente pour répondre à mes objectifs scientifiques. Dans la seconde partie de mes travaux de thèse, présentés dans la seconde partie de ce manuscrit, je me concentre sur l'analyse de ces metabarcodes.

L'assignation d'une taxonomie traditionnelle aux metabarcodes a pour objectif de lier des connaissances classiques de la biodiversité des dinoflagellés à des données et analyses de séquençage à très haut-débit. J'ai ainsi cherché à identifier des patrons de structuration de la biodiversité des communautés de dinoflagellés à travers différentes classes de tailles de protistes marins communs à plusieurs écosystèmes (Le Bescot *et al.*, submitted) (**chapitre 4**). Enfin, j'ai démarré une étude de l'influence des facteurs environnementaux sur l'organisation des méta-communautés de dinoflagellés identifiées par metabarcoding. Cette étude préliminaire vise à identifier certains facteurs physico-chimiques qui pourraient contribuer à la structuration fonctionnelle des communautés de dinoflagellés dans la zone photique de l'océan global (Le Bescot *et al.*, In prep) (**chapitre 5**).

## Références

- Adl SM, Hoppenrath M, Simpson AGB *et al.* (2012) The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, **59**, 429–493.
- Amann R, Binder BJ, Olson RJ *et al.* (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology*, **56**, 1919–1925.
- Arrigo KR (2004) Marine microorganisms and global nutrient cycles. *Nature*, **437**, 349–355.
- Bell JL, Hopcroft RR (2008) Assessment of ZoolImage as a tool for the classification of zooplankton. *Journal of Plankton Research*, **30**, 1351–1367.
- Cavalier-Smith T, Chao EE (2004) Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). *European Journal of Protistology*, **40**, 185–212.
- Culverhouse PF, Williams R, Benfield M *et al.* (2006) Automatic image analysis of plankton: future perspectives. *Marine Ecology Progress Series*, **312**, 297–309.
- Davis JI (1997) Evolution, Evidence, and the Role of Species Concepts in Phylogenetics. *Systematic Botany*, **22**, 373–403.
- Delwiche CF (1999) Tracing the Thread of Plastid Diversity through the Tapestry of Life. *The American Naturalist*, **154**, S164–S177.
- Dolan JR (2005) Biogeography of aquatic microbes. *Aquatic Microbial Ecology*, **41**, 39–48.
- Doucette GJ, Cembella AD, Boyer GL (1989) Cyst formation in the red tide dinoflagellate *Alexandrium tamarense* (dinophyceae): effects of iron stress. *Journal of Phycology*, **25**, 721–731.
- Ehara M, Inagaki Y, Watanabe KI, Ohama T (2000) Phylogenetic analysis of diatom *cox1* genes and implications of a fluctuating GC content on mitochondrial genetic code evolution. *Current Genetics*, **37**, 29–33.
- Evans KM, Wortley AH, Mann DG (2007) An Assessment of Potential Diatom “Barcode” Genes (*cox1*, *rbcL*, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). *Protist*, **158**, 349–364.
- Feinstein TN, Traslavina R, Sun MY, Lin S (2002) Effects of Light on Photosynthesis, Grazing, and Population Dynamics of the Heterotrophic Dinoflagellate *Pfiesteria piscicida* (Dinophyceae). *Journal of Phycology*, **38**, 659–669.
- Fensome RA, MacRae AR, Williams GL (1998) Évolution des dinoflagellés et variation de leur diversité dans le temps. *Marine Micropaleontology*, 1–6.
- Fensome RA, Taylor FJR, Norris G *et al.* (1993) A classification of living and fossil dinoflagellates. *Micropaleontology Press*, **7**, 1–351.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski PG (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, **281**, 237–240.
- Figueroa RI, Bravo I, Garcés E (2006) Multiple routes of sexuality in *Alexandrium Taylori* (dinophyceae) in culture. *Journal of Phycology*, **42**, 1028–1039.
- Figueroa RI, Garcés E, Camp J (2010) Reproductive plasticity and local adaptation in the host–parasite system formed by the toxic *Alexandrium minutum* and the dinoflagellate parasite *Parvilucifera sinerae*. *Harmful Algae*, **10**, 56–63.
- Fistarol GO, Legrand C, Rengefors K, Graneli E (2004) Temporary cyst formation in phytoplankton: a response to allelopathic competitors? *Environmental Microbiology*, **6**, 791–798.
- Foissner W, Hawksworth DL (2009) Protist diversity and geographical distribution. *Biodiversity and Conservation*, **17**, 243–259.
- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

- Garcés E, Delgado M, Masó M, Camp J (1998) Life history and in situ growth rates of *Alexandrium taylori* (dinophyceae, pyrrophyta). *Journal of Phycology*, **34**, 880–887.
- Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. **345**, 60–63.
- Gómez F (2012) A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity*, **10**, 267–275.
- Gómez F (2013) Problematic Biases in the Availability of Molecular Markers in Protists: The Example of the Dinoflagellates. *Acta Protozoologica*, **52**, 21–33.
- Graham LE, Wilcox LW (2000) *Algae*. Upper Saddle River: Prentice-Hall.
- Granéli E, Turner JT (2007) *Ecology of harmful algae*. Springer.
- Groben R, John U, Eller G, Lange M, Medlin LK (2004) Using fluorescently-labelled rRNA probes for hierarchical estimation of phytoplankton diversity – a mini-review. *Nova Hedwigia*, **79**, 313–320.
- Grosjean P, Picheral M, Warembourg C, Gorsky G (2004) Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES Journal of Marine Science: Journal du Conseil*, **61**, 518–525.
- Grzebyk D, Berland B (1996) Influences of temperature, salinity and irradiance on growth of *Prorocentrum minimum* (Dinophyceae) from the Mediterranean Sea. *Journal of Plankton Research*, **18**, 1837–1849.
- Hackett JD, Anderson DM, Erdner DL, Bhattacharya D (2004) Dinoflagellates: a remarkable evolutionary experiment. *American Journal of Botany*, **91**, 1523–1534.
- Hallegraeff GM (2003) Harmful algal blooms: a global overview. *Phycologia*, **11**, 25–49.
- Hamsher SE, Evans KM, Mann DG, Poulíčková A, Saunders GW (2011) Barcoding Diatoms: Exploring Alternatives to COI-5P. *Protist*, **162**, 405–422.
- Havskum H, Hansen PJ, Berdalet E (2005) Effect of turbulence on sedimentation and net population growth of the dinoflagellate *Ceratium tripos* and interactions with its predator, *Fragilidium subglobosum*. *Limnology And Oceanography*, **50**, 1543–1551.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- Hebert PDN, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S96–S99.
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biology*, **2**, e312.
- Heger TJ, Pawlowski J, Lara E *et al.* (2011) Comparing potential COI and SSU rDNA barcodes for assessing the diversity and phylogenetic relationships of cyphoderiid testate amoebae (Rhizaria: Euglyphida). *Protist*, **162**, 131–141.
- Henze K, Martin W (2003) Evolutionary biology: essence of mitochondria. *Nature*, **426**, 127–128.
- Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology*, **66**, 411–453.
- Hinnebusch AG, Klotz LC, Blanken RL, Loeblich AR III (1981) An evaluation of the phylogenetic position of the dinoflagellate *Cryptothecodinium cohnii* based on 5S rRNA characterization. *Journal of molecular evolution*, **17**, 334–347.
- Hollingsworth PM, Forrest LL, Spouge JL *et al.* (2009) A DNA barcode for land plants. *PNAS*, **106**, 12794–12797.
- Laabir M, Amzil Z, Lassus P *et al.* (2007) Viability, growth and toxicity of *Alexandrium catenella* and *Alexandrium minutum* (Dinophyceae) following ingestion and gut passage in the oyster *Crassostrea gigas*. *Aquatic Living Resources*, **20**, 51–57.

- Labry C, Erard Le Denn E, Chapelle A *et al.* (2008) Competition for phosphorus between two dinoflagellates: A toxic *Alexandrium minutum* and a non-toxic *Heterocapsa triquetra*. *Journal of Experimental Marine Biology and Ecology*, **358**, 124–135.
- LaJeunesse TC, Lambert G, Andersen RA, Galbraith DW, Coffroth M-A (2005) Symbiodinium (Pyrrhophyta) Genome Sizes (Dna Content) Are Smallest Among Dinoflagellates. *Journal of Phycology*, **41**, 880–886.
- Leander BS, Keeling PJ (2003) Morphostasis in alveolate evolution. *Trends in Ecology & Evolution*, **18**, 395–402.
- Lee RE (2007) *Phycology, Fourth edition*. Cambridge.
- Lin S (2011) Genomic understanding of dinoflagellates. *Research in microbiology*, **162**, 551–569.
- Lin S, Zhang H, Hou Y, Zhuang Y, Miranda L (2009) High-Level Diversity of Dinoflagellates in the Natural Environment, Revealed by Assessment of Mitochondrial *cox1* and *cob* Genes for Dinoflagellate DNA Barcoding. *Applied and Environmental Microbiology*, **75**, 1279–1290.
- Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*, **438**, 86–89.
- Litaker RW, Vandersea MW, Kibler SR *et al.* (2002) Life cycle of the heterotrophic dinoflagellate *Pfiesteria piscicida* (dinophyceae). *Journal of Phycology*, **38**, 442–463.
- Litaker RW, Vandersea MW, Kibler SR *et al.* (2007) Recognizing dinoflagellate species using ITS rDNA sequences. *Journal of Phycology*, **43**, 344–355.
- Lopez-Garcia P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**, 603–607.
- Lundholm N, Moestrup Ø (2006) The biogeography of harmful algae. In: *Ecology of harmful algae*, pp. 23–35. Springer.
- Manghisi A, Morabito M, Bertuccio C *et al.* (2010) Is routine DNA barcoding an efficient tool to reveal introductions of alien macroalgae? A case study of *Agardhiella subulata* (Solieriaceae, Rhodophyta) in Cape Peloro lagoon (Sicily, Italy). *Cryptogamie Algologie*, **31**, 423.
- Mann DG (1999) The species concept in diatoms. *Phycologia*, **38**, 437–495.
- Marko PB, Lee SC, Rice AM *et al.* (2004) Fisheries: Mislabelling of a depleted reef fish. *Nature*, **430**, 309–310.
- Martin R, Quigg A (2013) Tiny Plants That Once Ruled the Seas. *Scientific American*.
- Massana R, Pedrós-Alió C (2008) Unveiling new microbial eukaryotes in the surface ocean. *Current opinion in microbiology*, **11**, 213–218.
- McFadden GI, Reith ME, Munholland J, Lang-Unnasch N (1996) Plastid in human parasites. *Nature*, **381**, 482.
- Metfies K, Valentin K, Medlin LK, Mehl H, Wiltshire K (2006) Picoeukaryotic plankton diversity at the Helgoland time series site as assessed by three molecular methods. *Microbial ecology*, **52**, 53–71.
- Nassonova E, Fahrni J, Smirnov AV, Pawlowski J (2010) Barcoding Amoebae: Comparison of SSU, ITS and COI Genes as Tools for Molecular Identification of Naked Lobose Amoebae. *Protist*, **161**, 102–115.
- Not F, Siano R, Kooistra WHCF *et al.* (2012) Diversity and Ecology of Eukaryotic Marine Phytoplankton. In: *Genomic insights into the biology of algae*, pp. 1–53. Elsevier.
- Ochoa J, Maske H, Sheinbaum J, Candela J (2013) Diel and lunar cycles of vertical migration extending to below 1000 m in the ocean and the vertical connectivity of depth-tiered populations. *Limnology And Oceanography*, **58**, 1207–1214.
- Pawlowski J, Simpson AGB, Audic S *et al.* (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biology*, **10**, e1001419.
- Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef–Australia. *Scientia Marina*, **70**, 7–12.
- Persson A, Smith BC, Wikfors GH, Quilliam M (2006) Grazing on toxic *Alexandrium fundyense* resting cysts and vegetative cells by the eastern oyster (*Crassostrea virginica*). *Harmful Algae*, **5**, 678–684.

- Pfiester LA, Anderson DM (1987) Dinoflagellate reproduction. In: *The Biology of Dinoflagellates*, pp. 611–648. The Biology of Dinoflagellates (edited by Tylor FJR).
- Pochon X, Putnam HM, Burki F, Gates RD (2012) Identifying and Characterizing Alternative Molecular Markers for the Symbiotic and Free-Living Dinoflagellate Genus *Symbiodinium*. *PLoS ONE*, **7**, e29816.
- Rowan R, Whitney SM, Fowler A, Yellowlees D (1996) Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *The Plant cell*, **8**, 539–553.
- Rynearson TA, Armbrust EV (2004) Genetic differentiation among populations of the planktonic marine diatom *ditylum brightwellii* (bacillariophyceae) 1. *Journal of Phycology*, **40**, 34–43.
- Saldarriaga JF, Taylor FJR, Cavalier-Smith T, Menden-Deuer S, Keeling PJ (2004) Molecular data and the evolutionary history of dinoflagellates. *European Journal of Protistology*, **40**, 85–111.
- Saldarriaga JF, Taylor FJR, Keeling PJ, Cavalier-Smith T (2001) Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *Journal of molecular evolution*, **53**, 204–213.
- Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1879.
- Saunders GW (2009) Routine DNA barcoding of Canadian Gracilariales (Rhodophyta) reveals the invasive species *Gracilaria vermiculophylla* in British Columbia. *Molecular Ecology Resources*, **9**, 140–150.
- Saunders GW, Kucera H (2010) An evaluation of *rbcl*, *tufA*, *UPA*, *LSU* and *ITS* as DNA barcode markers for the marine green macroalgae. *Cryptogamie Algologie*, **31**, 487–528.
- Saunders GW, Hill D, Sexton JP, Andersen RA (1997) Small-subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. In: *Origins of algae and their plastids*. Origins of Algae and their Plastids.
- Seifert KA, Samson RA, deWaard JR *et al.* (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3901–3906.
- Siddall ME, Fontanella FM, Watson SC, Kvist S, Erséus C (2009) Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. *Systematic Biology*, **58**, 445–451.
- Sieracki CK, Sieracki ME, Yentsch CS (1998) An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*, **168**, 285–296.
- Simon N, Campbell L, Örnólfsson E *et al.* (2000) Oligonucleotide Probes for the Identification of Three Algal Groups by Dot Blot and Fluorescent Whole-Cell Hybridization. *Journal of Eukaryotic Microbiology*, **47**, 76–84.
- Siokou-Frangou I, Christaki U, Mazzocchi MG *et al.* (2010) Plankton in the open Mediterranean Sea: a review. *Biogeosciences*, **7**, 1543–1586.
- Smayda TJ (1997) Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology And Oceanography*, **42**, 1137–1153.
- Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13486–13491.
- Stern RF, Andersen RA, Jameson I *et al.* (2012) Evaluating the Ribosomal Internal Transcribed Spacer (ITS) as a Candidate Dinoflagellate Barcode Marker. *PLoS ONE*, **7**, e42780.
- Stern RF, Horak A, Andrew RL *et al.* (2010) Environmental Barcoding Reveals Massive Dinoflagellate Diversity in Marine Environments. *PLoS ONE*, **5**, e13991.
- Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic acids research*, **35**, e14–e14.
- Takabayashi M, Santos SR, Cook CB (2004) Mitochondrial DNA Phylogeny of the Symbiotic Dinoflagellates (*Symbiodinium*, Dinophyta). *Journal of Phycology*, **40**, 160–164.

- Taylor FJR (2001) Harmful Algal Blooms 2000 : Proceedings of the Ninth International Conference, Hobart, Australia, 7-11 February 2000-2001. *Harmful Algal Blooms*, 1–5.
- Taylor FJR (2004) Extraordinary Dinoflagellates: Past and Present. 61–66.
- Taylor FJR, Hoppenrath M, Saldarriaga JF (2007) Dinoflagellate diversity and distribution. *Biodiversity and Conservation*, **17**, 407–418.
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**, 110–117.
- Veldhuis MJW, Cucci TL, Sieracki ME (1997) Cellular DNA Content of Marine Phytoplankton Using Two New Fluorochromes: Taxonomic and Ecological Implications1. *Journal of Phycology*, **33**, 527–541.
- Venter JC, Remington K, Heidelberg JF *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66–74.
- Wang ZH, Qi YZ, Yang YF (2007) Cyst formation: an important mechanism for the termination of *Scrippsiella trochoidea* (Dinophyceae) bloom. *Journal of Plankton Research*, **29**, 209–218.
- Weisse T (2007) Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. *Biodiversity and Conservation*, **17**, 243–259.
- Wiebe PH, Benfield MC (2001) Zooplankton Sampling With Nets And Trawls. In: *Encyclopedia of Ocean Sciences*, pp. 3237–3253. Elsevier.
- Zhang H, Bhattacharya D, Lin S (2007) A Three-Gene Dinoflagellate Phylogeny Suggests Monophyly of Prorocentrales and a Basal Position for Amphidinium and Heterocapsa. *Journal of molecular evolution*, **65**, 463–474.





# **Partie 1**

## **Stratégies d'échantillonnage et validation des outils d'analyse**



# Chapitre 1

## L'expédition Tara Océans : Un échantillonnage complet du plancton marin dans son contexte environnemental

Submitted to PLoSOne



# Sommaire

## Chapitre 1 : L'expédition Tara Océans : Un échantillonnage complet du plancton marin dans son contexte environnemental

|   |           |
|---|-----------|
| <b>Tara Océans</b>  | <b>49</b> |
| <b>1. Contexte du projet et objectifs</b>   | <b>49</b> |
| <b>2. Les principales problématiques de l'échantillonnage</b>   | <b>50</b> |
| 2.1. Les contraintes techniques   | 50        |
| 2.2. Les contraintes biologiques  | 51        |
| 2.3. La stratégie d'échantillonnage   | 51        |
| <br>  |           |
| <b>The Tara Oceans Expedition: A comprehensive sampling of marine plankton biota in their environmental context</b> | <b>53</b> |
| Abstract  | 53        |
| <br>  |           |
| <b>References</b>   | <b>67</b> |

## Tara Océans

### 1. Contexte du projet et objectifs

Les océans contiennent une masse considérable d'organismes microscopiques qui sont à l'origine de la vie, et de notre existence sur Terre. Les premiers organismes marins ont probablement été des bactéries et des virus, mais l'évolution a produit au fil du temps dans les océans des organismes d'une bien plus grande complexité appelés les « eucaryotes unicellulaires » ou aussi protistes marins. Cette explosion de la biodiversité marine est le résultat d'une période d'environ 1 milliard d'années, pendant laquelle se sont produites des fusions, des incorporations de divers types cellulaires, et d'importants échanges génétiques.

Les écosystèmes marins actuels tirent leur robustesse de cette importante diversité générée par l'évolution. Ces organismes microscopiques se reproduisent à grande vitesse et colonisent tous les écosystèmes depuis des régions chaudes vers des régions froides, depuis des régions acides vers des régions moins acides, depuis des zones anoxiques vers des zones riches en oxygène, ou encore depuis des zones pauvres en nutriments vers des régions fertiles. Nos connaissances actuelles nous amènent à supposer qu'il existe une importante variabilité de la composition en organismes et en gènes entre ces différents écosystèmes planctoniques marins au fil des différentes régions océaniques (Karsenti *et al.* 2011).

Le but du projet Tara Océans a été d'explorer la diversité des protistes marins, mais aussi, des virus, des bactéries et des métazoaires de la manière la plus complète et approfondie possible. Cette exploration utilisant plusieurs approches allant de la microscopie à l'imagerie à haut débit, ou encore du méta-barcoding à la méta-génomique, a intégré pour chaque prélèvement un contexte physico-chimique et concerné presque toutes les régions océaniques du monde. Le parcours de l'expédition a été planifié selon deux impératifs : une navigation dans des conditions météorologiques les plus favorables possibles (vents portants et conditions favorables au déploiement du matériel scientifique) et la traversée des principales grandes zones océaniques présentant des propriétés océanographiques un maximum contrastées. La goélette Tara a accompli un parcours qui a permis d'échantillonner la mer Méditerranée, la mer Rouge, le Nord de l'Océan Indien (une région océanique fortement acidifiée avec en profondeur des eaux pauvres en oxygène), le Sud de cet Océan avec le canal du Mozambique, l'Atlantique Sud, l'Antarctique, le Pacifique Sud (« un désert » océanique), la région des Galápagos (qui est également une région touchée par l'acidification), les régions équatoriales, le Pacifique Nord, le système du Gulf Stream et l'Atlantique Nord (Fig 1).

Pour mener à bien ce projet unique, il a fallu mettre au point une méthode constante d'échantillonnage permettant de collecter de façon quantitative (ou semi-quantitative) des

organismes ayant des tailles très diverses allant de moins d'un micromètre pour les virus et les bactéries, au centimètre pour les métazoaires. En parallèle, des outils de mesures physiques devaient relever en continu un certain nombre de paramètres capables de caractériser précisément les zones traversées. L'échantillonnage a été réalisé à trois niveaux de profondeurs : en surface au niveau de la zone photique (zone éclairée), entre 20 et 100 mètres de profondeur dans une zone appelée « Deep Chlorophyll Maximum » et dans la zone méso-pélagique (entre 300 et 400 mètres de profondeur), une zone où la lumière est absente. Cette stratégie a permis de prélever des organismes planctoniques dans des conditions physiques radicalement différentes et devrait fournir des données permettant d'examiner la nature des corrélations entre environnement et composition en organismes et/ou en gènes, caractéristiques à chaque écosystème océanique.

En deux ans et demi de traversée (de septembre 2009 à mars 2012), c'est au total, 153 stations qui ont été réalisées permettant de collecter 27 000 échantillons, dont environ la moitié est destinée à des analyses génomiques, et l'autre moitié à de l'imagerie quantitative. 455 profils CTD décrivent les paramètres physico-chimiques des différentes colonnes d'eau échantillonnées ainsi que 1 800 échantillons de sels nutritifs. 825 échantillons pour la chimie des carbonates et 1 500 échantillons pour la quantification des pigments chlorophylliens viennent compléter cette collection unique au monde.

## **2. Les principales problématiques de l'échantillonnage**

### *2.1. Les contraintes techniques*

Déployer un dispositif d'exploration de la vie microbienne marine sur une période aussi longue et à une échelle géographique aussi vaste demande une organisation spécifique et unique. Le protocole doit tenir compte des spécificités techniques du navire, une goélette de 36 m de long et de 10 m de large. Les manipulations doivent pouvoir être exécutées dans des conditions de navigation diverses, tout en garantissant une qualité constante des prélèvements. Le nombre, le stockage et le transfert des échantillons doivent être optimisés et garantis. Le navire Tara s'est donc vu modifié : des espaces ont été aménagés pour le stockage des échantillons, des cabines ont été converties en laboratoire, des systèmes de prélèvements continus ont été installés sous la coque et le pont arrière a été équipé d'un portique pour la mise à l'eau du matériel. L'utilisation d'appareils électroniques et optiques sensibles à l'humidité et d'outils de filtration, par définition exposés directement à l'eau de mer, a contraint l'aménagement de l'espace en deux laboratoires distincts. Un laboratoire dit « humide » situé sur le pont du navire en accès direct avec l'océan et un laboratoire dit « sec » à l'atmosphère plus confinée et contrôlée, située à l'intérieur du bateau. De même, des espaces de stockage spécifiques et sécurisés ont été conçus dans les laboratoires ou dans la cale du navire (réfrigérateurs, congélateurs, réservoirs d'azote liquide). Les dispositifs de prélèvements,

habituellement déployés sur de gros navires océanographiques, ont été repensés, parfois redessinés ou simplement miniaturisés.

### 2.2. Les contraintes biologiques

Le plancton est traditionnellement prélevé à partir de filets constitués d'un grand cercle métallique sur lequel est attachée une toile de nylon ou de soie de forme conique qui se termine par un collecteur. Ce dispositif parfaitement adapté à l'étude du nano-, micro-, méso- et macro-plancton, ne permet pas d'étudier le femto-plancton ( $< 0,8 \mu\text{m}$ ). L'objectif étant ici de prélever l'ensemble de la communauté, il a fallu imaginer un système de pompage et de filtration embarqué spécifique à cette fraction de taille. Constitué d'une succession de filets directement disposés sur le pont du navire (utilisant la loi de la gravité), comme le GPSS (Garvity Plankton Sieving System), ou de dispositifs de pompage (jusqu'à 50 m de profondeur) et de filtrations en séries, de nouveaux outils et méthodes ont été testés. Une rosette, équipée de bouteilles Niskin, a été spécialement conçue pour pouvoir prélever jusqu'à 100 L d'eau de mer en une seule plongée pouvant atteindre les 2 000 m de profondeur. Les volumes d'eau à prélever pour obtenir une couverture la plus complète possible de la diversité ont été identifiés à partir de la littérature. Pour chaque groupe d'organismes (et donc fractions de tailles) un inventaire de la diversité décrite et une évaluation théorique de la concentration des cellules par litre d'eau de mer ont été réalisés. Ainsi un volume théorique a pu être défini pour chaque outil de prélèvement. Une série d'échantillons, la plus complète et diverse possible, a été imaginée pour permettre une analyse morphologique et moléculaire des 4 fractions de tailles de protistes. Une large gamme de fixateurs (lugol, formol, éthanol) garantit une grande plasticité pour les analyses morphologiques à long terme. Chaque prélèvement a été conçu pour générer un duplicata d'échantillons qui pourra servir à des analyses complémentaires plus spécifiques ou à de l'archivage. De nombreux et très divers outils et méthodes ont été associés dans le cadre de ce projet.

### 2.3. La stratégie d'échantillonnage

Le choix des zones de prélèvements est extrêmement complexe. Dépendant des conditions météorologiques, mais également de la structuration de l'espace à étudier, chaque point de prélèvement a fait l'objet d'une analyse minutieuse de la part du personnel scientifique embarqué mais aussi d'équipes internationales situées dans des laboratoires à terre. Les caractéristiques connues des différents espaces océaniques de la planète ont servi à la sélection des zones d'intérêts puis ont été associées aux mesures faites *in situ* au moment de l'arrivée du bateau sur zone pour assurer une caractérisation la plus complète possible de l'environnement ainsi choisi. L'expédition a été conçue dans le but de décrire et de contextualiser la plus large variété d'écosystèmes possible, en dehors de toute notion de temporalité. Les périodes et les zones de navigation ont été calculées en fonction des



saisons et des conditions météorologiques et les régions d'escales sélectionnées pour leurs facilités d'expédition des échantillons vers les laboratoires situés en Europe et ailleurs dans le monde.

Ce projet a permis de démontrer qu'une petite unité (un voilier de 36 m de long) correctement organisée et équipée des dernières technologies était capable d'apporter un regard nouveau sur l'exploration de la vie marine et de collecter des données d'une valeur exceptionnelle.

Submitted to PLoSOne

## The Tara Oceans Expedition: A comprehensive sampling of marine plankton biota in their environmental context

Fabrice Not<sup>1\*</sup>, Stéphane Pesant<sup>2\*</sup>, Marc Picheral<sup>3\*</sup>, Stephanie Kandels-Lewis<sup>4\*</sup>, **Noan Le Bescot<sup>1\*</sup>**, Sarah Searson<sup>3,5</sup>, Céline Dimier<sup>1</sup>, Julie Poulain<sup>6</sup>, Sarah Romac<sup>1</sup>, Hervé Le Goff<sup>7</sup>, Margaux Carmichael<sup>1</sup>, Johan Decelle<sup>1</sup>, Jean-Baptiste Romagnan<sup>3</sup>, Isabel Ferrera<sup>8</sup>, Hugo Sarmiento<sup>9</sup>, Camille Clerissi<sup>10</sup>, Lucie Subirana<sup>10</sup>, Jennifer R. Brum<sup>11</sup>, Lionel Guidi<sup>3</sup>, Francois Roullier<sup>3</sup>, Léo Berline<sup>3</sup>, Joséphine Ras<sup>3</sup>, Lucie Bittner<sup>1</sup>, Ian Probert<sup>1</sup>, Sébastien Colin<sup>1</sup>, Sasha Tozzi<sup>12</sup>, Nicole Poulton<sup>13</sup>, Jean-Pierre Gattuso<sup>3</sup>, Uros Krzic<sup>4</sup>, Lee Karp-Boss<sup>14</sup>, Christian Sardet<sup>3</sup>, Shinichi Sunagawa<sup>4</sup>, Jeroen Raes<sup>15</sup>, Olivier Jaillon<sup>6</sup>, Patrick Wincker<sup>6</sup>, Michael Sieracki<sup>13</sup>, Emmanuel Boss<sup>16</sup>, Sabrina Speich<sup>17</sup>, Pierre Testor<sup>7</sup>, Fabrizio D'ortenzio<sup>3</sup>, Lars Stemmann<sup>3</sup>, Nigel H. Grimsley<sup>10</sup>, Pascal Hingamp<sup>18</sup>, Daniele Iudicone<sup>19</sup>, Emmanuel G. Reynaud<sup>20</sup>, Matthew B. Sullivan<sup>11</sup>, Silvia G. Acinas<sup>8</sup>, Hiroyuki Ogata<sup>21</sup>, Christopher Bowler<sup>22</sup>, Colomban de Vargas<sup>1</sup>, Gabriel Gorsky<sup>3</sup>, Eric Karsenti<sup>4</sup>, Etienne Bourgois<sup>23</sup>, and the Tara Oceans Consortium

\* These authors contributed equally to this work.

Corresponding author: not@sb-roscoff.fr

<sup>1</sup> CNRS (UMR7144), Université & Pierre et Marie Curie (Paris 06), Station Biologique de Roscoff, Roscoff, France, <sup>2</sup> MARUM—Center for Marine Environmental Sciences, Universität Bremen, Bremen, Germany, <sup>3</sup> CNRS-UPMC, Laboratoire d'Océanographie de Villefranche (LOV), Villefranche-sur-Mer, France, <sup>4</sup> European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, <sup>5</sup> Department of Oceanography, University of Hawai'i at Manoa, Honolulu, HI, USA, <sup>6</sup> CEA, Institut de Génomique, Genoscope, Evry, France, <sup>7</sup> Laboratoire d'Océanographie et de Climatologie: Expérimentations et Approches Numériques (LOCEAN), CNRS/UPMC/IRD/MNHN, Paris, France, <sup>8</sup> Department of Marine Biology and Oceanography, Institute of Marine Science (ICM), CSIC, Passeig Marítim de la Barceloneta, Barcelona, Spain, <sup>9</sup> Federal University of Rio Grande do Norte, Department of Oceanography and Limnology, 'PPG Ecologia', Natal, Brazil, <sup>10</sup> CNRS (UMR7232), Université Pierre et Marie (Paris 06), Observatoire Océanologique, Banyuls-sur-Mer, France

<sup>11</sup> University of Arizona, Tucson, Arizona, United States of America, <sup>12</sup> University of California Santa Cruz (UCSC), Ocean Sciences, Santa Cruz, United States of America, <sup>13</sup> Bigelow Laboratory for Ocean Sciences, Bigelow Drive, East Boothbay, United States of America, <sup>14</sup> School of Marine Sciences, University of Maine, Orono, ME, United States of America, <sup>15</sup> Vrije Universiteit Brussel (VIB), Department of Structural Biology, Brussel, Belgium, <sup>16</sup> Bigelow Laboratory for Ocean Sciences, Bigelow Drive, East Boothbay, United States of America

<sup>17</sup> Université de Bretagne Occidentale (UBO), Institut Universitaire Européen de la Mer (IUEM), Brest, France, <sup>18</sup> CNRS, Aix-Marseille Université, Laboratoire Information Génomique et Structurale (UMR 7256), Mediterranean Institute of Microbiology (FR 3479), Marseille, France, <sup>19</sup> Stazione Zoologica Anton Dohrn di Napoli, Napoli, Italy, <sup>20</sup> Earth Institute, UCD, Belfield, Dublin 4, Ireland, <sup>21</sup> Education Academy of Computational Life Sciences, Tokyo Institute of Technology, Tokyo, Japan, <sup>22</sup> Environmental and Evolutionary Genomics, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS (UMR 8197) INSERM (U1024), Paris, France, <sup>23</sup> Tara Expeditions, 12 rue Dieu, 75001 Paris, France

### Abstract

Oceanic life, and particularly plankton, helps regulate Earth's climate and yet remains largely undescribed due to sampling often restricted on either specific environments or organismal groups. As climate changes planktonic ecosystems are being affected in ways that will feedback onto the biogeochemistry that drives Earth's climate engine. Here we present the sampling strategy and methods from the circum-global Tara Oceans Expedition (2009-2012) as a model for baseline planktonic ecosystem sampling. At sea, extensive environmental contextual data and organisms from viruses to planktonic animals (e.g. jellyfish, fish larvae) were collected, with sampling and on-land processing enabling an unprecedented integration of genomics, morphology and environmental data. Overall this uniquely exhaustive and integrated Tara Oceans sampling program provides a roadmap for future biology-centered ocean ecosystem studies, as well as a unique initiative upon which global scale ecological models predicting responses of oceanic ecosystems to environmental change can be developed.

**Keywords:** Plankton, biodiversity, engineering, global scale, Tara-Oceans

Since the H.M.S. Challenger expedition (1872-1876), exploration of the oceans have led to breakthroughs in geology (e.g. continental drift), oceanography (e.g. major oceanic currents), biogeochemistry (e.g. Redfield ratio), biodiversity (e.g. hydrothermal vents), and biology (e.g. aequorin and Green Fluorescent Protein from jellyfish). Modern expeditions are few but are gradually revealing the roles of oceanic life, particularly plankton, which regulate Earth's climate (Falkowski 2012). Such expeditionary discoveries are often triggered by technological innovations to improve sampling rates and spatio-temporal resolution of physico-chemical parameters. In contrast, plankton sampling has only minimally evolved since the introduction of Nansen bottles (1910), plankton nets (1928), and the Continuous Plankton Recorder (CPR, 1931) (Wiebe & Benfield 2003). Notably, biological sampling methods have not been adapted to keep up with modern methods of high throughput species identification and functional analysis that can provide large numbers of new morphologies and much molecular information. Plankton, from viruses to metazoans, spans several orders of magnitude in size and often have tight biotic interactions, necessitating complex analyses, leaving the field in need of integrated, comprehensive, sampling strategies to best capture plankton community composition and ecology.

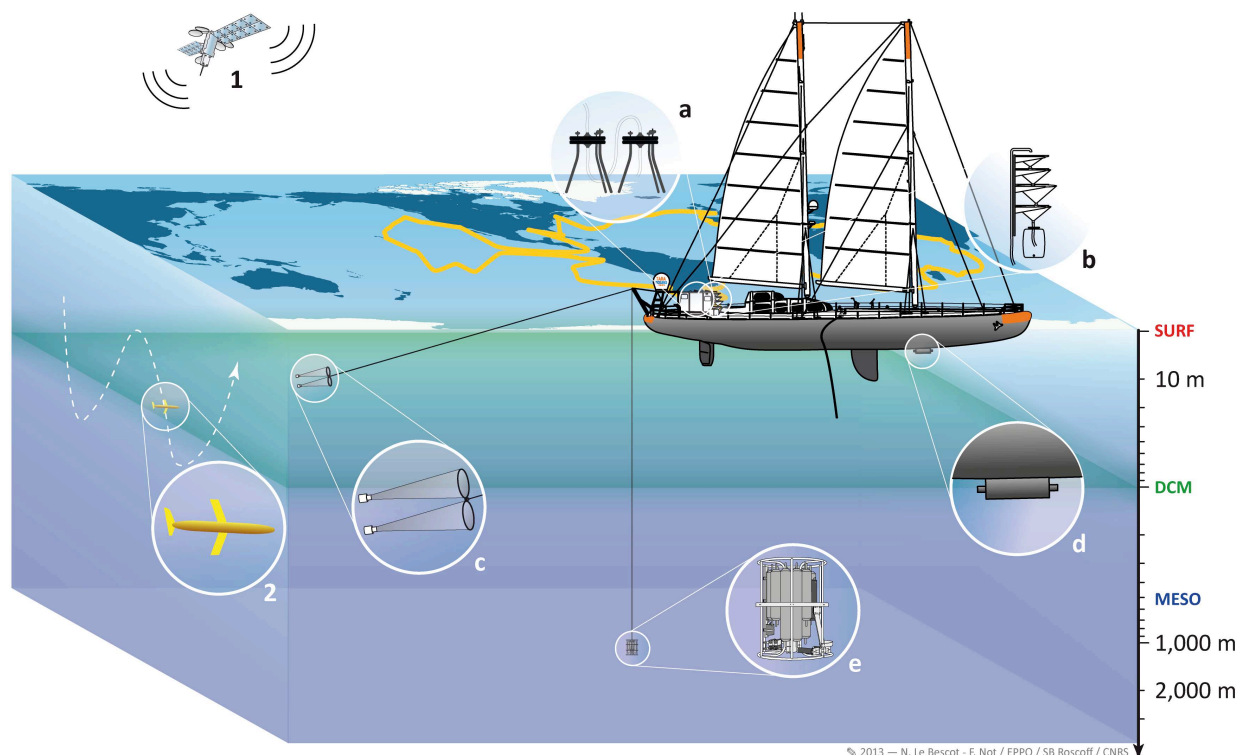


**Figure 1:** Tara Oceans cruise track (yellow) indicating the sampling station (red dots) and highlighting the topical experiments (light blue path and label) carried out during the expedition. Boundaries of Longhurst's biogeographical provinces are displayed in the background (Longhurst 2007). Detailed positions of station and station numbers are provided in supplementary figure 8 and supplementary .kml file (North Atlantic - 1 to 4 and 141 to 154, Mediterranean sea - 5 to 30, Red sea and Indian ocean - 31 to 66, south Atlantic and Antarctica - 67 to 89, south Pacific - 90 to 128, North Pacific - 129 to 140).

Global climate is changing at unprecedented rates, and with the oceans absorbing much of this change (warming and acidifying), plankton ecosystem structure and functioning will undoubtedly be impacted. Recent global ocean biodiversity surveys provide promising first steps towards developing a

baseline understanding of marine biota (<http://www.coml.org/>). Other surveys focus on a subset of organisms or environments. For instance, the Sorcerer II expedition (2004 - 2006) focused on surface prokaryotic metagenomes (Gross 2007), while the Malaspina expedition (2010-2011) emphasized diversity and process measurements in the dark bathypelagic ocean at 1,000 – 4,000m depth (Laursen 2011).

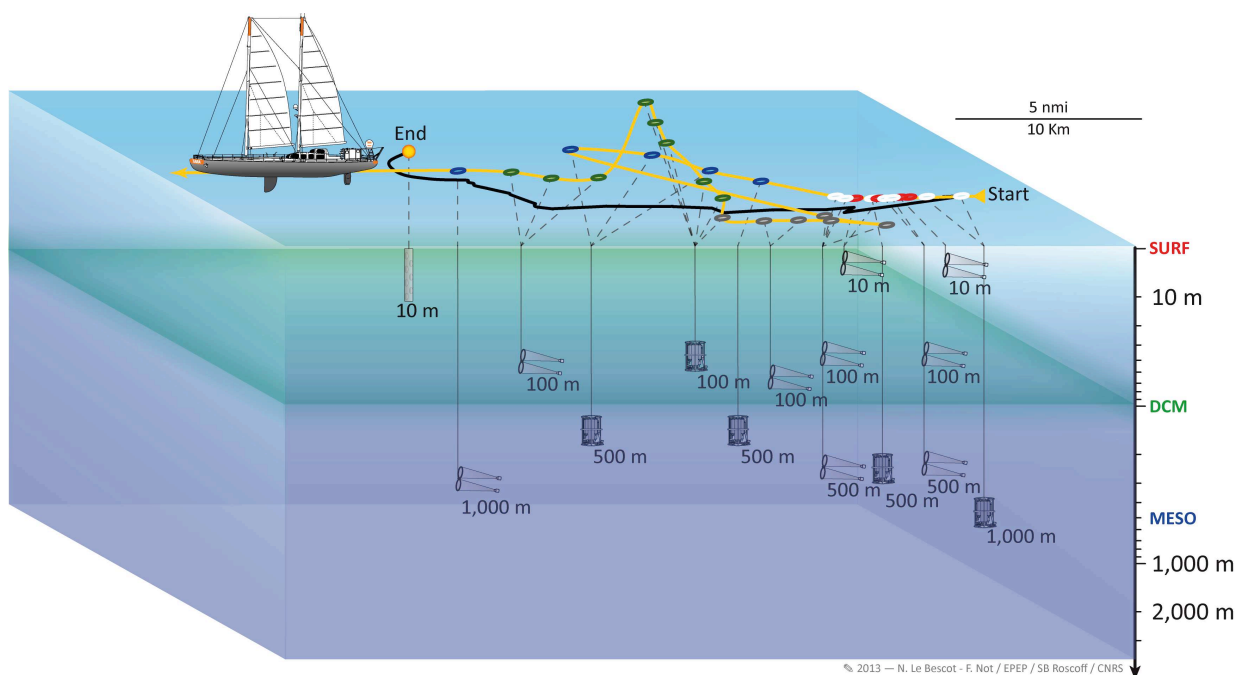
The Tara Oceans Expedition (2009-2012) was unique in several ways (Fig. 1): (i) use of a 36m long schooner, permitted a near pan-oceanic sampling at reasonable cost, although space and power were at a premium, (ii) it profiled the sunlit (0-200m) and mesopelagic (>400m) layers of the world oceans, (iii) organisms were targeted from viruses to the larger metazoans, (iv) modern sequencing technologies were combined with those for imaging, and (v), extensive contextual data were collected in association with biological samples. Together, this rather comprehensive and systematic sampling strategy should enable associations between biodiversity and function in marine planktonic ecosystem (Karsenti *et al.* 2011), a necessary step to integrate the biological complexity into predictive global-scale ecological models and management of oceanic ecosystems in response to environmental changes.



**Figure 2:** Schematic representation of the main sampling instruments used during the Tara Oceans Expedition, including (1) remote sensing devices to position the location of sampling stations, (2) in situ autonomous instruments (e.g. gliders), and on-board sampling devices: (a) peristaltic pump filtration pipelines, (b) Gravity Plankton Sieving System (GPSS) to concentrate plankton, (c) Plankton nets, (d) Underway continuous measurements on water pumped below the surface, (e) Niskin bottles and sensors mounted on a rosette. Depth of DCM and MESO sampling were variable and determined based on CTD profiles.

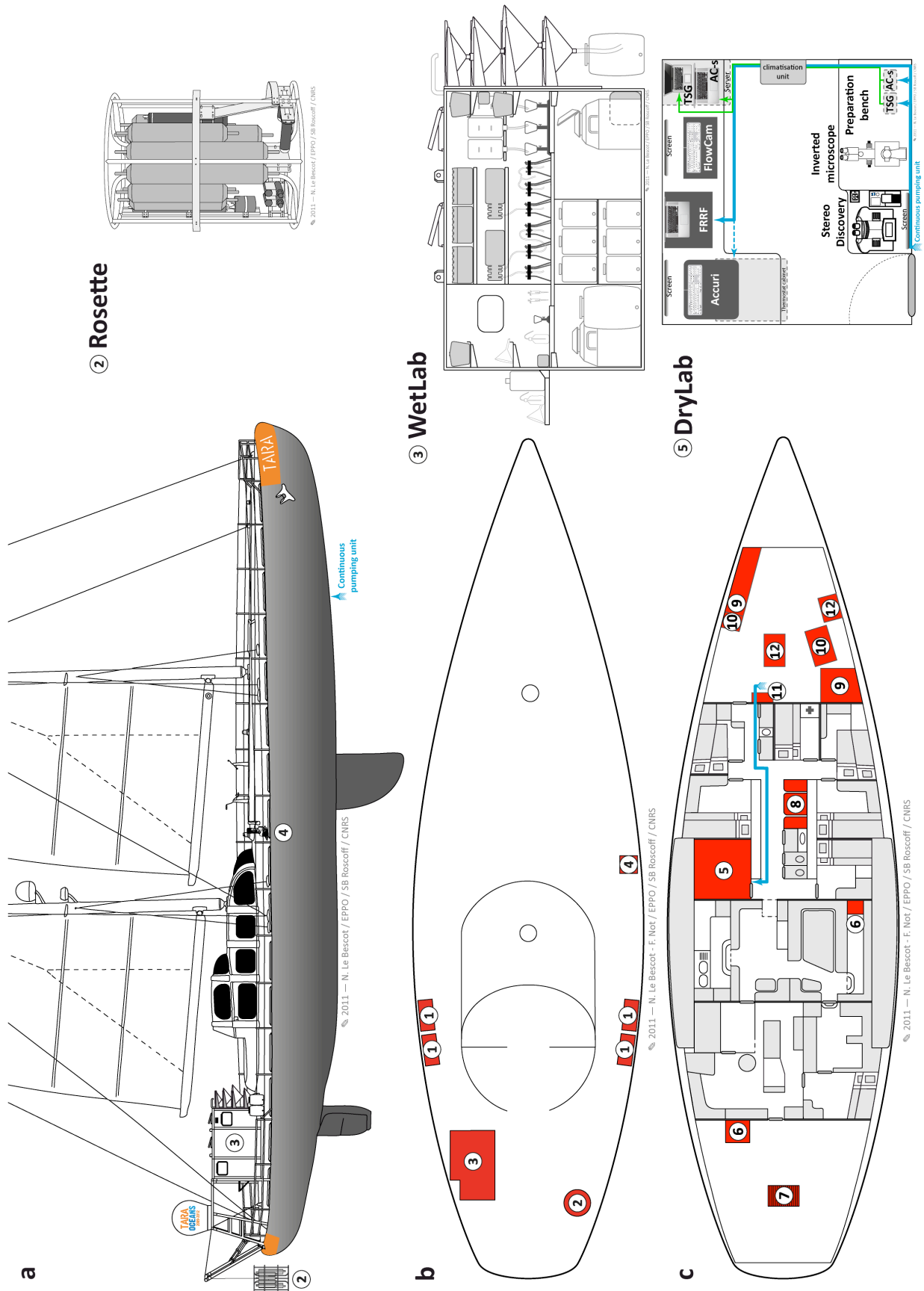
During the Tara Oceans Expedition, sampling was constrained by several factors such as sailing time between stop-overs and weather conditions. Sampling stations were carefully selected along the route to cover distinct marine ecosystems such as Longhurst's marine provinces (Longhurst 2007) or typical features such as Oxygen Minimum Zones (Fig. 1). The expedition was not designed to address temporal patterns, but rather to describe, with high data content, a variety of well-contextualized ecosystems. For each station we characterized the physical and biogeochemical context as precisely as possible using a suite of onboard and deployed sensors, while integrating also spatio-temporal information provided from satellites (Fig. 2). Within ecosystems, precise sampling locations were determined with respect to meso-scale features, such as eddies, fronts, or island wakes, so that all samples from a station were taken within a relatively homogeneous environment. Nonetheless local small spatial and temporal variability of key environmental parameters were assessed using continuous measurements on surface seawater pumped underway at  $\pm 1.5\text{m}$ , repeated deployments of sampling instruments, and at specific occasions using automated underwater vehicles (e.g. gliders (Tintore *et al.* 2010), ProvBio (Claustre *et al.* 2011)) and surface-tethered drifters (Fig. 2). Satellite data (Chlorophyll a, sea surface temperature, and altimetry) and real-time ocean model outputs (Mercator Ocean) were also used on a daily basis to revise sampling positions with respect to the selected oceanographic features. Weather permitting, each 48 hours long sampling station comprised exhaustive sampling at three depths of all 11 plankton size fractions for both morphological and meta-omics analyses (Fig. 3 and Fig. 5). For each sampling station, a full list of the samples collected and any associated information (e.g. depth, size fractions, etc...) is available at PANGAEA (<http://www.pangaea.de/>). Overall, each station generated a total of 370 distinct samples. In addition, specific scientific questions were addressed along the cruise track by performing topical experiments at relevant spatio-temporal scales (Fig. 1).

These ranged from the meso-scales (10's of km & hours/days), such as mapping of eddies, tracking of water masses, capturing nycthemeral dynamics and short-term perturbations, up to basin scales (1,000's of km & weeks/months), for example tracking Agulhas rings over the south Atlantic. Details for each of these topical experiments will be reported independently in specific contributions.



**Figure 3:** Chronology and spatial distribution of events during a station over 48 hours. The example presented corresponds to station 152 (Start at  $-016.8454^{\circ}\text{E}$ ,  $43.6850^{\circ}\text{N}$ , North Atlantic). The black track corresponds to the Argo drifter released at the beginning of the sampling procedures and the yellow line to the actual ship track. Along the ship track, red, green, and blue circles correspond to Surface, DCM, and Mesopelagic sampling, respectively. The white and grey circles represent deployment of gear to collect samples integrating the water column during day and night time, respectively.

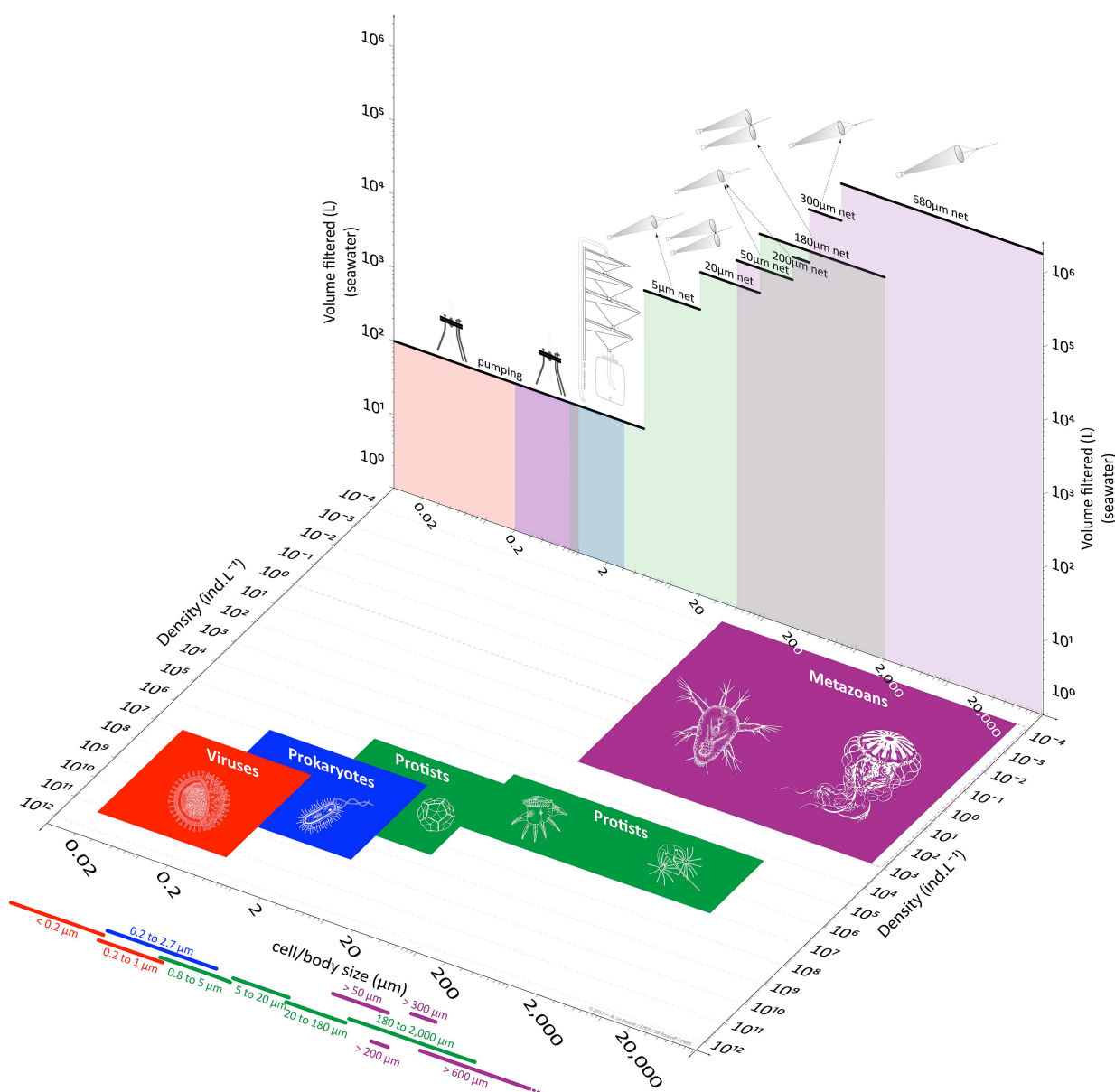
Alongside a crew of 5 sailors to run the schooner, a team of 6 scientists carried out the sampling operations. The scientific team rotated approximately every month and was composed of 1 chief scientist, 2 scientific engineers in charge of the deck operations, 2 scientists dedicated to the biological sampling, and 1 scientist in charge of the dry lab operations and imaging. The sampling program (detailed protocols in supplementary information) aimed to collect organisms ranging in size from viruses to fish larvae (i.e. viruses, prokaryotes, protists, and metazoans). In addition to classical biological oceanography sampling methods such as Niskin bottles casts and net deployments, new procedures were developed to match the requirements of state-of-the-art analytical processes. Working areas on board were specifically designed, installed, and organized according to the working constraints (Fig. 4). Storage and deployments of oceanographic instruments were performed from the deck, at the stern of the ship. A wet laboratory (WetLab) equipped with all filtration devices was designed and installed on the port side of the ship and a large peristaltic pump to sample surface and DCM water on starboard (Fig. 4 a, b). A winch was installed below deck under the central working platform near to the stern. A dry lab was setup inside on portside at the center of the ship and contained imaging and flow through instruments linked to a water flow taken underway from the front part of the boat (see details in supplementary information). The storage area was located in the forward hold (Fig. 4c).



**Figure 4:** Working areas distribution on-board Tara. These are shown in red on (a) side view, (b) bird eye view of the deck, and (c) inside view of the ship. Working areas include (1) storage boxes; (2) CTD-Rosette; (3) wet lab; (4) Peristaltic pump for large-volume sampling; (5) Dry lab; (6) Operational oceanography desks and computers; (7) Winch; (8) Mobile imaging area; (9) Room temperature storage areas; (10) Storage areas at +4°C and -20°C; (11) MilliQ water system (12) Diving gears. Water flow for continuous recording instruments is shown in light blue.



Based on knowledge of the abundance of organisms within each size fraction sampled, we used a variety of sampling devices to collect appropriate volume of seawater, ranging from Niskin bottles, to direct in situ pumping as well as plankton nets with mesh sizes ranging from 5µm up to 680µm for larger volumes (Fig. 5). The volumes of water sampled were also scaled across size-fractions in order to collect most of the diversity, based on species richness data currently available in the literature, for each of the size class considered (Supplementary Fig. 1). Whenever possible, samples were collected in duplicates for long-term storage, eventually allowing re-analysis using future technologies. This might be particularly relevant considering the exponentially growing power of -omics methods.



**Figure 5:** Size fractionation and seawater volumes used during the expedition. The plot shows the size distribution of planktonic organisms (X axis) and their range of abundance in 1L of seawater based on a literature survey (Y axis). For each of the size fractions considered, average volumes actually sampled during the expedition are represented (Z axis). It illustrates the match between densities of organisms in environmental samples and volume of water collected (e.g. metazoans exhibit the lowest densities, consequently larger volumes of water were sampled). The numerical data used to draw this plot are listed in supplementary Table 2.



Physico-chemical data are key parameters that provide an environmental context for the biological samples collected. A 12-bottle CTD-rosette was adapted to the size of Tara's A-frame and comprised 6 x 8L and 4 x 12L. In addition, it included state-of-the-art sensors: A Seabird 911 + CTD with two pairs of conductivity and temperature sensors, the ISUS (In Situ Ultraviolet Spectrophotometer) Nitrate sensor, oxygen sensor, fluorometers for chlorophyll and colored dissolved organic matter, transmissiometer (particles <5 $\mu$ m), back scattering (particles <15 $\mu$ m), and the Underwater Vision Profiler (Guidi *et al.* 2010) for particles >100 $\mu$ m. During the cruise a large set of parameters were also recorded using deployed instruments, underway devices, or autonomous sensors (Supplementary Table 3). An additional stand-alone Seabird 911 + CTD was mounted on a frame with a fluorometer and an oxygen sensor for use during harsh sea conditions when the deployment of the rosette was not safe. Downwelling irradiance and upwelling radiance were measured by a Hyperspectral Tethered Spectral Radiometer Buoy (TSRB) that was deployed for several minutes near midday when conditions were favorable (i.e. minimum cloudiness). A WetLab ac-S spectrophotometer was installed in flowthrough mode in the dry lab and provided (using an automated switching interval to measure dissolved properties) an estimate of the particulate absorption and attenuation coefficients (Tintore *et al.* 2010). Three proxies of biogeochemical properties were thereby derived; chlorophyll a concentration via the absorption line-height at 676nm, particulate organic carbon and a small-particle size parameter. Finally, a SeaBird SBE45 MicroTSG unit was also installed in flowthrough mode in the dry lab providing temperature and salinity data during transit between stations. CTD and optical sensors were calibrated at the manufacturer at least once a year.

A continuous recording of photosynthetic efficiency at sub-surface was performed by using a Fast Repetition Rate Fluorescence (FRRF) instrument installed in the dry lab. At each station, samples were collected from Niskin bottles for carbonate chemistry, nutrients, photosynthetic pigments, and on-board flow cytometry. Parameters of the carbonate system (dissolved inorganic carbon and total alkalinity) are useful to characterize the ecosystems with respect to ocean acidification. Nutrients gave us information about the primary production potential of the ecosystem, whilst pigment concentrations measured by HPLC were used in parallel to imaging and genomic methods to determine dominant phytoplankton taxa as well as to help validating and interpreting spectral measurements from satellites (Supplementary Fig. 2). A total of 6, 12, and 11 samples were consistently generated at each main sampling location for carbonates, nutrients, and pigments, respectively.

Ocean viruses are ubiquitous and abundant, and impact global biogeochemistry through mortality, horizontal gene transfer, and modulating microbial metabolism (Breitbart 2012). Here we describe a systematic, quantitative viral metagenomic sample-to-sequence pipeline tuned for small ship and tight sampling schedule (Duhaime & Sullivan 2012). This pipeline was evaluated using

replicated metagenomic experimental designs to optimize viral concentration and purification, DNA amplification, and downstream library processing and metagenomic sequencing platform choices (Solonenko *et al.* 2013).

To complement viral metagenomic sampling, ancillary samples were collected. First, two sets of viral size fraction (i.e.  $<0.22\mu\text{m}$ ) samples were collected for each qPCR analysis to target genes of interest (stored frozen), and virus cultivation and the new ‘viral tagging’ method which links viruses to hosts in a high-throughput manner (Yilmaz *et al.* 2012) (stored at  $4^{\circ}\text{C}$  in the dark to minimize damage to infectious viral particles). Second,  $20\mu\text{m}$  filtrate seawater samples were collected for total viral particle counts (SYBR-Gold staining and epifluorescence microscopy (Noble & Fuhrman 1998), morphological investigation of virus diversity using quantitative transmission electron microscopy, and lineage-specific assessment of viral infected cells using the newly developed phage FISH protocol (Amann *et al.* 2013). Third, “whole seawater” (i.e.  $200\mu\text{m}$  pre-filtered) samples were collected for flow cytometry preserved in either DMSO or glutaraldehyde to maximize the sample archives for capturing a diversity of microbes for downstream counting and genomics analyses. Overall, 8 distinct sampling procedures were carried out for sampling viruses, generating a total of 4,204 samples over the expedition.

While small viruses are commonly abundant, the presence of larger “giant viruses” or giruses is increasingly recognized in marine samples (Claverie *et al.* 2008; Hingamp *et al.* 2013). They are thought to target diverse eukaryotic hosts including microalgae and heterotrophic protists, and are likely impacting their host’s mortality and evolution (Subirana *et al.* 2011). To characterize the diversity and ecological roles of giruses, finely size fractionated samples were collected from 200L seawater per depth for deep metagenomic sequencing and PCR amplification of marker genes, using serial filtration devices ( $0.1\mu\text{m}$ - $0.2\mu\text{m}$ ,  $0.2\mu\text{m}$ - $0.45\mu\text{m}$  and  $0.45\mu\text{m}$ - $0.8\mu\text{m}$ ), alongside with additional samples for viral isolation in host cultures in the laboratory. Twenty-one samples were collected at each station (see supplementary information and supplementary Fig. 3 for details).

Prokaryotes are a fundamental component of the ocean’s food web and biogeochemical processes; they are responsible for 30% of the primary biomass production and 95% of the respiration of the ocean (Duarte & del Giorgio 2002). The majority of prokaryotes in the oceans are uncultivable and are connected in a network of planktonic organisms that range from viruses to zooplankton. In this expedition, we collected prokaryotic cells in two main size fractions: the  $0.22$ - $3\mu\text{m}$  fraction, mostly containing what is referred to as free-living prokaryotes and the  $3$ - $20\mu\text{m}$  fraction where prokaryotes are expected to be found either attached to larger cells, particles, or in aggregates (Acinas *et al.* 1999).

Samples to extract prokaryotic DNA for subsequent metagenomic investigations were collected regularly at 3 depths from 100L of seawater. RNA for metatranscriptomic studies was collected at the same depths using a reduced seawater volume of  $2 \times 10\text{L}$  in order to restrict the filtration time to a

maximum of 10-15min which is critical considering the rapid turnover of mRNA (Moran *et al.* 2012). Total abundances were estimated from samples specifically collected for flow cytometry and microscopy analysis. Quantification of specific taxa was achieved by Fluorescent In Situ Hybridization (FISH). Finally, samples for culturing and single cell genome sequencing were also collected and kept in DMSO and glycine betaine, respectively. A total of 54 samples were consistently generated at each sampling location (see supplementary information and supplementary Fig. 3 for details).

Unicellular eukaryotes, or protists, play primary and largely unexplored roles in functional ecology of marine plankton and biogeochemistry. Oceanic protists are very diverse, having representatives in all of the 8 super-groups of the eukaryotic tree of life and they encompass a broad size range. Given the fundamental knowledge-gap on marine protistan biodiversity, we developed a complex morpho-genetic pipeline including samples from 4 protistan size fractions, total DNA and RNA, and cell concentrates preserved in a series of buffers for subsequent morphological and genetic analyses. Samples were collected using either peristaltic pumping for the small size fraction (0.8-5 $\mu$ m) or plankton nets for the bigger size-fractions (5-20 $\mu$ m, 20-180 $\mu$ m, and 180-2,000 $\mu$ m), thus encompassing diversity across 3 orders of size magnitude (Fig. 5). Protistan total biomass (minimum of 100L for the small size fraction) from 3 depths was filtered onto polycarbonate membranes, flash-frozen, and stored in Liquid Nitrogen for subsequent high-throughput sequencing of protistan meta-barcode, meta-genomes, and meta-transcriptomes. For the largest size fractions, i.e. 20-180 $\mu$ m and 180-2,000 $\mu$ m, samples were fixed with EtOH for single cell morpho-molecular identification and potentially single-cell genomics. Plankton samples were also preserved in Lugol, formal, and Glutaraldehyde for morphological analysis using classical light and electron microscopy, as well as future development in confocal High-Throughput Fluorescent Microscopy (eHTFM), allowing to explore cellular features at the single-cell level. Samples were also preserved for Liquid FISH analysis that allows linking molecular identification of plankton diversity with quantitative morphological information. Further collection of live protist samples was carried out for onboard analysis by FlowCAM, and observation using stereo binocular or inverted microscope.

Arguably the most difficult fraction to collect and preserve is the 5-20 $\mu$ m, in which organisms are too diluted to be collected from non-concentrated seawater and too fragile to be accurately concentrated using nets. A compromise was reached between cell density and collection procedure. Whenever sea conditions allowed, a 5 $\mu$ m net was deployed to gently concentrate biomass of these planktonic cells. In harsh sea conditions a newly designed on-deck Gravity Plankton Sieving System (GPSS), consisting of a series of nets with decreasing mesh-size (Fig. 2b) was used to concentrate material from pumped water. For the smaller size-fractions (0.8-5 $\mu$ m and 5-20 $\mu$ m), samples were also collected and preserved in appropriate buffer for single cell genomics, FISH analysis, or SEM observations. Samples of total (not size fractionated seawater) were also collected for later meta-

omics and flow cytometry analyses, as well as specific microscopy techniques (e.g. automatic recognition of coccoliths by dynamical neural networks, SYRACO). Mesopelagic sampling was similar to epipelagic sampling, except that protocols were adjusted to the expected lower density of organisms and limited volume of water available from Niskin bottles meaning that size fractionation was restricted. Overall a total of 195 protistan samples were consistently generated at each sampling location (see supplementary information and supplementary Fig. 4 for details).

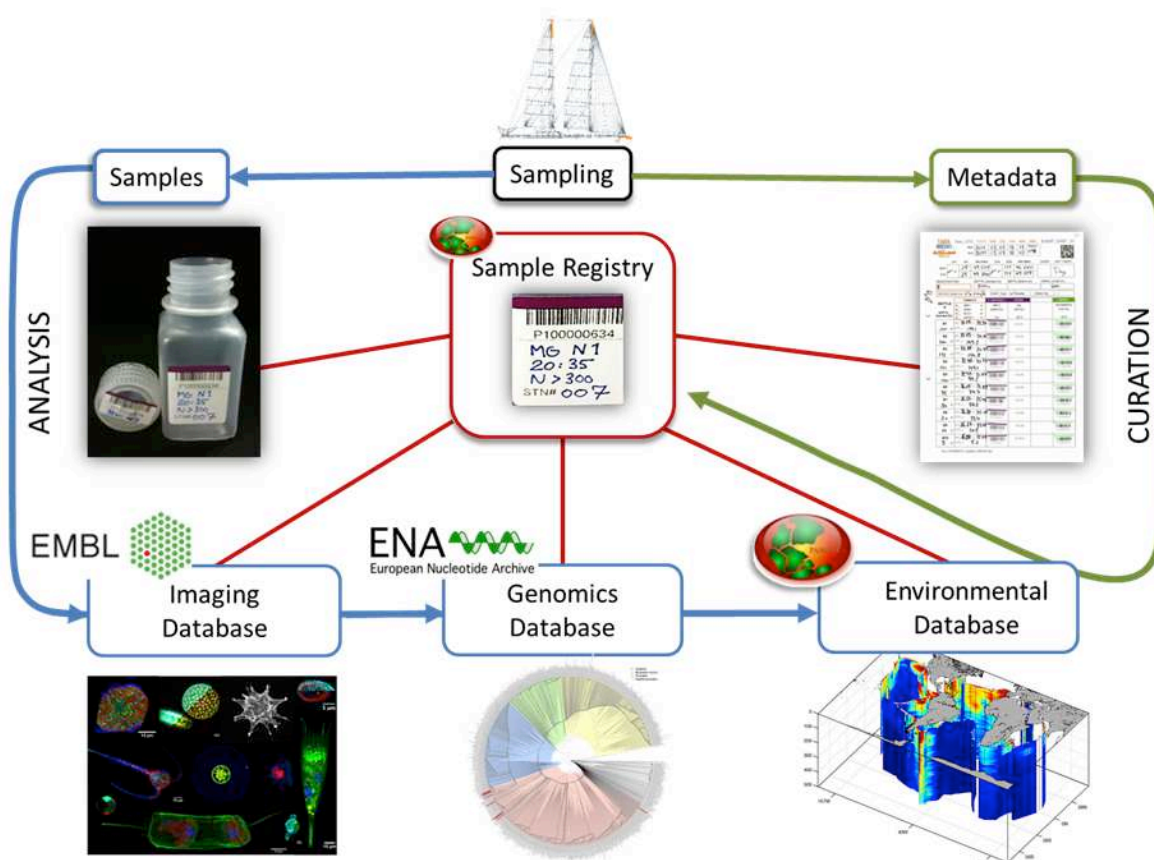
Zooplankton plays a pivotal role in both the transfer of energy to the higher trophic level such as fish and in the vertical export of particulate matter produced at the surface of the ocean (Banse 2013). Biodiversity and body-size are important properties affecting these two processes (Stemmann & Boss 2012). During the Tara Oceans Expedition, large planktonic organisms (both metazoans and large protists) were collected using nets of different mesh porosity (50 $\mu$ m, 200 $\mu$ m, 300 $\mu$ m, and 680 $\mu$ m), which target organisms of different sizes (Fig. 5). These nets were towed from various depths both during day and night in order to capture the nycthemeral vertical migrations. In addition a HYDROBIOS multinet (deployed whenever the sea state allowed) and the Underwater Vision Profiler mounted on the rosette were used to study the vertical distribution (down to 1,000m) of zooplankton >0.5mm. Zooplankton samples were preserved in ethanol, formol, and RNAlater for further morpho-genetic analyses ranging from traditional microscopy to quantitative imaging (ZooScan (Stemmann *et al.* 2010)) and from single specimen molecular identification, all the way to metagenomics. A total of 15 samples were consistently generated at each sampling location (see supplementary information and supplementary Fig. 5 for details). Besides these preservation procedures, regular collection of live samples was carried out for onboard observation using stereo binocular or inverted microscope and FlowCAM analysis.

The extensive sampling programme and the reduced workspace on-board required a robust system to label samples and to keep a thorough record of sampling conditions and of any deviation from standard protocols. The system also needed to be flexible, especially in the case of topical experiments that often led to exceptional sampling programmes and protocols. Our management system consisted of a series of log sheets designed as “bingo” cards on which scientists affix stickers with unique barcodes (Supplementary Fig. 6). The lines and columns of the log sheets corresponded to the various protocols, sampling depths and replicates. Log sheets included methodological details for each protocol (i.e. volume filtered, chemicals added, storage conditions), so that they were also used during sampling as a quick guide. Any deviation from standard protocols was recorded manually on the log sheets.

Barcode stickers had a colour band indicating the taxonomic group and shipping destination of samples. Stickers were made of BRADY<sup>®</sup> B-492 “Freezerbondz” white polyester, and those used for samples stored in liquid nitrogen had a long laminating band made of BRADY<sup>®</sup> B-461 transparent polyester. Both materials offered excellent adhesive-resistance to solvent and temperature-resistance

to liquid nitrogen or autoclaving. Barcodes were printed using thermal transfer ink ribbons (BRADY® R6400) that offered excellent print smudge resistance to most solvents. Barcode stickers affixed on the log sheets matched those affixed to the sample containers (i.e. vials, tubes, bottles, Petrislides). While stickers affixed to the log sheets were just large enough to print a barcode, those affixed to the sample had additional space to write the station number, protocol abbreviation, size fraction, sampling depth and in some cases the sampling time.

One goal of the Tara Oceans expedition is to generate online, open access datasets for present and future generations of scientists. Data management was carefully tailored to meet this goal by selecting permanent and authoritative data archives in the fields of genomics (European Nucleotide Archive; <http://www.ebi.ac.uk/ena/>), imaging (EMBL, <http://www.embl.de/>) and environmental science (PANGAEA, Data publisher for Earth and Environmental Science; <http://www.pangaea.de/>, and CORIOLIS, in situ data for operational oceanography; <http://www.coriolis.eu.org/>). In addition, some data will be further disseminated to specialised data portals such as OBIS for species biogeography (<http://www.iobis.org/>), and SeaDataNet for environmental data (<http://www.seadatanet.org/>). Sample barcodes act as unique identifiers and are the reference to link metadata from the log sheets, samples stored in the different laboratories, and data archives (Fig. 6). The registry of all stations, sampling events, and samples are available online at <http://www.pangaea.de/>.



**Figure 6:** Summary of the sample and data workflow showing data analysis and safeguarding outlined in blue, and metadata curation in green. The central sample registry, in red, is the key to linking samples (e.g., jars and tubes), data archived in a distributed network of databases, and metadata about sampling and analysis methodology.

Samples were stored onboard according to their respective storage temperature either in liquid nitrogen, freezer (-20°C), fridge (+4°C) or at room temperature. In particular the boat was equipped with three 30L liquid nitrogen dewar tanks, two of which were used for sample storage, while the third was used to re-fill the other two as needed. The storage capacity of the freezer and fridge were 392 and 360 liters, respectively. Due to the size of the boat the storage capacity was limited, which affected the sampling strategy and required a rigorous inventory system. The storage capacity was sufficient to contain samples from 8 to 10 stations and the shipping ports were carefully selected accordingly, also considering the customs regulations of countries along the route and the available infrastructures such as airports equipped with cooling facilities, storage warehouses etc. All samples were packed by a team of specialists (World Courier) who organized transportation to and from each shipping port. Samples were shipped in special temperature pre-conditioned containers to ensure that the cold chain was never interrupted. Frozen samples were packed on dry ice. Samples arrived in Europe (Frankfurt airport) within 36 hours after offloading from the vessel. Samples were immediately transferred to a temperature controlled storage warehouse (World Courier). Shortly after arrival, the logistics team unpacked, sorted and repacked the samples. These were then dispatched to the participating laboratories in France (Paris, Roscoff, Banyuls-sur-Mer, Marseille and Villefranche-sur-Mer), Spain (Barcelona), and the U.S.A. (Bigelow, Maine and Tucson, Arizona) where they are currently stored (Supplementary Fig. 7).

Overall, the Tara Oceans Expedition explored thoroughly 154 stations and generated nearly 28,000 biological samples for molecular (29%) and morphological (71%) analysis according to the above procedures. Complementing these biological samples are (i) 681 CTD profiles, (ii) ~1,500,000 in situ images including particle size spectra down to a maximum of 1,200m, (iii) ~310,000min averaged spectra of particulate absorption and attenuation, (iv) 62 independent remote-sensing reflectance spectra and (v) 4,400 biogeochemical samples for nutrients (42%), pigments (36%), and carbonate chemistry (22%). The Tara Oceans plankton sampling pipeline offers a starting point of standard integrated procedures for next global oceanographic sampling programs seeking value-added synergism through high inter-comparability with past, current, and future studies. Beyond global “snapshot” surveys future work will need to focus on understanding variations over smaller and larger time and space scales through (inter)national initiatives such as the Integrated Marine Biogeochemistry and Ecosystem Research (IMBER) or Global Ocean Observing System (GOOS) programs. Ultimately, provided a minimal standardization of the procedures allowing efficient merging of datasets, both approaches can be combined to provide a comprehensive understanding of life in the oceans and its impact on ecosystems structure and function.

Accurate descriptions of natural systems using interdisciplinary approaches have historically paved the way for discovery and towards conceptualization of new theories and laws of Nature (Darwin 1859; Newton 1972). The challenge for modern science is to overcome the incongruence of

specialized scientific knowledge and disconnected data sets. The integrated end-to-end Tara Oceans sampling strategy presented here strives to provide such interdisciplinary initiative in an oceanographic context. Future analyses of data from these samples will likely contribute to address global issues from a holistic perspective.

---

### Acknowledgments:

This research was supported by Région Bretagne, CNRS GDR3280 Tara Oceans, EMBL, CEA, Université Pierre et Marie Curie, Cap Lorient, Foundation Veolia Environment, ACRI-ST, Mercator Ocean, NASA, ESA, Carl Zeiss Microscopy GmbH, Fluid Imaging Inc., Science Foundation Ireland, and by the following grants: ANR POSEIDON (ANR-09-BLAN-0348), ANR TaraGirus (ANR-09-PCS-GENM-218), BACTERIOMICS (CTM2010-12317-E), MicroOcean PANGENOMICS (CGL2011-26848/BOS), Biosphere 2, BIO5, the Gordon and Betty Moore Foundation, the US NSF grant OCE0940390, and the program OCEANOMICS funded by the French government under the grant agreement "Investissement d'Avenir" ANR-11-BTBR-0008. We are also grateful to all members of the Tara crew and the Tara office team supporting the operations from Paris (France). In addition we would like to thank the following members of the consortium for their contribution: Abergel C.<sup>3,11</sup>, Arendt D.<sup>XX</sup>, Arslan D.<sup>3,11</sup>, Audic S.<sup>3,5,6</sup>, Aury J.M.<sup>15</sup>, Babic N.<sup>3,5,6</sup>, Beaufort L.<sup>3,11,23</sup>, Benzoni, F.<sup>XX</sup>, Bork, P.<sup>XX</sup>, Bourmaud, H.<sup>XX</sup>, Boutte C.<sup>3,5,6</sup>, Bricaud A.<sup>XX</sup>, Casotti R.<sup>14</sup>, Chaffron, S.<sup>XX</sup>, Chambouvet A.<sup>25</sup>, Chang P.<sup>3,6,13,26</sup>, Chica C.<sup>15</sup>, Claverie, J.M.<sup>XX</sup>, Cornejo-Castillo F.M.<sup>2</sup>, Da Silva C.<sup>15</sup>, De Monte S.<sup>3,4</sup>, Deng, L.<sup>XX</sup>, Desdevises Y.<sup>3,6,27</sup>, Duhaime M.B.<sup>XX</sup>, Dolan J.<sup>3,6,13</sup>, Duhaime M.<sup>9</sup>, Durrieu de Madron X.<sup>3,28</sup>, d'Ovidio F.<sup>6,29</sup>, Follows, M.<sup>XX</sup>, Garczarek L.<sup>3,6,30</sup>, Garet-Delmas M.J.<sup>3,5,6</sup>, Gasmi S.<sup>31</sup>, Gasol J.M.<sup>2</sup>, Grisoni, J.M.<sup>XX</sup>, Heilig R.<sup>15</sup>, Hurwitz, B.L.<sup>XX</sup>, Ignacio-Espinoza J.<sup>9</sup>, Jamet J.L.<sup>32</sup>, Katinka M.<sup>15</sup>, Khalili H.<sup>15</sup>, Kolber Z.<sup>33</sup>, Leymarie, E.<sup>XX</sup>, Lima-Mendez G.<sup>7,8</sup>, Mahé F.<sup>3,5,6</sup>, Marien, O.<sup>XX</sup>, Mazzocchi M.G.<sup>14</sup>, Montresor M.<sup>14</sup>, Morard, R.<sup>XX</sup>, Morin P.<sup>3,6,30</sup>, Noel B.<sup>15</sup>, Obolensky, G.<sup>XX</sup>, Ormerstad M.<sup>XX</sup>, Pedros-Alio C.<sup>2</sup>, Pelletier E.<sup>15</sup>, Perez Y.<sup>31,35</sup>, Piganeau G.<sup>3,6,27</sup>, Poirot O.<sup>3,11</sup>, Poulos, B.T.<sup>XX</sup>, Prejger F.<sup>3,6,13</sup>, Prihoda J.<sup>3,4</sup>, Rampal J.<sup>31</sup>, Reverdin G.<sup>3,6,34</sup>, Röttinger E.<sup>XX</sup>, Rouviere C.<sup>3,6,13,26</sup>, Samson G.<sup>15</sup>, Santini S.<sup>3,11</sup>, Sciandra A.<sup>3,6,13</sup>, Siano, R.<sup>XX</sup>, Solonenko S.<sup>9</sup>, Tanaka A.<sup>3,4</sup>, Thompson A.<sup>12</sup>, Tichanne-Seltzer V.<sup>3,11</sup>, Tirichine L.<sup>3,4,27</sup>, Toulza E.<sup>3,6,27</sup>, Troublé, R.<sup>XX</sup>, Valette, L.<sup>XX</sup>, Velayoudon D.<sup>XX</sup>, Veluchamy A.<sup>3,4</sup>, Weissenbach J.<sup>XX</sup>, Zingone A.<sup>14</sup>

This is Tara Oceans' publication #XXX

## References

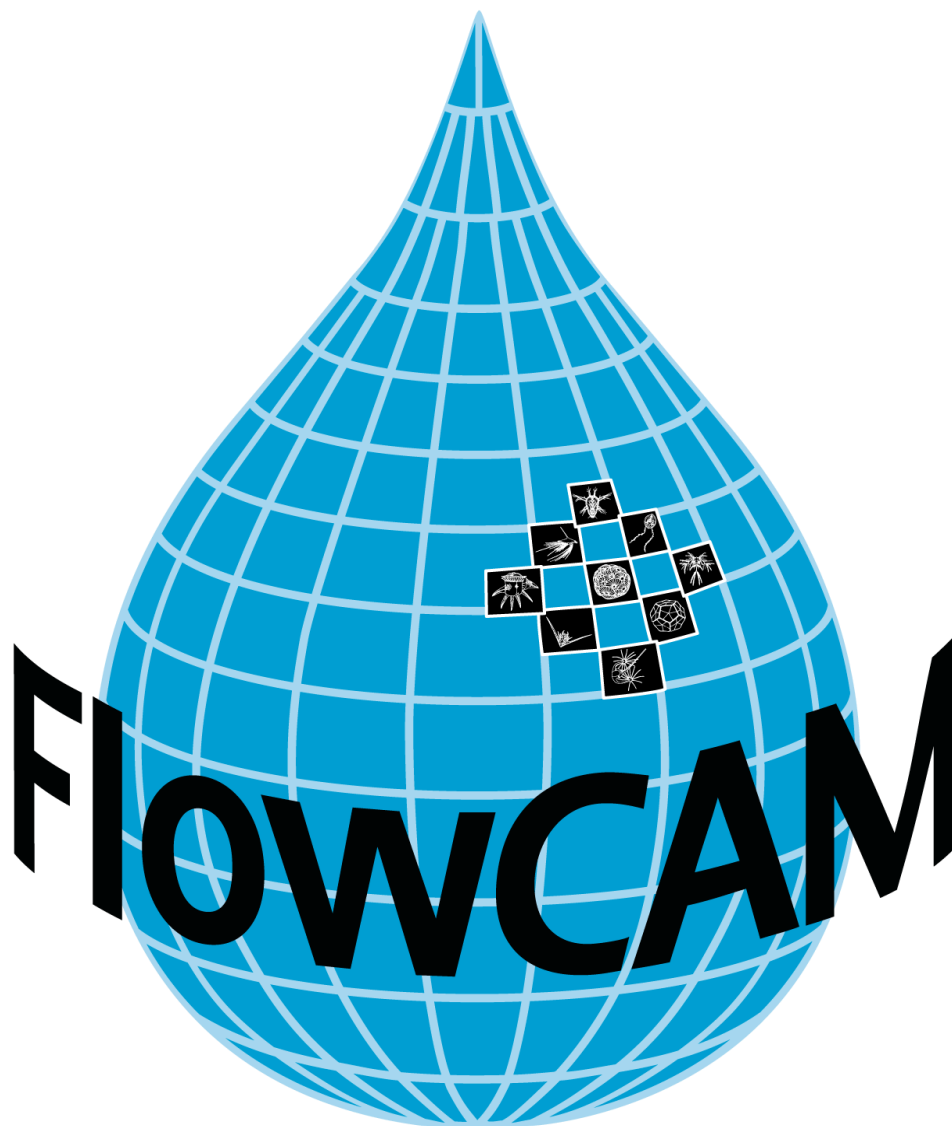
- Acinas SG, Antón J, Rodríguez-Valera F (1999) Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Environmental Microbiology*, **65**, 514–522.
- Amann R, Allers E, Moraru C *et al.* (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology*, **15**, 2306–2318.
- Banse K (2013) Reflections About Chance in My Career, and on the Top-Down Regulated World. *Annual Review of Marine Science*, **5**, 1–19.
- Breitbart M (2012) Marine Viruses: Truth or Dare. *Annual Review of Marine Science*, **4**, 425–448.
- Claustre H, Xing X, Morel A *et al.* (2011) Combined processing and mutual interpretation of radiometry and fluorimetry from autonomous profiling Bio-Argo floats: Chlorophyll a retrieval. *Journal of Geophysical Research*, **116**, C06020.
- Claverie J-M, Monier A, Ogata H (2008) Taxonomic distribution of large DNA viruses in the sea. *Genome Biology*, **9**, R106.
- Darwin C (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. New York: D Appleton.
- Duarte CM, del Giorgio PA (2002) Respiration in the open ocean. *Nature*, **420**, 379–384.
- Duhaime MB, Sullivan MB (2012) Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology*, **434**, 181–186.
- Falkowski PG (2012) Ocean Science: The power of plankton. *Nature*, **483**, S17–S20.
- Gross L (2007) Untapped Bounty: Sampling the Seas to Survey Microbial Biodiversity. *PLoS Biology*, **5**, e85.
- Guidi L, Picheral M, Stemmann L *et al.* (2010) The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology And Oceanography*, **8**, 462–473.
- Hingamp P, Acinas SG, Grimsley N *et al.* (2013) Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal*, **7**, 1678–1695.
- Karsenti E, Sullivan MB, Acinas SG *et al.* (2011) A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biology*, **9**, e1001177.
- Laursen L (2011) Spain's ship comes in. *Nature*, **475**, 16–17.
- Longhurst AR (2007) *Ecological geography of the sea*. Academic Press, New York.
- Moran MA, Satinsky B, Gifford SM *et al.* (2012) Sizing up metatranscriptomics. *The ISME Journal*, **7**, 237–243.
- Newton I (1972) *Philosophiae Naturalis Principia Mathematica* (mathematical principles of natural philosophy). London (1687).
- Noble RT, Fuhrman JA (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology*, **14**, 113–118.
- Solonenko SA, Hallam SJ, Ignacio-Espinoza JC *et al.* (2013) Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics*, **14**, 320.
- Stemmann L, Boss E (2012) Plankton and Particle Size and Packaging: From Determining Optical Properties to Driving the Biological Pump. *Annual Review of Marine Science*, **4**, 263–290.



- Stemmann L, Gorsky G, Ohman MD *et al.* (2010) Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research*, **32**, 285.
- Subirana L, Grimsley N, Thomas R *et al.* (2011) Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. *Environmental Microbiology*, **13**, 1412–1420.
- Tintore J, Testor P, Meyers G *et al.* (2010) Gliders as a component of future observing systems. **2**.
- Wiebe PH, Benfield MC (2003) From the Hensen net toward four-dimensional biological oceanography. *Progress in Oceanography*, **56**, 7–136.
- Yilmaz S, Deng L, Sullivan MB *et al.* (2012) Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging. *mBio*, **3**, e00373–12.

## Chapitre 2

Analyse morphologique par imagerie à haut débit  
des dinoflagellés marins



# Sommaire

|  |           |
|--|-----------|
| <b>Chapitre 2 : Analyse morphologique par imagerie à haut débit des dinoflagellés marins</b> | <b>69</b> |
| <b>1. Introduction</b>   | <b>71</b> |
| 1.1. <i>Histoire de l'étude des communautés planctoniques : l'enjeu de l'imagerie</i>        | 71        |
| 1.2. <i>Les contraintes de l'analyse d'image à haut débit</i>                                | 73        |
| 1.3. <i>Objectifs</i>  | 74        |
| <b>2. Matériel et méthode</b>  | <b>75</b> |
| 2.1. <i>Présentation et caractéristiques du dispositif FlowCAM</i>                           | 75        |
| 2.2. <i>Présentation et caractéristiques du logiciel Zoo/PhytoImage</i>                      | 76        |
| 2.3. <i>Méthodologie à bord de Tara</i>  | 78        |
| 2.4. <i>Traitement des données en laboratoire</i>  | 78        |
| 2.5. <i>L'analyse du genre Ceratium</i>  | 80        |
| <b>3. Résultats et discussion</b>  | <b>81</b> |
| 3.1. <i>Limitation technique de l'analyse</i>  | 81        |
| 3.2. <i>Analyse des dinoflagellés identifiables par FlowCAM/ZooPhytoImage</i>                | 82        |
| <b>4. Conclusion</b>   | <b>83</b> |
| <b>Références</b>  | <b>85</b> |

## 1. Introduction

Le plancton a de nombreux effets sur l'environnement, notamment à l'échelle globale (et ce, indépendamment de la petite taille des organismes qui le constitue). La composition de la communauté planctonique a une incidence sur les moyens par lesquels les micro-organismes qui la composent interagissent avec leur environnement. Le plancton peut varier rapidement en terme d'abondance et de biomasse en fonction des variations des conditions environnementales. La composition taxonomique, la distribution et l'abondance des groupes planctoniques sont ainsi des paramètres fondamentaux pour définir la structure et les mécanismes de l'écosystème. Pour bien comprendre l'hétérogénéité de la distribution du plancton, il est donc nécessaire d'étendre un maximum la résolution spatiale et temporelle des observations et de l'échantillonnage.

### *1.1. Histoire de l'étude des communautés planctoniques : l'enjeu de l'imagerie*

L'analyse de la communauté planctonique a commencé lorsque les techniques de microscopie à champ de lumière ont fait leur apparition permettant de définir une classification morphologique de base. La première réelle étude de la communauté planctonique fut réalisée en 1828 par le Dr J. Vaughan Thompson (Fraser 1968). Victor Hensen, un médecin, a ensuite développé le filet Hensen (encore populaire aujourd'hui) à partir des nombreuses réflexions et études menées sur la quantification et l'analyse du plancton. À la fin du 19e siècle, Teodor Cleve a commencé à utiliser des filets en soie pour étudier la distribution du phytoplancton dans la mer du Nord (Cleve 1899). Rapidement, l'intérêt croissant pour les stocks de poissons marins a encouragé de nombreux laboratoires scientifiques à entamer un examen détaillé des communautés planctoniques. Hans Lohmann a ainsi développé des techniques de centrifugation pour concentrer le plancton, permettant la découverte du nano-plancton (Lohmann 1911). Les premières tentatives pour déterminer des modèles mondiaux de la distribution des espèces planctoniques ont eu lieu au milieu des années 1950, avec l'apparition de la méta-analyse (Smayda 1958). Le développement des techniques de fluorescence pour l'identification du phytoplancton, qui débute en 1977, va de nouveau révolutionner le domaine, permettant une analyse beaucoup plus précise de cette composante majeure de la communauté planctonique. Il faudra attendre 1979 pour voir apparaître les premières silhouettes photographiques de micro-organismes planctoniques. Ces images ne sont, à cette époque, utilisées que pour limiter les problèmes liés à la conservation des échantillons (Ortner *et al.* 1979). Malgré leurs débuts comme outils d'archives, les caméras (dont les définitions d'acquisition ne cessent de s'améliorer) s'avèrent rapidement être de précieux outils pour l'analyse de la communauté planctonique *in situ*. Des caméras furent ainsi progressivement installées sur des submersibles (Wiebe & Benfield 2001).

Si les méthodes de collectes se sont perfectionnées, malheureusement, l'analyse (traditionnelle) des échantillons de plancton n'a elle que peu évolué au cours du temps et est encore aujourd'hui synonyme de longues séances de comptage des organismes planctoniques sous un microscope ou devant un écran d'ordinateur par un taxonomiste. Cette méthodologie entièrement manuelle impose un temps de traitement relativement long des échantillons et constitue une limitation, voire un goulot d'étranglement pour l'expansion de la taille et du nombre des échantillons analysés. Le nombre de taxonomistes capables de réaliser ce genre de travail est d'ailleurs de plus en plus réduit. Il existe cependant aujourd'hui une alternative pour observer et recueillir un plus grand nombre de données sur le plancton : l'analyse d'images assistée par ordinateur. Cette méthode, complémentaire aux taxonomistes, est considérée depuis les années 1980 (Ortner *et al.* 1979) mais a été pendant longtemps limitée par la qualité et la résolution des dispositifs d'acquisition.

Aujourd'hui, la combinaison de puissants ordinateurs avec des périphériques de numérisation de très haute qualité ainsi que l'efficacité croissante des algorithmes mathématiques pour faire de « l'apprentissage machine » fournit un outil d'analyse puissant pour les biologistes. L'analyse assistée par ordinateur du plancton peut fournir un recensement et une identification rapide d'un très grand nombre d'échantillons de plancton en un temps très limité. Ces nouvelles techniques, dites à haut-débit, permettent une reproductibilité plus élevée des identifications et des résultats (Embleton *et al.* 2003). Elles permettent également d'estimer le taux d'erreur de manière mathématique, car les erreurs d'un algorithme sont reproductibles, là où il a été montré qu'il était beaucoup plus complexe d'estimer le taux d'erreur d'un taxonomiste (Culverhouse *et al.* 2003; Benfield *et al.* 2007).

Pour aider les planctonologistes dans leurs travaux toute une série de matériels capables de prendre des photos du plancton, à la fois in situ et/ou à partir d'échantillons fixés, a ainsi été mise au point : le FlowCAM (Sieracki *et al.* 1998) (<http://www.fluidimaging.com>), le Laser Optical Plankton Counter (Herman *et al.* 2004), le Video Plankton Recorder (Davis *et al.* 1992), le Zooscan (Grosjean *et al.* 2004), l'Holocam (Craig *et al.* 2000), le CytoBuoy (Dubelaar & Gerritzen 2000), le Sipper (Samson *et al.* 2001), le HAB Buoy (Culverhouse *et al.* 2006a), l'ISIIS (Cowen & Guigand 2008), le Zoovis (Wiebe *et al.* 2010), etc., sans oublier l'utilisation d'un appareil photo numérique sur un microscope ou avec un objectif macro. Évidemment les images numériques du plancton ainsi obtenues ne sont que très peu utilisables en tant que telles. Elles doivent être analysées de manière à ce que les informations extraites de leurs pixels puissent produire de l'information biologique et écologique. Toutes les images digitalisées par ce type de matériel ont bien entendu toujours besoin de l'intervention et de la supervision de taxonomistes, notamment pour la mise en place de l'identification et de la classification, mais à un degré moindre.

La méthode reposant sur de « l'apprentissage machine » utilise les caractéristiques extraites depuis un lot représentatif de particules (issues de l'échantillon à analyser) pour identifier

automatiquement et classer l'ensemble des particules de l'échantillon en différents groupes morphologiques, taxonomiques ou écologiques. L'utilisation d'un logiciel capable de faire une telle analyse est donc une phase essentielle dans le traitement des images numériques. Différentes tentatives ont été déployées pour traiter ces images et en extraire de l'information. Le DiCANN (Dinoflagellate Categorization by Artificial Neural Network) (Culverhouse *et al.* 2003) est un logiciel développé pour analyser et classer les dinoflagellés avec des algorithmes de type réseau neuronal. Cette méthode peut distinguer six espèces avec 72% de précision. Le REFLICS (REal FLOW Imaging and Classification System) (Iwamoto *et al.* 2001) fournit un logiciel de reconnaissance automatique en temps réel pour les œufs de poisson, avec quelques faux négatifs, en utilisant l'algorithme du plus proche voisin. L'appareil FlowCAM est muni de Visual Spreadsheet (Brown 2008), un logiciel qui extrait des caractéristiques des images générées par l'appareil. Il est capable de créer des regroupements en fonction d'un spectre de taille défini. Il peut également reconnaître les particules en fonction de leurs similitudes. Le PVA (Plankton Visual Analyzer) est un logiciel gratuit qui compte et classe les particules par taille à partir d'images numériques (<http://www.azti.es/>), mais l'ensemble du processus prend beaucoup de temps. Plus récemment, le logiciel Visual Plankton a été développé pour analyser les images du Video Plankton Recorder. Enfin, il existe un logiciel gratuit conçu pour traiter des images issues du ZooScan : ZooProcess (<http://www.obs-vlfr.fr/LOV/ZooPart/ZooScan/>). Tous ces logiciels sont spécifiques à un ou quelques formats d'images et souvent spécifiques à un dispositif de numérisation.

Zoo/PhytoImage (<http://www.sciviews.org/Zoo/PhytoImage>) est un autre logiciel proposé en tant que solution libre pour l'analyse d'images du plancton dont les origines peuvent-être diverses. Il offre la possibilité de réaliser de nombreuses mesures en complément de celles réalisées par le programme d'acquisition de l'appareil (comme l'abondance, le spectre de taille partiel ou total, la biomasse totale ou partielle, etc.). Ce logiciel fournit un cadre général pour importer et analyser des images, puis exporter les résultats de l'analyse vers d'autres systèmes. Il est construit à partir de différents logiciels existants comme ImageJ (Abràmoff *et al.* 2004), R ou XnView.

### *1.2. Les contraintes de l'analyse d'image à haut débit*

Avant de commencer l'acquisition d'images depuis un échantillon de plancton, et quelque soit le type de dispositif d'acquisition que l'on sélectionne, il est nécessaire de connaître et de définir exactement le but de l'analyse d'images. Cela signifie qu'il faut déterminer le seuil de reconnaissance que l'on souhaite atteindre pour les particules que l'on va imager (l'espèce, le genre, l'ordre, ou encore un groupe morphologique). Même s'il est préférable de numériser les particules à la résolution maximale, il est important de se souvenir que la résolution taxonomique dépend de la taille des pixels, et donc qu'une classification simple (l'ordre par exemple) ne nécessite pas une acquisition à très haute

définition. Les fichiers générés sont ainsi de plus petites tailles et permettent un traitement plus rapide pour la suite du processus d'analyse.

Après avoir été prélevé, le plancton peut être analysé de deux manières : soit directement par imagerie en temps réel (en « live »), soit ultérieurement après une fixation chimique. Le temps qui s'écoule entre l'instant du prélèvement et le moment où l'analyse est lancée dicte souvent si l'imagerie en « live » est possible ou non. En effet, bien qu'il puisse être plus instructif d'étudier les organismes lorsqu'ils sont vivants, le temps nécessaire pour cette analyse fait de la fixation un atout, permettant une observation sur une plus longue période et potentiellement plus en détail. Cependant, des difficultés apparaissent rapidement lors de l'utilisation d'échantillons fixés. En effet les cellules peuvent rétrécir, se déformer et certaines structures se dégrader plus ou moins rapidement au cours du temps, ayant alors une incidence directe sur les analyses taxonomiques (Zarauz & Irigoien 2008).

Divers dispositifs existent pour capturer des images depuis un échantillon de plancton frais et/ou fixé comme nous l'avons vu précédemment, mais je n'en décrirai ici qu'un, le FlowCAM, car c'est ce dispositif d'imagerie à haut débit qui a été choisi pour l'expérience Tara Océans. Le FlowCAM est un procédé d'imagerie à haute résolution. Les images sont acquises avec un matériel basé sur le principe de la cytométrie en flux. Les particules sont numérisées dans un flux d'eau et les images stockées sur un ordinateur. L'enregistrement sur l'ordinateur des images empêche la perte de données et permet une post-analyse. Un traitement plus approfondi des images peut ainsi être réalisé sur un ordinateur distant à l'aide d'un logiciel conçu spécialement pour faire des analyses géométriques plus complexes.

### *1.3. Objectifs*

Dans le cadre de mon travail de thèse j'ai voulu explorer la possibilité d'utiliser le FlowCAM associé au logiciel Zoo/PhytoImage pour étudier la diversité et la distribution spatiale des dinoflagellés présents dans les échantillons de la mer Méditerranée prélevés pendant l'expédition Tara Océans.

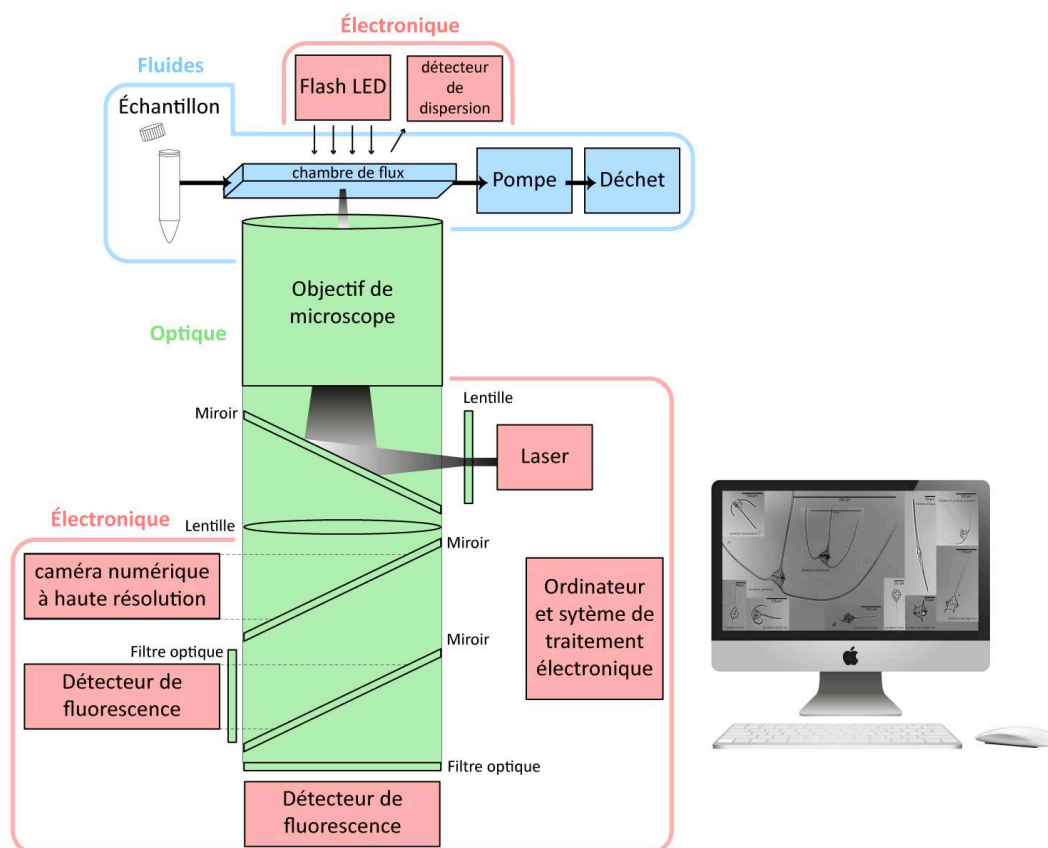
Je me suis exclusivement intéressé aux données issues du FlowCAM installé sur le bateau Tara et je traite ici exclusivement des images faites en « live » sur Tara.

Ce chapitre vise à présenter la méthodologie développée en partenariat avec l'université de Mons en Belgique (avec le Dr Philippe Grosjean et son étudiant Kevin Denis) et l'université de Dublin en Irlande pour l'acquisition et l'analyse des données FlowCAM (Dr Emmanuel Reynaud et son étudiante Maureen Williams).

## 2. Matériel et méthode

### 2.1. Présentation et caractéristiques du dispositif FlowCAM

Le FlowCAM (Flow Cytometer And Microscope) combine deux technologies : la cytométrie en flux et la microscopie photonique reliée à une caméra numérique (Fig.1).



**Figure 1 :** Principe de fonctionnement du FlowCAM. Trois unités sont à distinguer : le système optique (vert), les dispositifs électroniques (rouge) et l'espace de circulation des fluides (bleu). Le tout est couplé à un ordinateur servant au pilotage de l'appareil et au stockage des images.

L'échantillon d'eau de mer à analyser est transporté par une pompe péristaltique vers une cellule passante en verre de dimensions parfaitement connues (appelée chambre de flux), positionnée devant un objectif de microscope lui-même relié à une caméra vidéo. La caméra réalise une acquisition d'images des particules selon des paramètres préalablement définis : soit suite à la détection par un faisceau laser, soit à intervalle de temps régulier. Le laser permet de détecter la particule elle-même, ou bien sa fluorescence (deux canaux sont disponibles : un pour la chlorophylle, un pour la phycoérythrine). À chaque particule détectée, selon les paramètres initialement spécifiés, correspond une photo à laquelle est associée un ensemble de 21 mesures (surface, diamètre, etc.) stockées dans un fichier texte tabulé.



Le FlowCAM utilisé en routine à bord du bateau Tara est équipé d'une caméra noir et blanc haute résolution (FireWire Progressive Scan) et d'un objectif de grossissement 4X, qui s'utilise avec des cellules en verre calibrées de profondeur 300µm ou 600µm. Cet assemblage garantit une profondeur de champ idéale pour une netteté optimale des particules circulant dans le flux. (Un appareillage identique est disponible au laboratoire pour toute analyse complémentaire). Visual SpreadSheet (v2.2.7), le logiciel de paramétrage et d'exploitation des données acquises à l'aide du FlowCAM, permet de définir les paramètres d'acquisitions et de mesures des images. Les principaux paramètres de réglages du logiciel Visual SpreadSheet sont définis avant l'acquisition.

Ils concernent essentiellement :

- L'illumination de la partie de la cellule calibrée (intensité, homogénéité lumineuse),
- Les paramètres vidéo d'acquisition des images (taille, nombre d'images par seconde, sauvegarde des fichiers de sortie...),
- Les paramètres de détection des particules (filtres de taille, critères de teinte...).

### *2.2. Présentation et caractéristiques du logiciel Zoo/PhytoImage*

Le logiciel Zoo/PhytoImage est un programme d'analyse et de classification automatique d'images. Basé sur le principe du « machine learning », il permet de réaliser les différentes étapes du processus qui conduit à la classification automatisée, ou semi-automatisée, d'un ensemble d'objets géométriques, à partir d'un jeu d'images donné (Fig. 2).

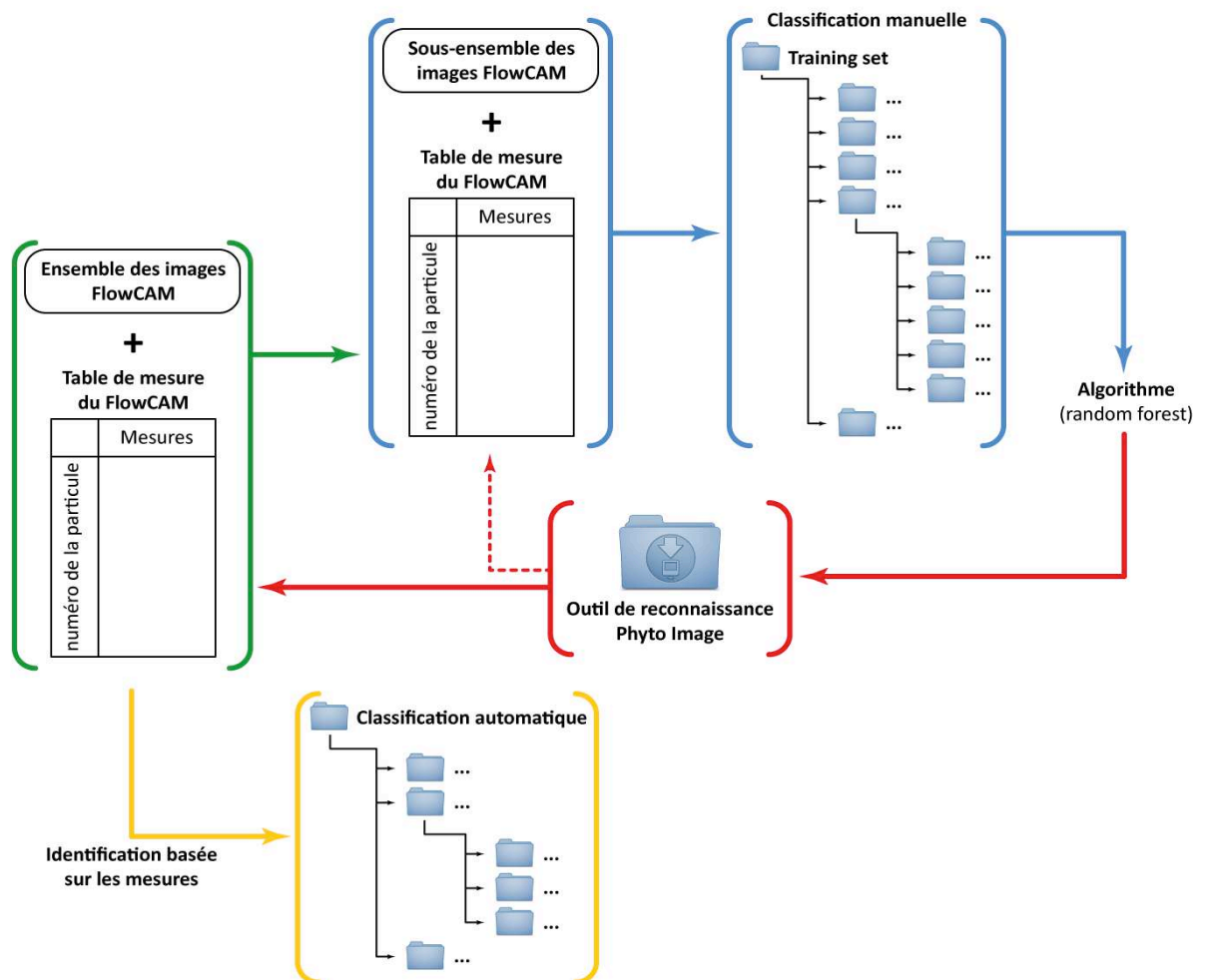
Une étape préliminaire indispensable à celle de l'apprentissage est la reconnaissance des particules étudiées. Cet apprentissage (à travers une interface) se fait en deux phases :

- ↪ La première phase consiste à réaliser un set d'apprentissage constitué à partir d'une banque d'images issues de l'appareil de numérisation (ici le FlowCAM) et représentatives des particules rencontrées dans les échantillons à analyser ultérieurement. Ces images sont identifiées visuellement par un observateur (taxonomiste) puis classées dans autant de dossiers ou sous-dossiers que nécessaire pour représenter au mieux les niveaux d'identifications recherchés.
- ↪ La seconde phase pour la réalisation de l'outil de reconnaissance automatique, ou semi-automatique, utilise le set d'apprentissage précédemment créé pour déterminer un algorithme de classification (supervisée) capable de reconnaître la nature des particules sur la base des mesures obtenues sur les images. Pour que cet apprentissage soit optimal, plusieurs étapes sont nécessaires :
  - ✓ choisir le meilleur algorithme possible,
  - ✓ optimiser les (parfois nombreux) paramètres de cet algorithme sur la base de validations croisées ayant pour objectif de quantifier et de maximiser le taux de reconnaissance global,

voire le taux de reconnaissance spécifique d'un ou plusieurs groupes taxonomiques (selon l'objectif visé),

- ✓ étudier l'effet du regroupement progressif des taxons en vue d'obtenir un meilleur compromis entre le niveau de détail et les performances de classification,
- ✓ enfin, évaluer les performances obtenues, en particulier le taux d'erreur entre les différents taxons, la robustesse par rapport à des échantillons nouveaux, le nombre de particules identifiées par unité de temps, etc.

Une fois l'outil de reconnaissance automatique optimisé, celui-ci peut être utilisé sur un nouveau jeu d'images pour classer automatiquement les particules dans les catégories qui ont été définies manuellement dans le set d'apprentissage initial.



**Figure 2 :** Principe de la classification supervisée. Un ensemble d'images, acquis grâce au FlowCAM (vert), est échantillonné pour former un sous-ensemble représentatif d'images (bleu) qui est alors trié manuellement. Ce tri est soumis à un algorithme mathématique pour créer un outil de reconnaissance (rouge). Cet outil une fois validé peut être utilisé sur un ensemble d'images FlowCAM (vert) pour générer une identification automatique (jaune).

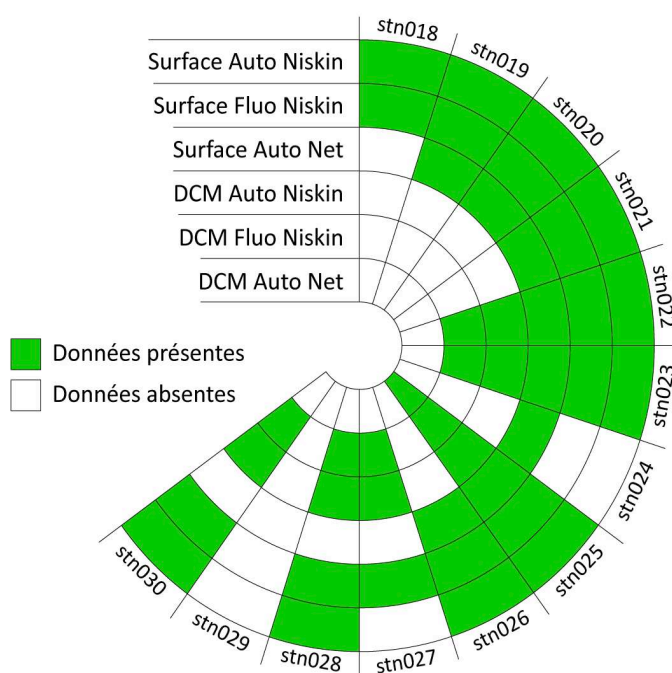
### *2.3. Méthodologie à bord de Tara*

Pour ce travail entamé en décembre 2010, avant la fin de la mission Tara Océans, seuls les échantillons de mer Méditerranée (13 stations) étaient à disposition et seront donc analysés et présentés. Les méthodes d'échantillonnage déployées (cf. Chapitre 1) ont été conçues pour collecter un large éventail de taille d'organismes. Pour chaque station, deux méthodes d'échantillonnage, le filet à plancton et la rosette équipée de bouteilles Niskin, ont été utilisées dans le but de faire de l'imagerie. Les bouteilles Niskin et les filets ont été généralement déployés à deux profondeurs, en surface et à la profondeur du maximum en chlorophylle (DCM), zone où se situe habituellement la plus forte concentration en plancton.

Les échantillons destinés à l'analyse FlowCAM et à l'imagerie en « live » ont été réalisés à bord par le personnel scientifique embarqué très peu de temps après leurs collectes, évitant ainsi une acquisition après fixation. Les échantillons ont été analysés selon deux modes d'acquisition distinctes, le mode automatique et le mode de détection par fluorescence (Álvarez & López-Urrutia 2011). Les images ont été acquises à un grossissement de 4x et les échantillons traités ont été détruits à la fin de l'analyse (un duplicata d'échantillon ayant été généré au cours de l'échantillonnage cf. Chapitre 1). Les données sont acquises avec le logiciel Visual Spreadsheet, qui fournit des informations en temps réel sur l'échantillon. Il renseigne l'utilisateur sur l'emplacement de la capture (coordonnées GPS), la distribution de taille au sein de l'échantillon, et affiche une synthèse des mesures géométriques. Les images générées sont également affichées en temps réel ainsi que les vignettes obtenues par segmentation des particules. Les sorties-écrans sont combinées en temps réel avec les images brutes et enregistrées sur l'ordinateur relié au FlowCAM. Ces données sont ensuite transférées dans un format compressé sur des disques durs externes pour une analyse ultérieure en laboratoire.

### *2.4. Traitement des données en laboratoire*

L'analyse en laboratoire des données FlowCAM commence par un inventaire des données présentes. L'utilisation de plusieurs disques durs et de plusieurs opérateurs à bord a en effet conduit à un ensemble complexe et multiple de conventions de nommage des données. Chaque station présente ensuite une combinaison unique de profondeurs échantillonnées, de méthodes de déclenchement et de méthodes d'échantillonnage. Une représentation claire des données disponibles sous la forme de diagrammes circulaires a ainsi été générée par Maureen Williams fin 2012 (Williams 2012) (Fig. 3).



**Figure 3** : Inventaire des données FlowCAM disponibles pour chaque station échantillonnée en Mer Méditerranée. (Adapté de Maureen Williams, 2012).

L'analyse géométrique commence par l'importation des données issues du FlowCAM (acquises avec le logiciel Visual Spreadsheet) dans le logiciel Zoo/PhytoImage. Celui-ci traite alors chaque vignette générée par le FlowCAM comme l'image d'un organisme planctonique. Des fonctions R, créées dans TinnR (Faria 2013), permettent de lancer la procédure Zoo/PhytoImage qui prépare les images avant leur exportation vers ImageJ (Abràmoff *et al.* 2004). ImageJ est ensuite utilisé avec le plug-in FIT-VIS pour améliorer les images en créant un fichier binaire, un masque, et une image en niveaux de gris avec un seuil automatique. Le plug-in FIT-VIS d'ImageJ calcule ensuite 26 paramètres pour chaque vignette (ou particule) (tableau 1). Ces mesures sont enregistrées avec les vignettes, et les masques. Toutes les données relatives aux particules de chaque échantillon, y compris les paramètres environnementaux et les informations issues de Visual Spreadsheet, sont combinées en un fichier unique, un fichier ZID. Ce fichier compressé contient les données mesurées et les images. Un fichier ZID est généré pour chaque échantillon dans chaque station.

Pour constituer l'outil de reconnaissance semi-automatique, le choix a été fait de trier manuellement dans l'interface d'XnView l'ensemble des particules contenues dans les fichiers ZID de 21 échantillons de la Mer Méditerranée (Fig. 4). Le choix des échantillons est basé sur la qualité de l'acquisition (volume échantillonné, qualité des images) et le volume de données à analyser. Au total, ce sont 66 691 particules qui ont ainsi été triées manuellement.

2.5. L'analyse du genre *Ceratium*

Les échantillons ont été collectés par filet en surface entre le 16 novembre et le 15 décembre 2009, et immédiatement numérisés en mode automatique sans fixation chimique avec le FlowCAM installé à bord du navire Tara. Le volume imagé est équivalent pour l'ensemble des échantillons. L'assignation taxonomique à l'espèce a été réalisée par un taxonomiste (Dr Alina Tunin Ley) après un traitement de classification semi-automatique des échantillons.

| Parameter   | Description   |
|-------------|---|
| ECD         | Equivalent circular diameter                                    |
| Area        | Surface Area  |
| Mean        | Mean of the gray scale of the pixels                            |
| Skew        | The third-order moment, about the mean of the gray scale        |
| Kurt        | The fourth-order moment, about the mean of the gray scale       |
| StDev       | Standard deviation of the gray scale of the pixels              |
| Mode        | Mode of the gray scale of the pixels                            |
| Median      | Median of the gray scale of the pixels                          |
| Min         | Minimum of the gray scale of the pixels                         |
| Max         | Maximum of the gray scale of the pixels                         |
| IntDen      | Sum of the gray scale of the pixels                             |
| XM          | Coordinate horizontal of the gray scale centre of the pixels    |
| YM          | Coordinate vertical of the gray scale centre of the pixels      |
| Perim       | Perimeter   |
| Width       | Width of the rectangle containing the particle                  |
| Height      | Height of the rectangle containing the particle                 |
| Major       | Longest axis of the ellipsis containing the particle            |
| Minor       | Shortest axis of the ellipsis containing the particle           |
| Circ        | Circularity   |
| Feret       | Diameter of longest distance between two points of the particle |
| Temperature | Surface temperature during collection                           |
| Salinity    | Salinity of the sample  |
| Depth       | Depth of sample collection                                      |
| Latitude    | Latitude of sample collection                                   |
| Longitude   | Longitude of sample collection                                  |

Tableau 1 : Liste des 26 paramètres analysés par ImageJ avec le logiciel Zoo/PhytoImage. (depuis (Fernandes *et al.* 2008)).

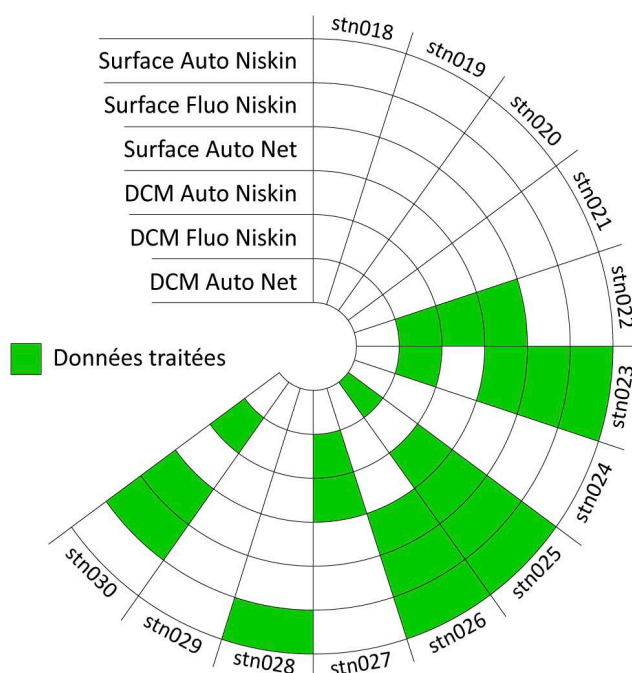


Figure 4 : Liste des échantillons traités manuellement pour la constitution de l'outil de reconnaissance.

### 3. Résultats et discussion

#### 3.1. Limitation technique de l'analyse

Sur 66 691 particules triées, seules 9 238 (14%) ont pu être correctement identifiées (57 453 — 86% — n'ont pas pu être classées). En effet, même si la classification semi-automatique d'images planctoniques fournit un grand nombre d'images classées et de précieuses informations, elle produit aussi des pourcentages très élevés d'images « non identifiables ». Il est, avec ce procédé d'analyse des images, parfois très difficile d'identifier et d'isoler de manière fiable la neige marine, les détritiques, les bulles, ou les particules non planctoniques (Benfield *et al.* 2007). La présence d'agrégats (des regroupements de plusieurs individus), parfois associés à des détritiques ou à de la matière organique, peut aussi compliquer l'analyse. Certains micro-organismes peuvent être visibles dans ces agrégats et peuvent être identifiés lors d'un comptage manuel par un taxonomiste, mais sont exclus lors d'un comptage semi-automatisé. En effet, la géométrie de l'objet détecté est dans ce cas très éloignée de la géométrie de l'organisme identifié. Un nombre élevé d'agrégats peut ainsi fausser de manière significative les proportions relatives des taxons observés. Certains taxons forment plus fréquemment des agrégats que d'autres.

La plupart des outils utilisés pour l'analyse d'images semi-automatisée reposent sur une technologie en deux dimensions (2D) de capture puis d'analyse des images. Pourtant, à cause du système de pompe péristaltique, le mouvement de l'échantillon de plancton est en trois dimensions (3D) au moment de l'acquisition (mise en mouvement d'un fluide). Dans ces conditions, le plancton peut être imagé sous n'importe quel angle de vue, ce qui multiplie de manière quasi infinie les géométries associables à un taxon et rend très difficile la reconnaissance par un appareil sur ces seuls critères (Culverhouse *et al.* 2006b). Les mouvements tridimensionnels des particules dans le fluide lors de la capture des images provoquent également des images floues ce qui limite une nouvelle fois la capacité d'identification (Culverhouse *et al.* 2006b).

D'autres problèmes sont également rencontrés pour l'identification semi-automatique lorsque l'on utilise la longueur, la largeur, la taille (ou forme) grande ou petite des particules. Une même espèce peut avoir des individus dont la taille diffère au cours de son cycle de vie (Pizay *et al.* 2009). Dans ce cas, le programme risque de mal classer ou de considérer ces objets comme des taxons distincts (Bell & Hopcroft 2008). Certaines espèces planctoniques peuvent également former des colonies aux formes et aux dimensions variables qu'il est très difficile de caractériser et qui seront identifiées comme des objets uniques.

Cette technologie est donc, en l'état, très peu adaptée pour réaliser une observation détaillée d'une communauté dans son ensemble. Ce constat est d'autant plus clair pour la communauté des dinoflagellés qui présente des morphologies caractéristiques, mais parfois radicalement différentes au

cours du temps. Ceci étant, malgré ces nombreux problèmes de classification, l'analyse de certains membres de la communauté planctonique est plutôt couronnée de succès.

### 3.2. Analyse des dinoflagellés identifiables par FlowCAM/ZooPhytoImage

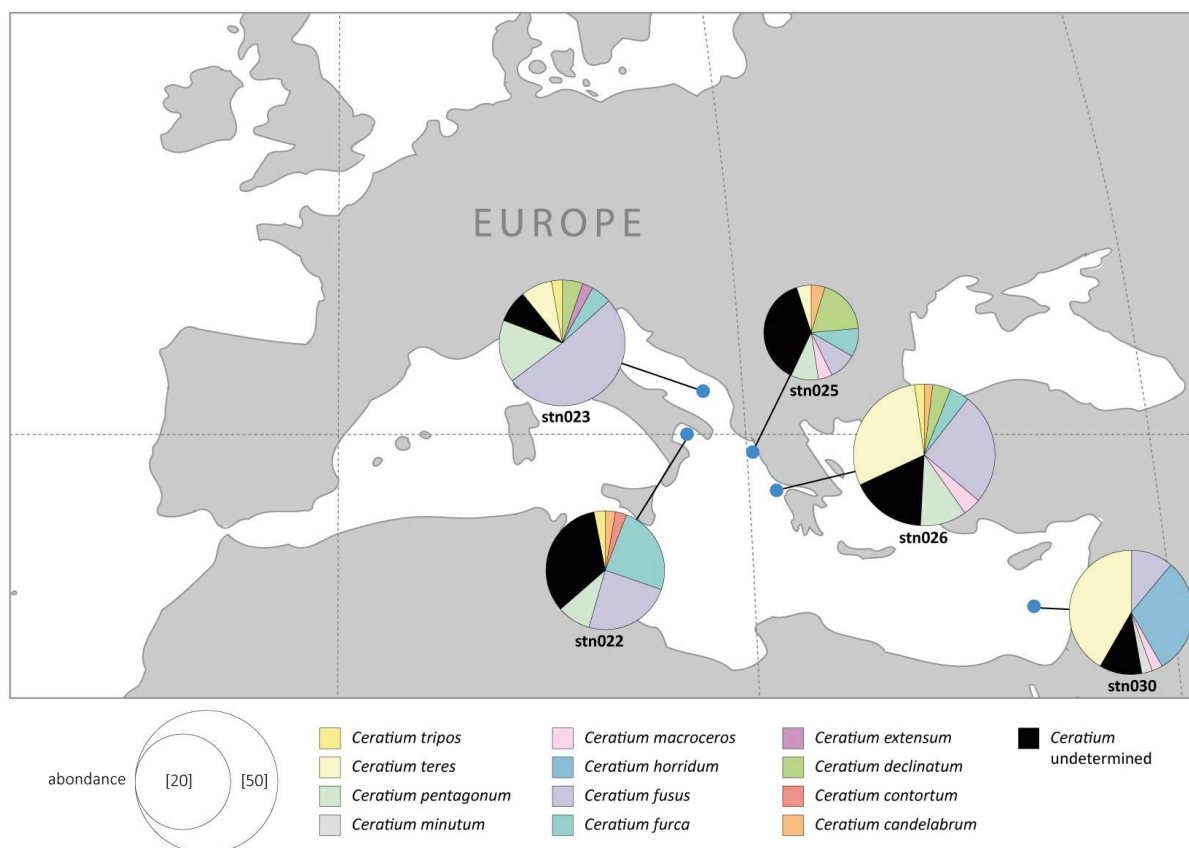
Peu de dinoflagellés sont identifiables par l'analyse FlowCAM/ZooPhytoImage. En effet, la plupart des dinoflagellés sont de forme sphéroïdale et présentent assez peu de caractéristiques spécifiques au niveau de l'espèce ou du genre. Certains dinoflagellés comme *Karenia* (Buskey & Hyatt 2006) ou *Ceratium* ont cependant des morphologies suffisamment caractéristiques pour qu'il soit possible de les identifier précisément au niveau de l'espèce à partir d'images produites par le FlowCAM. Pour le genre *Ceratium*, il a été possible de créer un outil de reconnaissance efficace capable de discriminer plusieurs espèces à la morphologie très caractéristique. Les échantillons de surface de la mer Méditerranée ont ainsi révélé une diversité non négligeable du genre *Ceratium* (Fig. 5).

174 images de *Ceratium* spp. ont pu être isolées. Douze espèces ont été identifiées avec précision, et un groupe de cellules regroupant des cellules de *Ceratium* indéterminables au niveau spécifique ajouté. *Ceratium teres* et *N. fusus* sont les espèces les plus abondantes représentant à elles deux 45% des cellules observées. *Ceratium teres* domine la partie Est du bassin méditerranéen alors que *N. fusus* domine la partie Ouest.

La station 30 présente une diversité moins importante que les autres stations observées dans cette étude avec une forte abondance de *N. horridum* une espèce que l'on n'observe pas ailleurs. Cette station est située dans le tourbillon anticyclonique Shikmona (qui se forme au-dessus du mont sous-marin Ératosthène) ce qui lui confère peut-être des caractéristiques physico-chimiques clairement différentes des autres masses d'eau voisine et favorise cette espèce.

Ces résultats restent malheureusement insuffisants pour réaliser une étude robuste de la diversité du genre *Ceratium* en Mer Méditerranée puisque le nombre de points échantillonnés et le volume d'eau analysé sont insuffisants pour obtenir une couverture exhaustive de la diversité. Une analyse plus approfondie de la station 30, un volume 10 fois supérieure au volume protocolaire y a été analysé, a en effet montré que nous ne saturons pas en terme de diversité. Avec un seuil à 100 000 particules détectées contre 10 000 habituellement le nombre d'espèces de *Ceratium* identifié passe de 17 à 25 espèces.

Ce travail, qui n'a pas été poursuivi, a cependant montré des résultats encourageants.



**Figure 5** : Répartition de la diversité et de l'abondance des *Ceratium* en mer Méditerranée. L'identification a été réalisée par un taxonomiste à partir d'images issues du FlowCAM.

#### 4. Conclusion

La méthode semi-automatique d'identification du plancton basée sur des critères géométriques ne permet pas de faire une couverture suffisante de la diversité des dinoflagellés et n'est donc pas adaptée à une étude globale des communautés de dinoflagellés. Une approche en trois dimensions des particules devrait être envisagée pour une meilleure identification. Le développement de l'outil de reconnaissance qui est basé sur un échantillon représentatif du set de données finales observées est encore trop complexe à réaliser. En effet, l'importante variabilité au sein des observations, qui peut être due à la dispersion géographique, à la saisonnalité ou aux variations morphologiques intra spécifique augmente considérablement les variations des critères de discrimination. La fiabilité de l'outil de reconnaissance est dans ce cas réduite, augmentant les erreurs de classification. Le temps nécessaire pour produire une identification de qualité des particules reste trop important (égale voir supérieur au comptage traditionnel par un taxonomiste).

Cet outil présente tout de même des avantages. Il permet de faire un prétraitement relativement rapide des organismes présents dans un échantillon et la détection d'espèces cibles. Cette approche peut s'avérer particulièrement intéressante pour l'exploitation d'un groupe cible,



comme le genre *Ceratium*, dans les échantillons de la mission Tara Océans, permettant de cibler les échantillons d'intérêts pour un autre type d'analyses plus spécifiques.

Le FlowCAM peut également être envisagé comme un outil capable de faciliter la détection de certaines espèces planctoniques aux caractéristiques morphologiques marquantes sur de longues périodes de temps et à de grandes échelles géographiques. Le suivi de certaines espèces toxiques (comme *Dinophysis* spp.) en environnement côtier pourrait ainsi être optimisé.

---

### Remerciements :

Je souhaite remercier mes collaborateurs qui tout au long de ce travail m'ont conseillé et formé. Je pense tout d'abord au Dr Philippe Grosjean et à son étudiant Kevin Denis de l'université de Mons qui m'ont chaleureusement accueilli dans leur laboratoire en décembre 2010, mais également à Alina Tunin Ley qui m'a reçu à la station marine d'Arcachon (début 2011). Je remercie également le Dr Emmanuel Reynaud de l'université de Dublin et son étudiante Maureen Williams qui ont repris une partie de mes analyses et poursuivi le traitement de ces données en 2012. Merci également à Sébastien Colin pour sa contribution non négligeable pour la constitution de l'outil de reconnaissance, et à Jean-Baptiste Romagnan pour nos échanges d'expériences en set d'apprentissage FlowCAM.

## Références

- Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. *Biophotonics*.
- Álvarez E, López-Urrutia Á (2011) How to effectively sample the plankton size spectrum? A case study using FlowCAM. *Journal of Plankton Research*, **33**, 1119–1133.
- Bell JL, Hopcroft RR (2008) Assessment of Zoolmage as a tool for the classification of zooplankton. *Journal of Plankton Research*, **30**, 1351–1367.
- Benfield MC, Grosjean P, Culverhouse PF *et al.* (2007) RAPID: research on automated plankton identification. *Oceanography*, **20**, 172–187.
- Brown L (2008) Particle Image Understanding—A Primer. *Fluid Imaging Technologies*.
- Buskey EJ, Hyatt CJ (2006) Use of the FlowCAM for semi-automated recognition and enumeration of red tide cells (*Karenia brevis*) in natural plankton samples. *Harmful Algae*, **5**, 685–692.
- Cleve PT (1899) *Plankton: collected by the Swedish expedition to Spitzbergen in 1898*. P.A. Norstedt & Söner, Stockholm.
- Cowen RK, Guigand CM (2008) In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. *Limnol Oceanogr Methods*, **6**, 126–132.
- Craig G, Alexander SJ, Anderson S *et al.* (2000) HoloCam: a subsea holographic camera for recording marine organisms and particles. *Proceedings of SPIE*, **4076**, 111–119.
- Culverhouse PF, Williams R, Simpson B (2006a) HAB Buoy: a new instrument for in situ monitoring and early warning of harmful algal bloom events. *African Journal of Marine Science*, **28**, 245–250.
- Culverhouse PF, Williams R, Benfield M *et al.* (2006b) Automatic image analysis of plankton: future perspectives. *Marine Ecology Progress Series*, **312**, 297–309.
- Culverhouse PF, Williams R, Reguera B, Herry V, Gonzales-Gil S (2003) Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, **247**, 17–25.
- Davis CS, Gallager SM, Berman MS, Haury LR, Strickler JR (1992) The Video Plankton Recorder (VPR): design and initial results. *Archiv für Hydrobiologie Beiheft Ergebnisse der Limnologie*, **36**, 67–81.
- Dubelaar G, Gerritzen PL (2000) CytoBuoy: a step forward towards using flow cytometry in operational oceanography. *Scientia Marina*, **64**, 255–265.
- Embleton KV, Gibson CE, Heaney SI (2003) Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *Journal of Plankton Research*, **25**, 669–681.
- Faria JC (2013) Resources of Tinn-R GUI/Editor for R Environment. *Ilheus*.
- Fernandes JA, Irigoien X, Boyra G, Lozano JA, Inza I (2008) Optimizing the number of classes in automated zooplankton classification. *Journal of Plankton Research*, **31**, 19–29.
- Fraser JH (1968) *The history of plankton sampling*. UNSECO.
- Grosjean P, Picheral M, Warembourg C, Gorsky G (2004) Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES Journal of Marine Science: Journal du Conseil*, **61**, 518–525.
- Herman AW, Beanlands B, Phillips EF (2004) The next generation of optical plankton counter: the Laser-OPC. *Journal of Plankton Research*, **26**, 1135–1145.
- Iwamoto S, Checkley DM Jr, Trivedi MM (2001) REFLICS: Real-time flow imaging and classification system. *Machine Vision and Applications*, **13**, 1–13.

- Lohmann H (1911) Über das nannoplankton und die Zentrifugierung kleinster Wasserproben zur Gewinnung desselben in lebendem Zustande. *Internationale Revue gesamten Hydrobiologie*, **4**, 1–38.
- Ortner PB, Cummings SR, Aftring RP, Edgerton HE (1979) Silhouette photography of oceanic zooplankton. *Nature*, **277**, 50–51.
- Pizay M-D, Lemée R, Simon N *et al.* (2009) Night and Day Morphologies in a Planktonic Dinoflagellate. *Protist*, **160**, 565–575.
- Samson S, Hopkins T, Remsen A *et al.* (2001) A system for high-resolution zooplankton imaging. *IEEE Journal Of Oceanic Engineering*, **26**, 671–676.
- Sieracki CK, Sieracki ME, Yentsch CS (1998) An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*, **168**, 285–296.
- Smayda TJ (1958) Biogeographical Studies of Marine Phytoplankton. *Oikos*, **9**, 158.
- Wiebe PH, Benfield MC (2001) Zooplankton Sampling With Nets And Trawls. In: *Encyclopedia of Ocean Sciences*, pp. 3237–3253. Elsevier.
- Wiebe PH, Ashjian C, Sutor M, Lavery A, Fratantoni D (2010) Spatial and Temporal Variability of Zooplankton Thin Layers: The Effects of Composition and Orientation on Acoustic Detection of Layers.
- Williams M (2012) *Organization and Preliminary Analysis of Global Planktonic Community Flowcam Images From the Tara Oceans Expedition*.
- Zarauz L, Irigoien X (2008) Effects of Lugol's fixation on the size structure of natural nano-microplankton samples, analyzed by means of an automatic counting method. *Journal of Plankton Research*, **30**, 1297–1303.

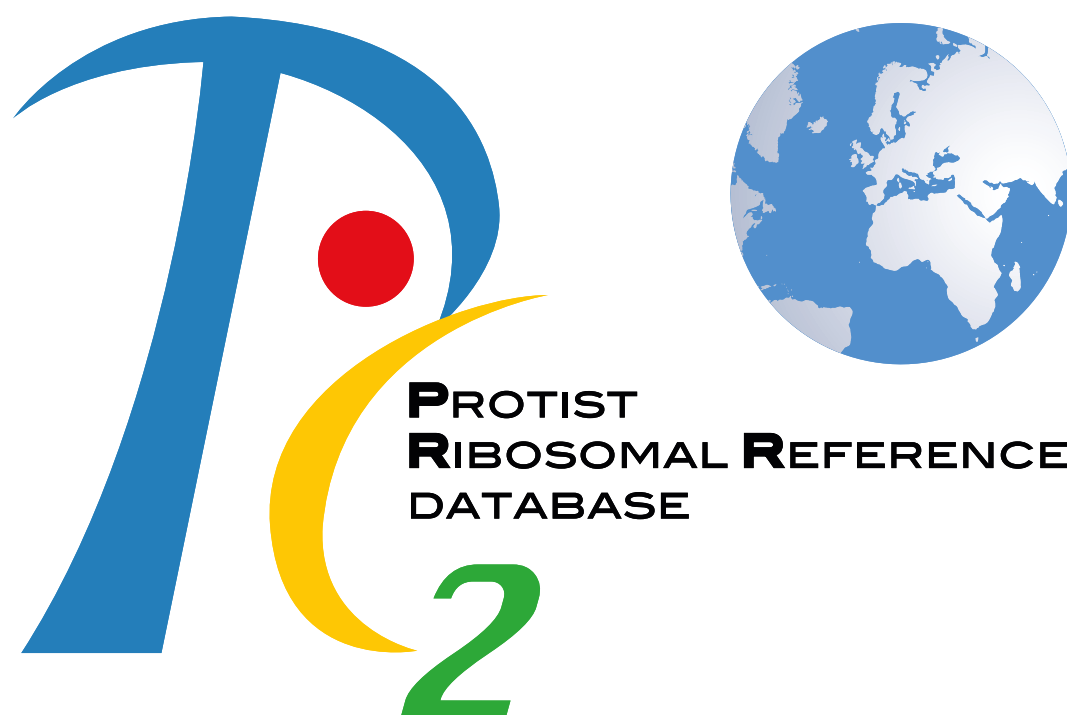
## Chapitre 3

### Création et validation d'une base de référence pour le traitement des données génétiques

[Cf. Annexe Chapitre 3]

**The Protist Ribosomal Reference database (PR<sup>2</sup>): un catalogue des petites séquences de la sous-unité de l'ARNr chez les eucaryotes unicellulaires et leur taxonomie validée**

Published in *Nucleic Acids Research*, 2013, Vol. 41



# Sommaire

|  |            |
|--|------------|
| <b>Chapitre 3 : Création et validation d'une base de référence pour le traitement des données génétiques</b> | <b>87</b>  |
| <b>Introduction</b>  | <b>89</b>  |
| <b>Matériel et méthode</b>   | <b>91</b>  |
| 1. <i>Échantillonnage</i>  | 91         |
| 2. <i>Traitement des données génétiques</i>  | 92         |
| 3. <i>Constitution de la base de référence</i>   | 92         |
| 4. <i>Évaluation de la qualité des assignations</i>  | 93         |
| 5. <i>Analyse des données morphologiques</i>   | 93         |
| <b>Résultats</b>   | <b>95</b>  |
| 1. <i>Base de référence</i>  | 95         |
| 2. <i>Phylogénie</i>   | 95         |
| 3. <i>Évaluation des séquences de référence</i>  | 95         |
| 4. <i>Comparaison metabarcoding V9 / morphologie</i>   | 98         |
| 5. <i>Comparaison metabarcoding V9 / observations FlowCAM pour le genre Neoceratium</i>                      | 99         |
| <b>Discussion</b>  | <b>100</b> |
| 1. <i>La base de référence V9</i>  | 100        |
| 2. <i>Phylogénie</i>   | 100        |
| 3. <i>Comparaison metabarcoding V9 / morphologie</i>   | 101        |
| <b>Conclusion</b>  | <b>102</b> |
| <b>Références</b>  | <b>103</b> |

## Introduction

La dernière décennie a vu une explosion sans précédent du nombre de gènes et de génomes séquencés (<http://genomesonline.org>). Beaucoup de bases de données, accessibles au public, ont été créées pour stocker et distribuer cette information. Les séquences publiées offrent un point de vue unique sur la diversité des organismes et sur leur fonctionnement métabolique, ce qui permet d'effectuer des études de biodiversité et d'écologie fonctionnelle.

Pour être représentative de la réalité cette mine d'informations doit être à la fois précise, fiable, accessible de manière claire et cohérente. Néanmoins, la création et la maintenance d'une base de données de séquences de référence sont sujettes à l'erreur. L'ampleur de ces erreurs ne se mesure pas aisément et l'impact de leurs utilisations est parfois complexe à mesurer.

Plusieurs sources d'erreurs sont possibles : le format des fichiers contenus dans la base de données, l'information complémentaire associée qui caractérise la séquence ou la séquence elle-même qui peut contenir des erreurs scientifiques, syntaxiques ou typographiques. Ces erreurs prennent toute leur ampleur lorsque l'on considère que leur utilisation dans le cadre des projets de recherche peut engendrer de mauvaises interprétations scientifiques.

Ces séquences de référence erronées peuvent par exemple être utilisées pour annoter des séquences nouvellement découvertes. Il est souvent plus facile et plus rapide de caractériser une nouvelle séquence environnementale, par sa comparaison à des séquences homologues plutôt que de compter sur une méthode expérimentale pour la caractériser. C'est ainsi qu'une annotation originale erronée vient rapidement à se diffuser à d'autres séquences engendrant d'autres informations erronées.

La correction d'erreurs dans une base de référence est rare et difficile à apporter. Si, par exemple, des preuves scientifiques indiquent que certaines des conclusions d'une étude sont peut être incorrectes ou incomplètes, une correction n'est pas nécessairement apportée aux séquences publiées. Des émendations de taxa peuvent intervenir suite à la découverte de nouveaux taxons ou à de nouvelles analyses phylogénétiques, mais la mise à jours de leur assignation taxonomique dans les bases de données n'est pas automatiquement apportée. Ce problème se retrouve accentué par le fait que dans plusieurs bases de données de séquences nucléotidiques (comme par exemple NCBI) les données de séquences et les annotations ne peuvent être modifiées que par la personne qui a initialement présenté la séquence. Jusqu'à présent, aucune étude à grande échelle n'a évalué le nombre de ces erreurs dans les bases de données de séquences publiques ou la vitesse à laquelle ces erreurs peuvent se propager.

Lorsque nous avons décidé d'étudier la diversité des séquences environnementales de l'expédition Tara Océans, il est apparu évident qu'avant d'utiliser les informations extraites depuis des

bases de données publiées, il était indispensable d'effectuer un travail de curation et d'homogénéisation minutieux des séquences. Ainsi on pourra disposer d'une référence robuste.

L'objectif de cette partie de mon travail de thèse a été la constitution d'une base de référence, robuste et curée, de séquences pour la région V9 de l'ADN ribosomiale 18S chez les dinoflagellés marins à partir de séquences complètes de 18S déposées dans la base publique de données NCBI (<http://www.ncbi.nlm.nih.gov/>) (Geer *et al.* 2009). Une validation de cette base de référence a été effectuée en comparant les résultats des assignations taxonomiques des séquences obtenues par metabarcoding (en utilisant la région V9) à des listes d'espèces (réalisées par des taxonomistes) et à des comptages cellulaires de dinoflagellés (réalisés par FlowCAM). Ces données ont été obtenues à partir de l'analyse d'échantillons vivants ou fixés, prélevés en même temps que les échantillons génétiques dans le cadre du projet Tara Océans. Les analyses en microscopie optique ont été effectuées et mises à disposition par le Dr. Eleonora Scalco et le Dr. Adriana Zingone de la Stazione Zoologica de Naples, membres du consortium Tara.

## Matériel et méthode

### 1. Échantillonnage

Pour cette étude, nous avons utilisé des échantillons issus de 33 stations de l'expédition Tara Océans collectés pendant une période de 2 ans allant du 15 Septembre 2009 au 13 mai 2011. Ces échantillons portent sur 4 fractions de taille de protistes, de la plus petite à la plus grande (0,8-5 $\mu$ m ; 5-20 $\mu$ m ; 20-180 $\mu$ m ; 180-2000 $\mu$ m) et couvrent 7 bassins océaniques (l'Océan Atlantique Nord (1 station), la mer Méditerranée (6), la mer Rouge (4), l'océan Indien (8), l'océan Atlantique Sud (8), l'océan Austral (2) et l'océan Pacifique Sud (4)). Les échantillons ont été recueillis en surface (< 5m) à l'aide d'une pompe péristaltique pour la plus petite fraction de taille (0,8-5 $\mu$ m), et d'une série de filets à plancton avec un maillage adapté pour les trois grandes fractions de taille (5 à 2 000 $\mu$ m). L'eau pompée a été pré-filtrée à travers un système de filtration, nommé Gravity Plankton Sieving System (GPSS), constitué d'une superposition de trois filets, aux mailles successives de 20 $\mu$ m, 5 $\mu$ m, et de nouveau 5 $\mu$ m (Not et al., soumis), dans le but de bien séparer la plus petite fraction du plancton (inférieures à 5 $\mu$ m) des cellules les plus grosses. Pour récupérer la fraction de taille 0,8-5 $\mu$ m, pour chaque station, deux bidons de 100L ont ainsi été collectés puis filtrés en parallèle à travers deux membranes en polycarbonate de porosité 0,8 $\mu$ m et de diamètre 142mm. Ces membranes (4 par échantillon) ont alors été immédiatement placées dans un cryotube de 5mL et stockées directement dans de l'azote liquide. Pour les trois fractions de taille supérieures à 5 $\mu$ m, des filets à plancton de maille 5, 20, et 180 $\mu$ m ont été trainés pendant une période d'environ 15min, puis remontés à bord et rincés de l'extérieur avec de l'eau de mer filtrée (0,1 $\mu$ m). Ces échantillons de plancton ont été passés sur des tamis appropriés (5, 20 ou 180 $\mu$ m) et versés dans une bouteille Nalgen de 8L. Le volume a été ajusté à 3L avec l'eau de mer filtrée (0,1  $\mu$ m). Un volume de 0,5L de ce plancton concentré a alors été filtré à travers deux membranes en polycarbonate de 47mm (de porosité 5, 20 ou 180 $\mu$ m) lesquelles sont ensuite récupérées et placées dans un cryotube de 5mL immédiatement stocké dans de l'azote liquide. Tous les cryotubes ont été stockés à -80°C lors de leur arrivée dans les laboratoires à terre jusqu'à leur traitement. Le protocole détaillé est disponible dans Not et al. (soumis) (cf. chapitre 1).

Pour l'analyse en microscopie, 50mL d'eau collectés depuis le filet 20 $\mu$ m ont été versés dans des bouteilles en plastique opaques de 60mL pré-remplies avec 1mL de Lugol et stockées à 4°C à bord du bateau, puis au laboratoire. Une fixation au formol a également été réalisée, 45mL d'eau ont été versés dans un tube Falcon de 50mL pré-rempli avec 5mL de formol neutralisé à 37%. Les échantillons ont été stockés à température ambiante à bord du bateau, puis à 4°C au laboratoire.

Les échantillons analysés pour construire les 2 jeux de données utilisés dans cette étude, un morphologique et un génétique, ont été obtenus synchroniquement à partir d'un unique trait de filet de 20 $\mu$ m.



### 2. Traitement des données génétiques

L'extraction de l'ADN a été faite avec le kit Nucleospin® DNA II (Macherey-Nagel, Hoerd, France). Les régions hypervariables V9 du gène de l'ARN ribosomique 18S ont été amplifiées avec les amorces eucaryotes 1389f 5'-TTGTACACACCGCCC-3' et 1510r 5'-CCTTCYGCAGGTTACCTAC-3' (Amaral-Zettler *et al.* 2009). Les amplifications ont été réalisées avec la DNA Polymerase à haute fidélité Phusion® (Finnzymes). La réaction PCR a été effectuée dans un volume final de 25µL contenant 5ng de matrice à une concentration finale de 0.35µM de chaque amorce, 3% de DMSO et 2X de tampon GC Phusion Master Mix (Finnzymes). Les amplifications ont été réalisées en suivant un programme de PCR constitué par une étape de dénaturation initiale à 98°C pendant 30 sec, suivie de 25 cycles de 10 sec à 98°C, 30 sec à 57°C, 30 sec à 72°C, et d'une étape d'élongation finale à 72°C pendant 10 minutes. Chaque échantillon a été amplifié en trois exemplaires pour obtenir des quantités suffisantes d'amplicons. Les produits PCR sont contrôlés par migration sur gel d'agarose à 1,5% pour vérifier le succès de l'amplification et la longueur des fragments obtenus. Les amplicons ont ensuite été regroupés et purifiés en utilisant le kit NucleoSpin® Extrait II (Macherey-Nagel, Hoerd, France) et envoyés au CEA Génoscope d'Evry (France) pour séquençage. Les fragments amplifiés ont été pyro-séquencés en utilisant un système Genome Analyser Iix (Illumina, San Diego, CA, USA) double sens (*paired ends*).

Les séquences utilisées pour les analyses ont été comparées à toutes les séquences de référence de la base de données en utilisant un algorithme d'alignement exact global par paires (Needleman & Wunsch 1970). Chaque séquence a reçu l'assignation taxonomique de son plus proche voisin dans la référence (ou du dernier ancêtre commun en cas d'égalité) de la base de données PR2 (Guillou *et al.* 2013). Les alignements ont été calculés avec GGSearch, un outil de la version 36 du programme FASTA (Pearson & Lipman 1988), en utilisant les paramètres par défaut. Les résultats ont été stockés dans une base de données SQLite.

### 3. Constitution de la base de référence

Toutes les séquences de dinoflagellés déposées sur la base de données NCBI (<http://www.ncbi.nlm.nih.gov/>) (Geer *et al.* 2009) avant le 30 Juin 2012 ont été récupérées pour construire une base de données de référence des régions V4 et V9 de l'ADNr 18S. Un alignement de toutes les séquences (3 339 séquences) de dinoflagellés présents dans la base NCBI a été effectué à l'aide du logiciel Mafft (v.7 en mode automatique : <http://mafft.cbrc.jp/alignment/server/>) (Katoh & Toh 2010; Katoh & Standley 2013). Seules les séquences contenant les régions V4 et V9 ont été conservées (1 191 séquences). Dans le cadre du projet Tara Océans, seule la région V9 de l'ADNr 18S a été séquencée (Karsenti *et al.* 2011), mais le fragment V4 a été conservé pour la référence dans le but d'avoir une plus grande résolution pour l'assignation taxonomique. Afin d'identifier la présence

éventuelle de chimères dans la région V9, toutes les séquences ont été soumises à l'outil KeyDNATools (<http://keydnatools.com>) (Guillou *et al.* 2013). L'assignation taxonomique des séquences a été manuellement et au cas par cas vérifiée, puis a été contrôlée à partir des bases de données taxonomiques publiques de référence telle que AlgaeBase (<http://www.algaebase.org>) (Guiry & M 2012) et WoRMS (<http://www.marinespecies.org/>) (WoRMS Editorial Board 2012).

Un arbre phylogénétique en Neighbor Joining (NJ) a été construit (méthode des distances sans gap) avec le logiciel SeaView (Gouy *et al.* 2010). Les clades de séquences générés par cette phylogénie ont été nommés hiérarchiquement par ordres, familles, genres et espèces. En cas de taxons polyphylétiques, les clades ont été nommés avec le même nom hiérarchique mais avec des numéros successifs et finalement unifiés pour faciliter la suite des analyses. Ces séquences forment la base de référence DinR2 (Dinoflagellate ribosomal reference database). Ces séquences de référence ont également été incluses dans la base de données Protist Ribosomal Reference (PR2) (Guillou *et al.* 2013) (cf Annexe — Chapitre 3). L'intégration de la taxonomie y a été réalisée sans les informations de clades.

#### 4. *Évaluation de la qualité des assignations*

La longueur de chaque séquence V9 contenue dans la base de référence a été calculée. Pour chaque ordre et genre un dénombrement par longueur possible de séquence a été réalisé. Une normalisation par pourcentage a été réalisée.

Les séquences environnementales obtenues après séquençage ont été utilisées pour évaluer leur pourcentage d'identité avec les références. Pour chaque pourcentage d'identité (compris entre 100 et 90) et pour chaque ordre ou genre, le nombre de séquences environnementales a été comptabilisé et pondéré par leur abondance. Les valeurs obtenues par ordre ou genre ont ensuite été normalisées en pourcentage.

#### 5. *Analyse des données morphologiques*

Une identification taxonomique et un comptage des dinoflagellés présents dans une série de 12 échantillons environnementaux prélevés par filet en surface dans l'océan Atlantique Sud ont été réalisés. Ces échantillons de la fraction 20-180 $\mu$ m ont été fixés au formol. Pour chaque observation, 3mL d'échantillon ont été placés dans une chambre Utermöhl avec une goutte de colorant calcofluor (1:100.000), qui colore la cellulose permettant ainsi de mieux détecter et identifier les espèces de dinoflagellés. Les cellules, tombant dans 2 ou 4 transects de la chambre, ont été identifiées et dénombrées. Comme ces échantillons ont été prélevés par filet (Bongo), seule une estimation de l'abondance (%) a pu être obtenue. Les espèces ont été identifiées et dénombrées à l'aide d'un microscope optique inversé (Carl Zeiss Axiophot200) à un grossissement de 400x. L'identification a été

effectuée au niveau de l'espèce, si possible. Les espèces ont été regroupées en 4 groupes : les diatomées, les dinoflagellés, les coccolithophoridés et les autres cellules du plancton.

Ces données morphologiques de la fraction 20-180µm ont été analysées par le Dr. Adriana Zingone et Dr. Eleonora Scalco de la Stazione Zoologica de Naples (Italie) et mises à notre disposition (Fig 3) sous la forme d'un tableau bilan Excel.

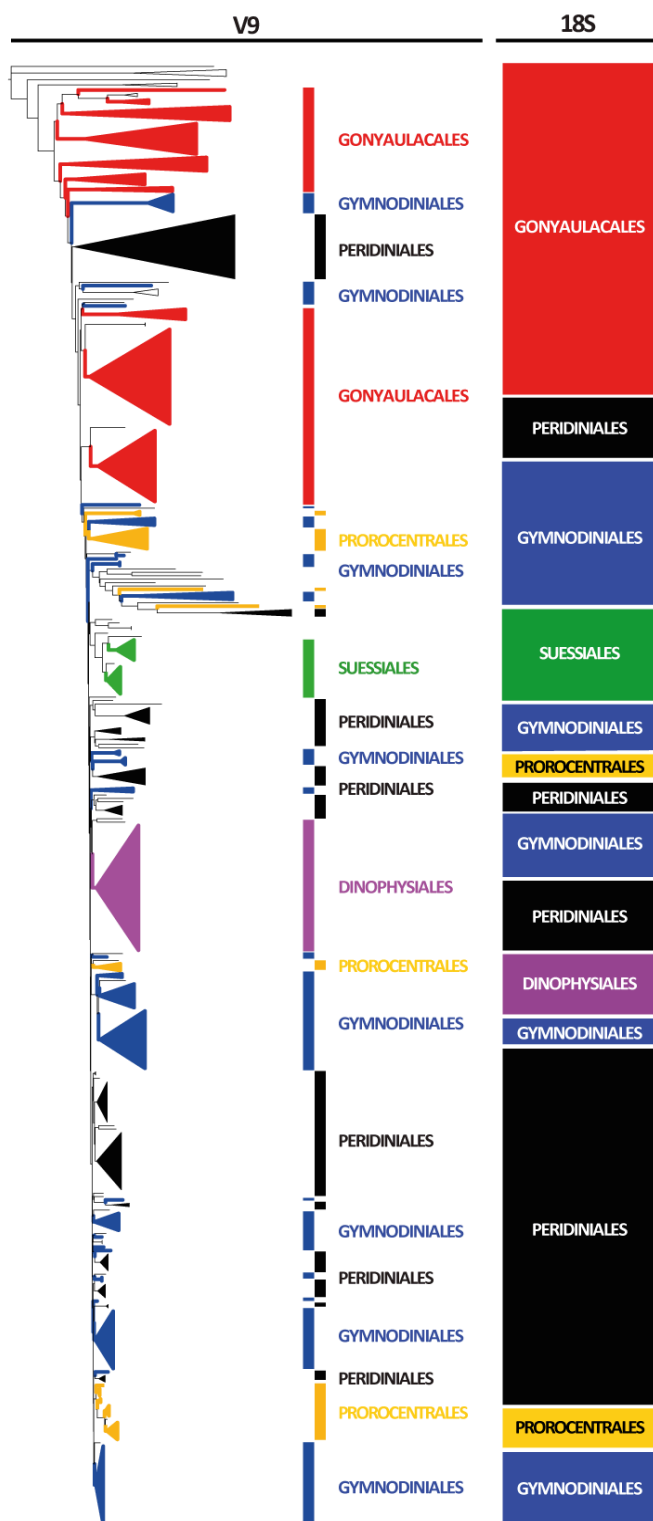


Figure 1 : Comparaison de la topologie de l'arbre V9 et 18S complet chez les dinoflagellés.

## Résultats

### 1. Base de référence

La base de référence est composée de 329 espèces réparties dans 97 genres et 13 ordres : Blastodinales (2 genres|2 espèces), Dinophysiales (7|34), Gonyaulacales (14|78), Gymnodinales (25|68), Lophodinales (1|2), Noctilucales (4|5), Oxyrrhinales (1|1), Peridinales (30|93), Phytodinales (3|4), Prorocentrales (1|22), Pyrocystales (1|2), Suessiales (5|15), Thoracosphaerales (1|1). Une catégorie Uncultured (21|21) et Dinophyceae (2|2) viennent compléter cet ensemble. Les séquences identifiées comme Uncultured correspondent à des séquences environnementales appartenant à la classe des Dinophyceae formant des clades clairement distinguables. Le groupe Dinophyceae correspond à deux genres connus de dinoflagellés pour lesquels aucun ordre n'est admis.

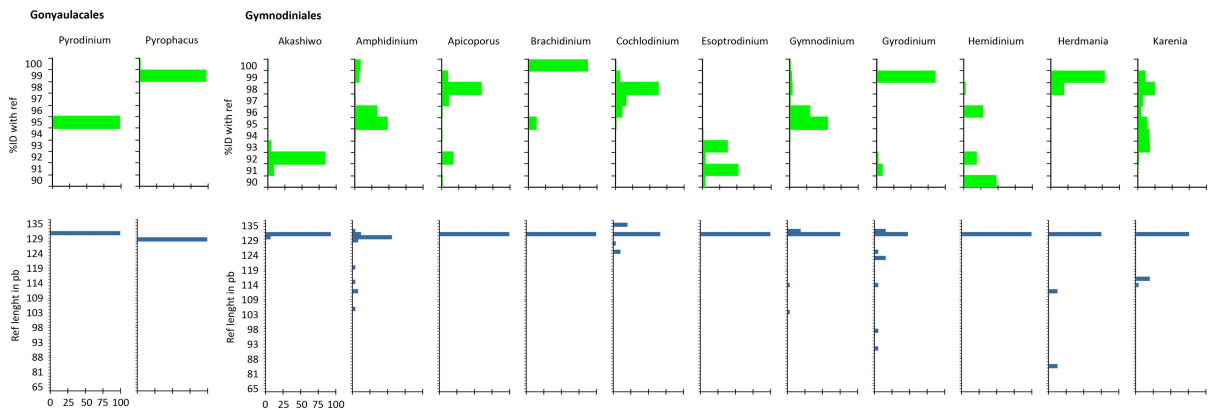
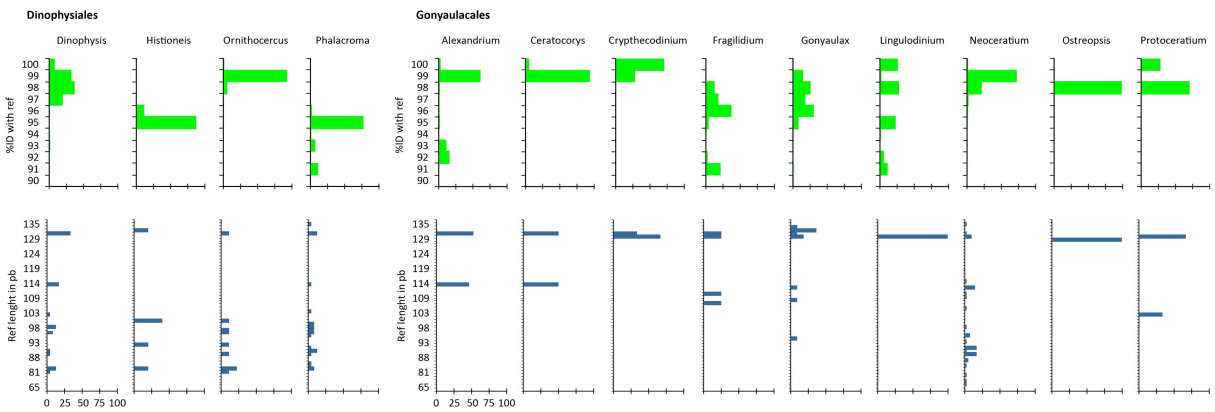
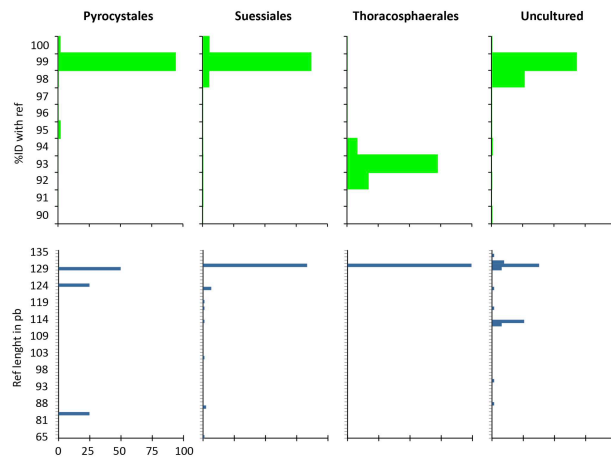
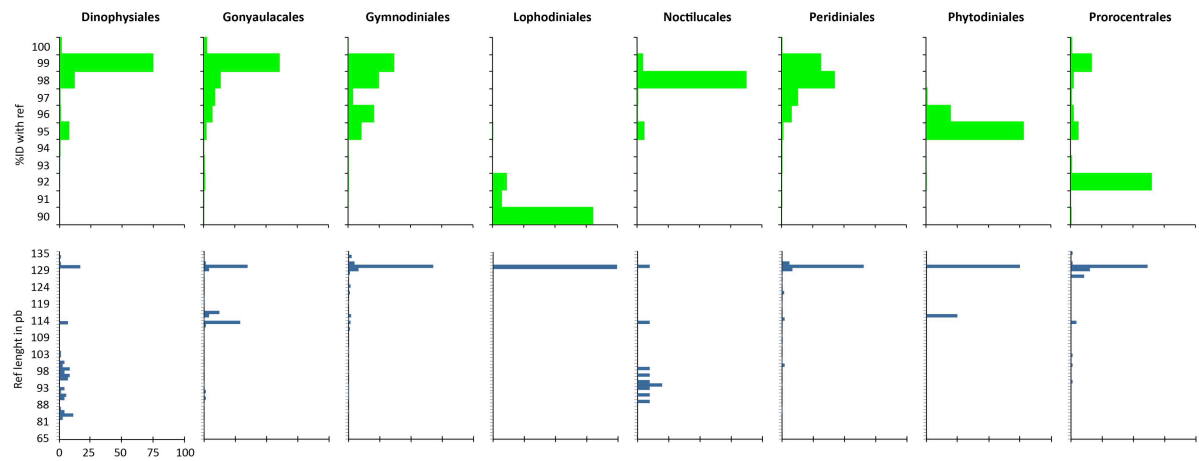
### 2. Phylogénie

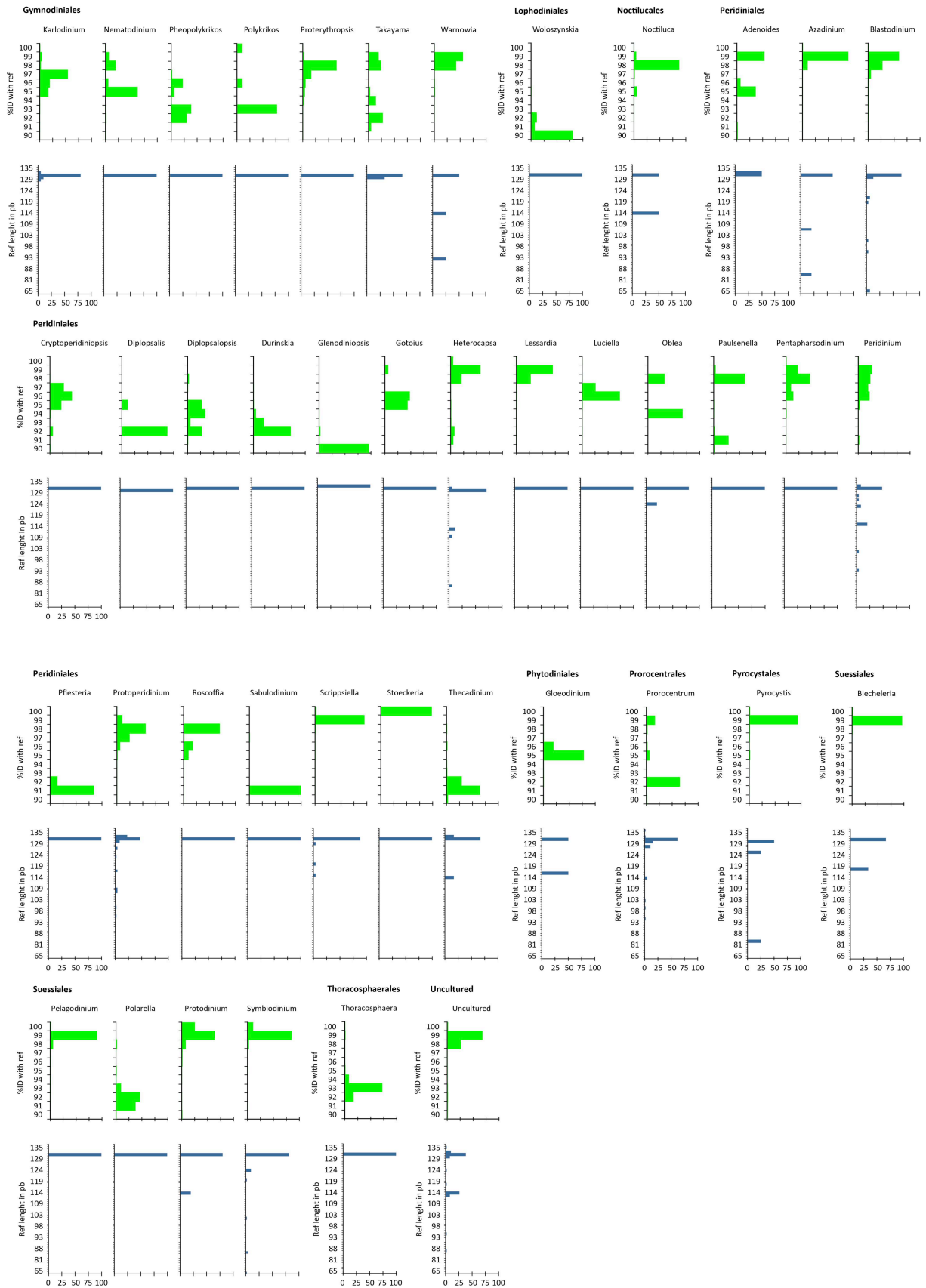
La topologie de l'arbre obtenue à partir de ces séquences de référence V9 de l'ADNr 18S montre une phylogénie polyphylétique pour les ordres Gymnodinales, Prorocentrales, Peridinales et Gonyaulacales. Chez l'ensemble des dinoflagellés, à l'exception de Gonyaulacales, la polyphylie au niveau de l'ordre est évidente même lorsque l'on considère l'ensemble du 18S (Fig. 1).

### 3. Évaluation des séquences de référence

La qualité de l'assignation taxonomique des code-barres obtenue au niveau des différents ordres et genres a été évaluée. Pour chaque ordre et genre la distribution du pourcentage d'identité des séquences environnementales avec les séquences de référence a été déterminée (graphiques verts Fig. 2). La variabilité des longueurs des séquences de référence a également été répertoriée (graphiques bleues Fig. 2).

Trois ordres (sur onze) présentent des pourcentages d'identité à moins de 95% : Lophodinales, Thoracosphaerales et Prorocentrales. Pour ces trois ordres les séquences de référence sont cependant dans la moyenne des longueurs de V9 à savoir  $\pm 130$ pb. Au niveau du genre, 16 genres sur 67 présentent des pourcentages d'identité majoritairement inférieurs à 95%. Ce phénomène est observé dans plusieurs genres de plusieurs ordres comme : *Akashiwo*, *Esoptrodinium*, *Hemidinium*, *Pheopolykrikos*, *Polykrikos* (Gymnodinales) ; *Diplopsalis*, *Diplopsalopsis*, *Durinskia*, *Glenodiniopsis*, *Pfiesteria*, *Sabulodinium*, *Thecadinium* (Peridinales) ; ou encore *Polarella* (Suessiales).

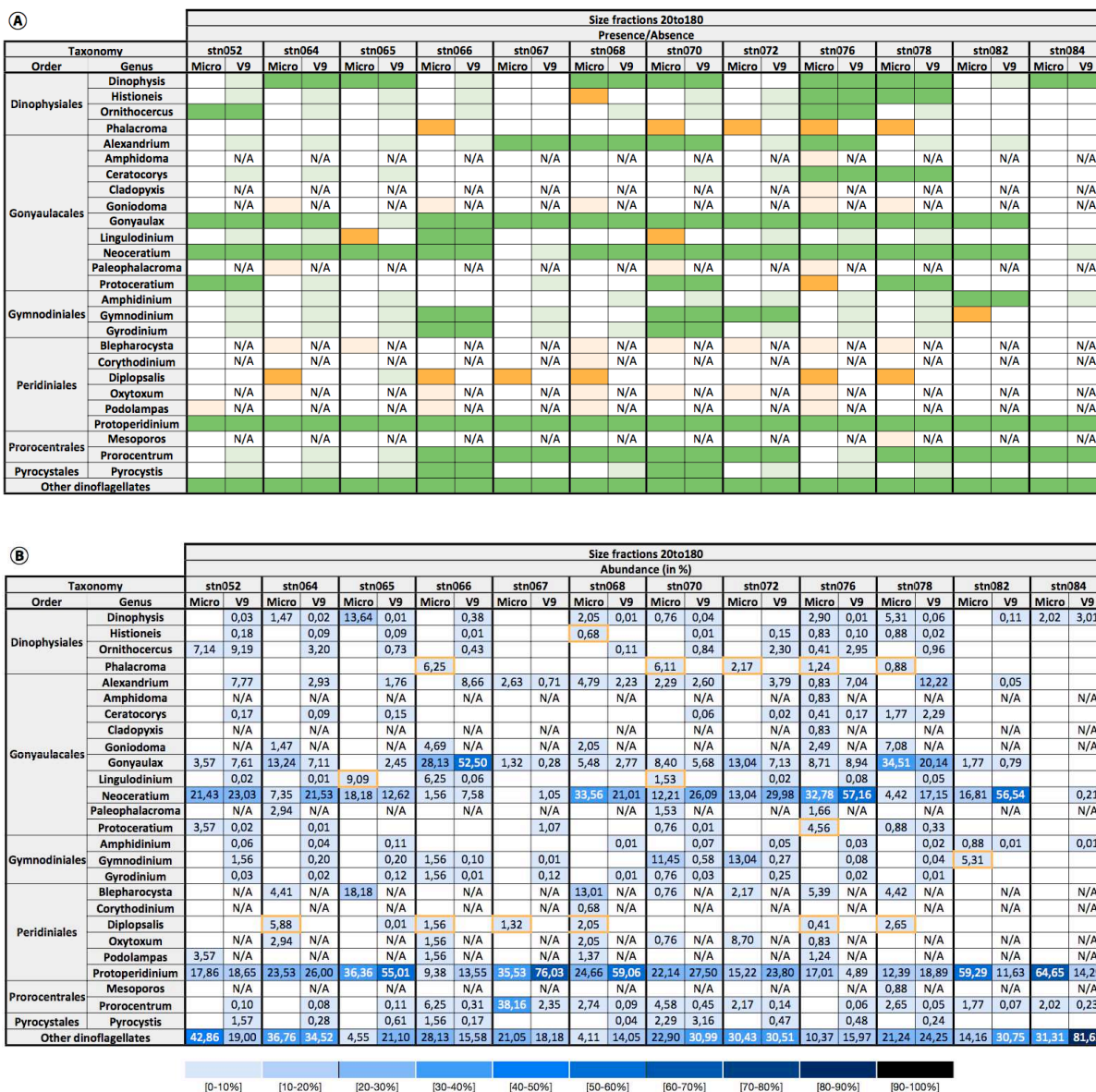




**Figure 2 :** Distribution des pourcentages d'identité des séquences environnementales par ordre et genre (graphiques supérieurs en vert) et longueurs des séquences de référence (pb) correspondantes (graphiques inférieurs en bleu). L'axe X est en pourcent.

4. Comparaison metabarcoding V9 / morphologie

La comparaison des jeux de données génétiques et morphologiques repose sur la liste des espèces identifiées par observation morphologique dans 12 stations situées dans l'océan Atlantique Sud sur une période de 9 mois (du 17 mai 2010 au 3 janvier 2011). Les espèces non identifiées par microscopie sont rassemblées sous la dénomination « Other dinoflagellates ». Le nombre de genres détectés par l'outil moléculaire (metabarcoding) est en général supérieur au nombre de genres identifiés par l'observation visuelle (microscopie) des échantillons.



**Figure 3 :** **A**) Identification en présence/absence des différents genres de dinoflagellés. En vert foncé sont représentés les genres observés en metabarcoding et en microscopie, en vert clair uniquement en metabarcoding, en orange foncé uniquement en microscopie, en orange clair uniquement en microscopie mais en absence de référence pour le metabarcoding. **B**) Abondance relative par station des différents genres de dinoflagellés. Un bleu soutenu indique une forte abondance du genre. Les cases clérées de orange marquent les genres uniquement détectés en microscopie.

Sur l'ensemble des échantillons considérés, 15 genres sur les 26 analysés sont détectés dans au moins une station par les deux méthodes (Fig. 3A). Sur ces 11 genres restants, 9 n'ont pas de références génétiques et donc n'ont pas pu être identifiés dans le jeu de données moléculaires. Les deux genres disposant d'une référence et non détectés par les deux méthodes (au sein d'une même station) sont les genres *Phalacroma* et *Diplopsalis*. Le genre *Phalacroma* est uniquement observé par microscopie. Le genre *Diplopsalis*, régulièrement identifié en microscopie, a lui été détecté une seule fois par metabarcoding dans la station 65 où il apparaît comme absent par l'analyse morphologique.

Deux ordres sont identifiés comme dominants pour l'ensemble des échantillons de cette analyse et par les deux méthodes abordées (Fig. 3B) : les Gonyaulacales et les Peridinales. L'ordre des Peridinales est majoritairement dominé par le genre *Protoperidinium* présent dans l'ensemble des échantillons. L'ordre des Gonyaulacales est majoritairement représenté par les genres *Gonyaulax* et *Neoceratium* présents dans plus de 80% des échantillons.

#### 5. Comparaison metabarcoding V9 / observations FlowCAM pour le genre *Neoceratium*

Trois échantillons, de la fraction 20-180µm, issus d'un trait de filet en surface en mer Méditerranée ont été analysés au niveau spécifique pour le genre *Neoceratium* par deux approches distinctes : le FlowCAM (cf protocole dans le Chapitre 2) et le metabarcoding. Sur les 13 espèces de *Neoceratium* identifiées par FlowCAM aucune ne dispose de séquences de référence dans notre base de référence V9 (Tab 1). Seule trois séquences de *Neoceratium* sont disponibles en V9 : *N. longipes*, *N. massiliense* et *N. tenue*. La comparaison de la résolution entre les deux approches au niveau spécifique est donc impossible à faire, mais on remarque que dans les trois stations la présence du genre *Neoceratium* a bien été détectée par l'approche moléculaire.

|                                 | Stn023  |       | Stn025  |       | Stn030  |       |
|---------------------------------|---------|-------|---------|-------|---------|-------|
|                                 | FlowCAM | V9    | FlowCAM | V9    | FlowCAM | V9    |
| <i>Neoceratium candelabrum</i>  | 0       | N/A   | 1       | N/A   | 0       | N/A   |
| <i>Neoceratium contortum</i>    | 0       | N/A   | 0       | N/A   | 0       | N/A   |
| <i>Neoceratium declinatum</i>   | 2       | N/A   | 4       | N/A   | 0       | N/A   |
| <i>Neoceratium extensum</i>     | 1       | N/A   | 0       | N/A   | 0       | N/A   |
| <i>Neoceratium furca</i>        | 2       | N/A   | 2       | N/A   | 0       | N/A   |
| <i>Neoceratium fusus</i>        | 19      | N/A   | 2       | N/A   | 4       | N/A   |
| <i>Neoceratium horridum</i>     | 0       | N/A   | 0       | N/A   | 11      | N/A   |
| <i>Neoceratium macroceros</i>   | 0       | N/A   | 1       | N/A   | 1       | N/A   |
| <i>Neoceratium minutum</i>      | 0       | N/A   | 0       | N/A   | 1       | N/A   |
| <i>Neoceratium pentagonum</i>   | 6       | N/A   | 2       | N/A   | 0       | N/A   |
| <i>Neoceratium undetermined</i> | 3       | N/A   | 8       | N/A   | 4       | N/A   |
| <i>Neoceratium teres</i>        | 3       | N/A   | 1       | N/A   | 15      | N/A   |
| <i>Neoceratium tripos</i>       | 1       | N/A   | 0       | N/A   | 0       | N/A   |
| <i>Neoceratium longipes</i>     |         | 0     |         | 0     |         | 0     |
| <i>Neoceratium massiliense</i>  |         | 0     |         | 0     |         | 0     |
| <i>Neoceratium tenue</i>        |         | 52961 |         | 19576 |         | 43057 |

**Tableau 1** : Comparaison des abondances en *Neoceratium* par identification en imagerie à haut débit et en metabarcoding.



### Discussion

#### 1. Phylogénie

Il a été montré que l'extrémité 3' (sur une longueur pouvant aller jusqu'à 72pb) de l'ADNr 18S chez les dinoflagellés est relativement conservée (Ki 2011). Cette faible proportion de caractères informatifs en fin de V9 associée à sa longueur assez courte ( $\pm 130$ pb) aboutit à une faible résolution phylogénétique pour l'ensemble du groupe des dinoflagellés (Saunders *et al.* 1997; Saldarriaga *et al.* 2004). La polyphylie qui est observée dans l'arbre V9 est ainsi due aux taux d'évolution très faibles du 18S (Gómez 2013). Cependant, cette région V9 reste l'une des plus informative de l'ADNr 18S (Ki 2011) devancée par la région V2.

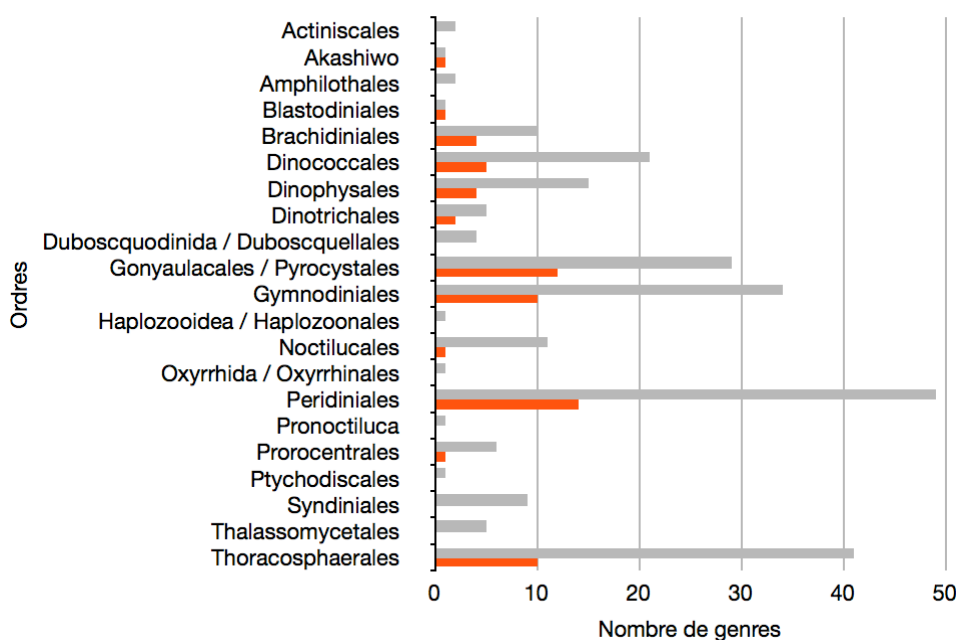
#### 2. La base de référence DinR2

Sur les 2 377 espèces reconnues de dinoflagellés (Gómez 2012a; b; 2013) seules 329 sont représentées dans notre base de référence (Fig. 4). Ce nombre relativement bas s'explique par le fait que pour beaucoup de dinoflagellés nous ne disposons que des informations morphologiques. Aujourd'hui la description taxonomique d'une espèce est associée au dépôt dans une base de données publique d'une ou plusieurs séquences d'identification, mais cette association entre données génétiques et morphologiques n'existe pas pour la plupart des espèces de dinoflagellés qui ont été décrites morphologiquement il y a un certain temps. Encore aujourd'hui, l'observation d'espèces de dinoflagellés, dans le cadre de campagnes océanographiques, de projets de surveillance ou de recherche, est majoritairement réalisée par des observations en microscopie d'échantillons planctoniques. Il est relativement rare que des observations microscopiques soient associées à des analyses génétiques. Une telle stratégie permettrait pourtant clairement de renforcer le nombre de séquences à disposition et enrichirait les bases de données de séquences publiques. Notre base de séquences de référence (DinR2) couvre cependant une large majorité des principaux ordres connus de dinoflagellés ainsi que les principaux genres (les plus divers) ce qui garantit une identification correcte du groupe dans son ensemble (Fig. 4). Le faible nombre de séquences de référence pour certains genres, comme le genre *Neoceratium* par exemple, permet tout de même de détecter leur présence au sein d'un échantillon (Tab. 1).

#### 3. Qualité de l'assignation

Les plus faibles pourcentages d'identité des séquences environnementales obtenus pour certains ordres ou genres avec la base de référence DinR2 peuvent probablement s'expliquer par un manque de diversité dans nos références (Fig. 2 et 4). En effet, la présence de peu ou d'une unique séquence de référence pour décrire un ordre ou un genre peut ne pas représenter sa diversité totale.

Cependant, il est intéressant de noter que les genres qui ont un faible pourcentage d'identité sont tout de même identifiés et qu'ils restent donc exploitables dans l'analyse des metabarcodes.



**Figure 4 :** Nombre de genres par ordre de dinoflagellés (aligné sur la classification de Gomez). En gris son représentés les genres connus de dinoflagellés pour chaque ordre (Gómez 2012a) et en rouge les genres présents dans la base de référence PR2 (Guillou *et al.* 2013).

#### 4. Comparaison metabarcoding V9 / morphologie

Sur l'ensemble des données analysées, une très large majorité des genres a été identifiée par l'analyse moléculaire. Si on enlève les 9 genres pour lesquels on n'a pas de références génétiques, 15 genres sur 17 sont repérés par metabarcoding. L'approche moléculaire est plus précise que l'analyse morphologique car un nombre plus élevé de genres a été identifié. Ce nombre est plus élevé même si plusieurs références moléculaires sont manquantes dans la base de référence DinR2. Ces références manquantes ne concernent que des genres non dominants du point de vue de l'abondance (Fig 3B). Les genres qui ont été caractérisés par l'approche moléculaire mais qui n'ont pas été observés en microscopie pourraient s'expliquer par une moins bonne conservation de leur structure externe (et donc une réelle difficulté à les identifier sur des critères morphologiques) dans des échantillons fixés ou par un seuil de détection bien plus élevé (précis) dans le cas de l'analyse moléculaire.

Cependant, dans certaines stations des genres identifiés par morphologie n'ont pas été identifiés par metabarcoding.

Les groupes taxonomiques non détectées par l'outil génétique (alors que la référence existe) pourraient être expliqués par plusieurs hypothèses : la première serait due à l'assignation que l'on

pourrait qualifier d'« intuitive » (donc imprécise) faite par le taxonomiste, comme pour le genre *Diplopsalis* qui est identifié comme un groupe de genres plus que comme un véritable genre indépendant. La seconde pourrait être un problème d'amplification lié à une moins bonne fixation des amorces ou à un faible nombre de copies du V9 chez certains genres.

### Conclusion

La comparaison d'échantillons analysés à la fois par microscopie et par metabarcoding montre que presque la totalité des genres identifiés visuellement a été retrouvée par l'approche génétique. Ce résultat prouve l'efficacité de l'approche moléculaire et que le travail d'annotation des séquences a été efficace. L'approche par metabarcoding montre une caractérisation plus complète de la diversité des dinoflagellés dans les échantillons analysés. En effet, un plus grand nombre de genres ont ainsi été identifiés. Cependant, un certain nombre de genres, qui peuvent être distingués en microscopie, n'ont pas pu être identifiés par metabarcoding car ceux-ci ne sont pas représentés dans notre base de référence. Cette absence, qui est due à un manque de séquences pour un certain nombre de genres dans les bases publiques (comme NCBI), devra être résolue. Les genres les plus abondants sont cependant correctement représentés.

La précision de la mesure de l'abondance au niveau des genres peut être discutée, mais on observe une bonne corrélation au niveau des ordres. Une analyse plus approfondie à partir de cette base de référence V9 de l'ADN 18S au niveau de l'ordre semble donc pertinente.

---

### Remerciements

Je remercie tout particulièrement le Dr. Adriana Zingone et Dr. Eleonora Scalco pour leur contribution importante à ce travail. Sans eux la validation de la base de référence n'aurait pas été possible. Merci également à Marie-José Garet pour son aide lors de la collecte des séquences de référence sur NCBI.

## Références

- Amaral-Zettler L, McCliment EA, Ducklow HW, Huse SM (2009) A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, **4**, e6372.
- Geer LY, Marchler-Bauer A, Geer RC *et al.* (2009) The NCBI BioSystems database. *Nucleic acids research*, **38**, D492–D496.
- Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular biology and evolution*, **27**, 221–224.
- Gómez F (2012a) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *CICIMAR Océánides*, **27**, 65–140.
- Gómez F (2012b) A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity*, **10**, 267–275.
- Gómez F (2013) Problematic Biases in the Availability of Molecular Markers in Protists: The Example of the Dinoflagellates. *Acta Protozoologica*, **52**, 21–33.
- Guillou L, Bachar D, Audic S *et al.* (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic acids research*, **41**, D597–D604.
- Guiry MD, M GG (2012) Algaebase: Listing the World's Algae. *World-wide electronic publication, National University of Ireland, Galway. www.algaebase.org searched on August 2012.*
- Karsenti E, Sullivan MB, Acinas SG *et al.* (2011) A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biology*, **9**, e1001177.
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution*, **30**, 772–780.
- Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, **26**, 1899–1900.
- Ki J-S (2011) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. *Journal of Applied Phycology*, **24**, 1035–1043.
- Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, **48**, 443–453.
- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 2444–2448.
- Saldarriaga JF, Taylor FJR, Cavalier-Smith T, Menden-Deuer S, Keeling PJ (2004) Molecular data and the evolutionary history of dinoflagellates. *European Journal of Protistology*, **40**, 85–111.
- Saunders GW, Hill D, Sexton JP, Andersen RA (1997) Small-subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. In: *Origins of algae and their plastids*. Origins of Algae and their Plastids.
- WoRMS Editorial Board (2012) WoRMS - World Register of Marine Species. Available from [www.marinespecies.org](http://www.marinespecies.org) at VLIZ. Accessed August 2012.



## **Partie 2**

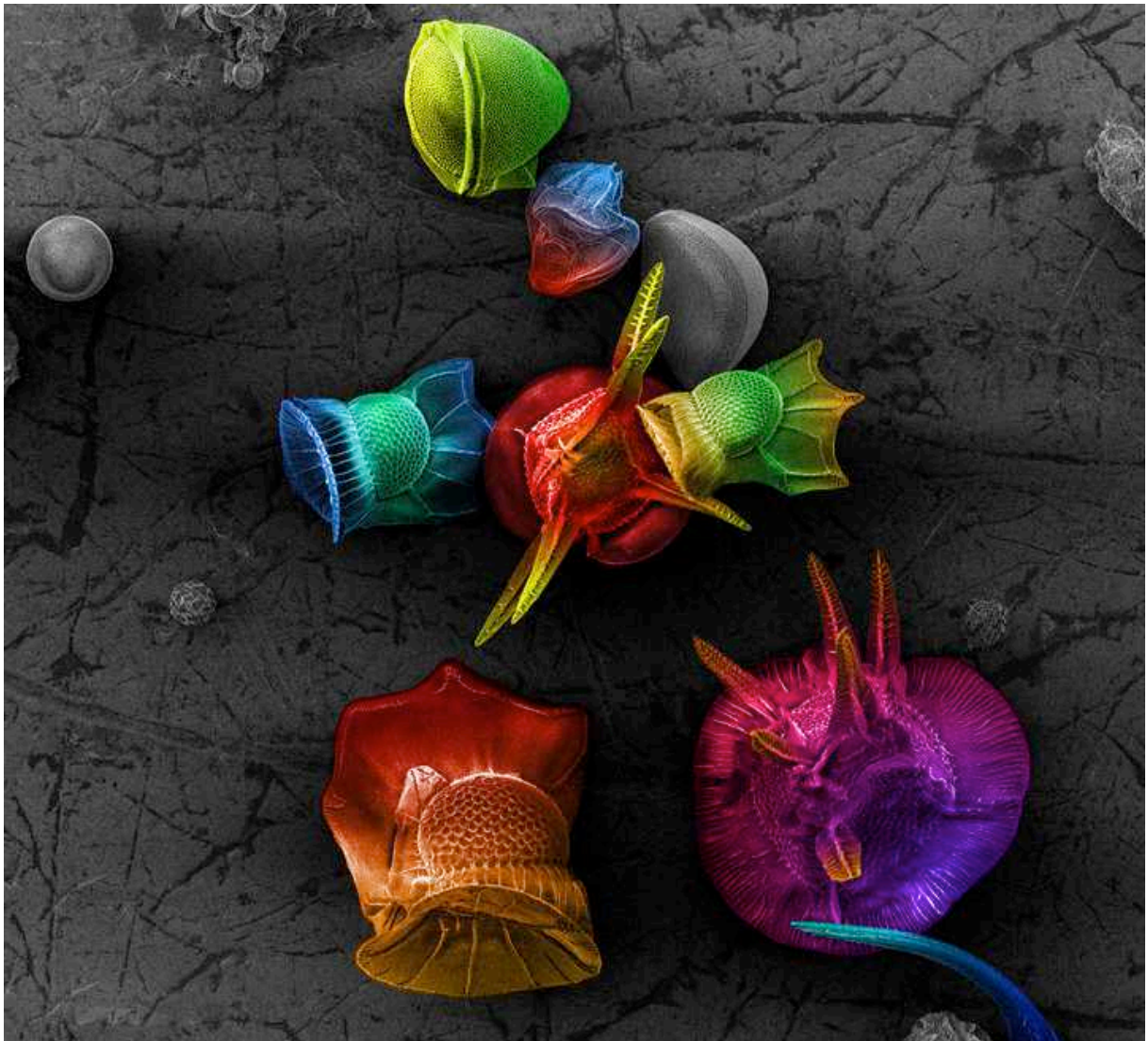
# **Analyse de la biodiversité et de l'écologie des dinoflagellés planctoniques marins par metabarcoding**



# Chapitre 4

## Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding

Submitted to *Environmental Microbiology*



© André Rochon - Université du Québec à Rimouski



# Sommaire

|   |            |
|---|------------|
| <b>Chapitre 4 : Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding</b> | <b>107</b> |
| <b>Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding</b>              | <b>109</b> |
| <b>Abstract</b>   | <b>109</b> |
| <b>Introduction</b>   | <b>110</b> |
| <b>Experimental procedures</b>  | <b>112</b> |
| Sampling  | 112        |
| Genomic DNA extraction, purification and sequencing   | 113        |
| Database of reference dinoflagellate 18S rDNA and metabarcoding annotation  | 115        |
| Assessment of molecular biodiversity  | 115        |
| <b>Results</b>  | <b>116</b> |
| Dinoflagellates versus protists   | 116        |
| Dinoflagellate community characterization   | 117        |
| Pico-dinoflagellates  | 120        |
| Nano-dinoflagellates  | 120        |
| Micro-dinoflagellates   | 121        |
| Dinoflagellate diversity in meso-plankton   | 121        |
| <b>Discussion</b>   | <b>125</b> |
| <b>Concluding remarks</b>   | <b>132</b> |
| <b>References</b>   | <b>133</b> |

---

Submitted to *Environmental Microbiology*

## Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding

Noan Le Bescot<sup>1</sup>, Frédéric Mahé<sup>2</sup>, Stéphane Audic<sup>1</sup>, Dimier Céline<sup>1</sup>, Marie-José Gareth<sup>1</sup>, Mikael Katinka<sup>3</sup>, Patrick Wincker<sup>3</sup>, Colombran de Vargas<sup>1</sup>, Raffaele Siano<sup>4</sup>

<sup>1</sup> CNRS UMR 7144 & Université Pierre & Marie Curie, EPEP - Evolution of Protists and Pelagic Ecosystems, Station Biologique de Roscoff, 29682 Roscoff, France

<sup>2</sup> Department of Ecology Technische Universität Kaiserslautern Erwin-Schrödinger Str. 14 D-67663 Kaiserslautern, Germany

<sup>3</sup> Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux CP5706 91057 Evry cedex, France

<sup>4</sup> IFREMER, Centre de Brest, DYNECO/Pelagos, BP 70, 29280 Plouzané, France

**Correspondence:** Raffaele Siano, raffaele.siano@ifremer.fr. Tel: 00 33 2 98 22 42 04  
Colombran de Vargas, vargas@sb-roscoff.fr, Tel: 00 33 2 98 29 25 28

---

### Abstract

Dinoflagellates (Alveolata) are one of the most complex and ecologically important groups of modern phytoplankton, making assessment of their global diversity and community structure difficult. Here we used massive sequencing of >170 million V9 18S ribosomal DNA sequences from 106 size-fractionated plankton communities collected across the world's surface oceans during the *Tara-Oceans* expedition (2009-2012) in order to assess patterns of pelagic dinoflagellate diversity and community structuring over global taxonomic and ecological scales. Our data suggest that dinoflagellate diversity has been largely underestimated, representing overall about half of total protistan rDNA OTU richness in the world's surface oceans. Dinoflagellate OTU richness and abundance display strong and differential patterns of taxonomic composition at the order level across size fractions. While the pico- and nano-planktonic communities are composed of an extreme diversity of metabarcodes assigned to Gymnodiniales or simply undetermined, most micro-dinoflagellate OTUs relate to the well-referenced Gonyaulacales and Peridiniales, and a lower abundance and diversity of essentially symbiotic Peridiniales is unveiled in the meso-plankton. Our analyses show the power of metabarcoding to quantitatively assess biodiversity from compartment of plankton ecosystems that still escape microscopy investigations, and illustrate its promises to understand and model the dynamics and function of pelagic biodiversity.

---

**Keywords:** Phytoplankton · biodiversity · dinoflagellates · metabarcoding · picoplankton · symbiotic dinoflagellates · Global Ocean · Tara Oceans

### Introduction

Dinoflagellates are a class of flagellated protists belonging to the eukaryotic super-group Alveolata, and form one of the largest lineages of modern phytoplankton (Guiry 2012). They are amongst the most fascinating group of protists due to their outstanding complexity in terms of life mode (autotrophic, mixotrophic, heterotrophic, parasitic, symbiotic) and cycle (e.g. cyst formation), ecology (pelagic and/or benthic of marine and freshwater habitats), and capacity for adaptation to environmental stressors (Schnepf & Elbrächter 1992; Jeong 1999; Gómez 2012b). They display significant variability in morphology, pigment composition, photosynthetic activity, along a large cell size spectrum ranging from  $<2\mu\text{m}$  to  $\pm 2\text{mm}$  in 'giant' species such as *Noctiluca*. In world pelagic ecosystems, they are part of the so-called nano-, pico-, micro-, and meso-planktonic size-fractions and are responsible for a core part of primary production (Gaines & Elbrächter 1987). Some species produce massive blooms well known for the red, green, or brownish water discolorations they induce, and they can also produce potent toxins noxious to fish and humans once accumulated in shellfish (Faust & Gullede 2002).

Genetic and morphological data allow the distinction of nine clear-cut major dinoflagellate orders: Gonyaulacales, Peridinales, Gymnodinales, Suessiales, Prorocentrales, Dinophysiales, Phytodinales, Noctilucales, and Pyrocystales. The phylogenetic position of Lophodinales and Thoracosphaerales is uncertain and they may actually belong to Peridinales. The orders Gonyaulacales and Peridinales are characterized by the presence of five latitudinal, one cingular and one sulcal series of cellulose-like thecal plates within the cortical alveoli of the cell covering (amphiesma). The Prorocentrales and the Dinophysiales instead share the division of the theca into lateral halves joined by a sagittal suture. The Gymnodinales, or unarmored dinoflagellates, is a clearly polyphyletic group distinguished by the absence of thecal plates within the cortical alveoli. The Phytodinales include poorly understood genera characterized by a shift from a coccoid cell or continuous-walled colonial stage to a vegetative stage. Similar life cycle shifts have been observed in genera of the order Suessiales that includes essentially symbiotic species. The Noctilucales is an early-diverging order containing aberrant dinoflagellates characterized by a highly motile ventral tentacle, which is missing in typical dinoflagellates and other alveolates (Not *et al.* 2012). The use of various species concepts had a strong influence on the number and biogeography of dinoflagellate species (Lundholm & Moestrup 2006). Gómez (Gómez 2012a; b; 2013) recognized 2,294 dinokaryotic dinoflagellates (including Noctilucales) belonging to 238 genera, while dinoflagellates *sensu lato* comprised 2,377 species belonging to 259 genera with 1,555 free-living marine morpho-species. In many cases classical morpho-species can be further split into cryptic or pseudo-cryptic species detected by DNA sequencing-based methods. One extreme example is the single marine genus *Symbiodinium* that displays degrees of genetic diversity corresponding to orders (Rowan & Powers 1992; Coffroth &

Santos 2005). Considering also that the smallest, benthic, and parasitic/symbiotic dinoflagellate diversity is still largely uncharacterized (Moreira & Lopez-Garcia 2002; Coffroth & Santos 2005; Worden 2006), the real number of dinoflagellates species may well be significantly bigger than the number of currently morphologically recognized taxa.

Dinoflagellate biogeography is similar to the spatial distribution of other aquatic protists within the same size range (Dolan 2005). The entire phytoplanktonic community of northern and southern temperate oceans is supposedly virtually identical (Taylor 2004). Likewise the same dinoflagellate morphospecies seem to occur in similar climatic zones in both northern and southern hemispheres (Taylor 2001). Dinoflagellates are often the dominant group of protists in tropical oceanic regions where they also show their greatest diversity (Taylor *et al.* 2007). Dinoflagellates communities differentiate clearly along inshore to offshore gradients. They reach highest abundances in estuaries and coastal marine waters, where nutrient supply from land sources and/or deep-water upwelling is high. Overall, dinoflagellates have clearly evolved multiple adaptive strategies rather than a main common one as observed in other phytoplankton groups. Basic abiotic factors, such as turbulence or nutrient availability, can select dominant life forms (Smayda & Reynolds 2003), however the dominant species or genera within life form communities are hardly predictable (Not *et al.* 2012).

Dinoflagellates' ecology and diversity are still largely based on light microscopy observations. Besides being partly subjective, light microscopy cannot discriminate species in groups that lack clear morphological features such as the pico- (<2 $\mu$ m) and nano- (<20 $\mu$ m) dinoflagellates. In the last decade the sequencing of environmental clone libraries of PCR-amplified ribosomal genes highlighted the presence of many novel dinoflagellates within pico-/nano-plankton assemblages (Moon-van der Staay *et al.* 2001). Nevertheless, the taxonomic and ecological significance of small-sized dinoflagellates is hidden behind the lack of quantitative data and taxonomic information (Siano *et al.* 2009; Siokou-Frangou *et al.* 2010; Siano *et al.* 2010).

The overarching goal of this study was to re-explore, using an objective and semi-quantitative metabarcoding (Taberlet *et al.* 2012) approach, the overall biodiversity and community structure of pelagic dinoflagellates on global taxonomical and biogeographical scales. Our molecular ecology protocol was thus constrained by the need to target all dinoflagellate lineages of the Earth surface oceans. In dinoflagellates, the classical DNA barcodes COI and *Cob* (Cytochrome Oxidase I and B) were tested on a wide taxonomic range, however none of these mitochondrial markers could be PCR-amplified from all dinoflagellate strains or could resolve common ambiguous genera at the species level (Stern *et al.* 2010). In particular COI was successful in identifying 70% of cultured species, but when applied to environmental samples it revealed a much higher diversity of uncharacterized species (Stern *et al.* 2010). Although *Cob*-based metabarcoding suffers from a significant lack of reference sequences from well-identified strains (Lin *et al.* 2009), this genetic marker has been used recently to

assess to the total richness of benthic dinoflagellates (Kohli *et al.* 2013). The use of relatively fast-evolving nuclear ribosomal DNA loci (LSU, ITS rDNA) (Orr *et al.* 2012) is generally preferred to the commonly used protistan 18S rDNA marker or to mitochondrial genes to assess dinoflagellate biodiversity (Edwardsen *et al.* 2003; Litaker *et al.* 2007). ITS-1 and ITS-2 rDNA barcodes were successful in identifying cultured species, and also reveal hidden diversity in strains from culture collections. However the use of these markers for metabarcoding is biased mainly by the difficulty of PCR-priming across dinoflagellate lineages, and to a lesser degree by the presence of paralogues, the potential for unidentifiable chimeras (Stern *et al.* 2012). Despite its relatively low variability, 18S rDNA remains advantageous for a comprehensive, first-order assessment of dinoflagellate biodiversity within a broader context of pan-eukaryotes diversity (Pawlowski *et al.* 2012), thanks to its ease to PCR target all dinoflagellate and most other protistan lineages, and its large representativeness in reference database (Not *et al.* 2009) which allows taxonomic annotation of environmental barcodes. The 18S rDNA contains eight hyper variable regions (V1 to V9, the V6 being present only in prokaryotes) that differ in length. The relevance of these hypervariable regions for dinoflagellate barcoding was evaluated using 77 cultured strains, confirming that intra-species genetic variation is low compared to interspecies divergence (Ki 2011). The V9 is considered as the third most informative 18S rDNA region to infer dinoflagellate diversity (Ki 2011). We used this barcode to assess overall dinoflagellate diversity at global scale, because its relatively short length of  $\pm 130$ bp allowed the use of high-throughput *HiSeq Illumina* sequencing of the 106 protistan communities analysed herein close to saturation. Our objective was not to characterize environmental dinoflagellate communities at species level, but rather to seek for general patterns of pelagic dinoflagellate biodiversity and community structure across global spatial and time scales, and in particular within relatively underexplored eukaryotic size classes.

## Experimental procedures

### Sampling

Samples analyzed in this study were collected in the frame of the 30 month-long international expedition *Tara Oceans* (Karsenti *et al.* 2011), which explored 7 oceanic regions and 12 Longhurst provinces (Longhurst 2007) allowing it to collect worldwide plankton samples from a complete range of planktonic ecosystems (coastal, tropical, oceanic, upwelling, etc.) and through a wide inter-annual and seasonal variability. Plankton from 11 size classes, from viruses to fish larvae, was collected together with a series of physico-chemical contextual data (Not *et al.*, submitted). For this study focusing on dinoflagellates, we used samples collected at 33 *Tara Oceans* stations over a 2-year period from September 15th, 2009 to May 13th, 2011. These samples concern 4 organismal size-fractions from the smallest to the largest protists (0.8-5 $\mu$ m; 5-20 $\mu$ m; 20-180 $\mu$ m; 180-2000 $\mu$ m) and cover 7

oceanic basins (the North Atlantic Ocean (1 station), the Mediterranean Sea (6), the Red Sea (4), the Indian Ocean (8), the South Atlantic Ocean (8), the Southern Ocean (2) and the South Pacific Ocean (4) (Table 1)). For convenience, we will call these 4 size-fractions respectively pico-, nano-, micro-, and meso- plankton, although rigorous size-boundaries are classically set at 2, 20, and 200  $\mu\text{m}$ . Samples were collected in subsurface (<5m) waters using an industrial peristaltic pump for the smallest size-fraction (0.8-5 $\mu\text{m}$ ), and a series of plankton nets with different mesh size for the 3 larger size fractions (5 to 2,000 $\mu\text{m}$ ). The pumped water was pre-filtered through a Gravity Plankton Sieving System (GPSS, a superposition of 3 nets with successive mesh sizes of 20 $\mu\text{m}$ , 5 $\mu\text{m}$ , and 5 $\mu\text{m}$  again (Not et al., submitted), in order to carefully separate the larger and often fragile plankton from the cells <5 $\mu\text{m}$ . This system was not expressly designed to assess dinoflagellate, but rather total protist diversity. The risk of any filtering strategy is that it may create a bias via the disruption of cells and their dispersion across smaller size fractions. This risk is particularly high for dinoflagellates, which contain large and fragile unarmored species. However GPSS was designed to concentrate large volumes of water as gently as possible in order to obtain sufficient biomass and nucleic acids to explore plankton diversity even in ultra-oligotrophic water masses. Two samples of 100L of recovered <5 $\mu\text{m}$  sieved seawater were then filtered through two parallel 142mm-diameter, 0.8 $\mu\text{m}$ -porosity polycarbonate membranes to recover the [0.8-5 $\mu\text{m}$ ] organismal size-fraction. The membranes (4 per sample) were immediately folded into 5mL barcoded cryotubes stored in liquid nitrogen on board. For the three >5 $\mu\text{m}$  size-fractions, 5, 20, and 180 $\mu\text{m}$ -meshed plankton nets were towed for ~15min in subsurface waters, and rinsed from the outside with filtered (0.1 $\mu\text{m}$ ) seawater when back on board. Plankton samples were then poured from the cod-ends through the appropriate sieves into 8L bottles. The volume was adjusted to 3L with 0.1 $\mu\text{m}$  filtered seawater. 0.5L of concentrated plankton was then filtered through two 47mm polycarbonate membranes and recovered together into a single 5mL barcoded cryotube which was immediately stored in liquid nitrogen. All cryotubes were stored at -80°C back in the laboratory until further molecular processing. Detailed protocols are available in Not et al. (submitted).

#### *Genomic DNA extraction, purification and sequencing*

Total nucleic acids (DNA + RNA) were extracted from 106 samples, using the Nucleospin® DNA II kit (Macherey-Nagel, Hoerdt, France). The V9 hyper-variable regions of the 18S ribosomal RNA gene were PCR-amplified with the eukaryotic primers 1389f 5'-TTGTACACACCGCCC-3' and 1510r 5'-CCTTCYGCAGGTTACCTAC-3' (Amaral-Zettler *et al.* 2009). Amplifications were conducted with the Phusion® High-Fidelity DNA Polymerase (Finnzymes). Briefly, the PCR mixture (final volume of 25 $\mu\text{L}$ ) contained 5ng of template with forward and reverse primers at a final concentration of 0.35 $\mu\text{M}$ , 3% of DMSO and 2X of GC buffer Phusion Master Mix (Finnzymes). Amplifications consisted in an initial denaturation step at 98°C for 30sec followed by 25 cycles of 10sec at 98°C, 30sec at 57°C, 30sec at 72°C, and a final elongation step at 72°C for 10min. Each sample was amplified in triplicates to obtain

sufficient PCR products, which were systematically checked on 1.5% agarose gels for positive bands of the expected length. Amplicons were then pooled, purified using the NucleoSpin® Extract II kit (Macherey-Nagel, Hoerdtt, France) and pyro-sequenced using a Genome Analyser Iix system (Illumina, San Diego, CA, USA) from both sides. Sequence reads were quality-checked and only those having exact forward and reverse primer match were considered. A total of 177,821,359 high-quality and complete V9 rDNA protistan sequences were recovered, including 4,402,312 distinct barcodes.

| Station Number | Latitude       | Longitude      | Sampling day | Season | Longhurst Province ID/name | Ecosystem  |
|----------------|----------------|----------------|--------------|--------|----------------------------|------------|
| stn004         | N 036° 34' 22" | W 006° 32' 19" | 15 sep. 2009 | Summer | NASE                       | Westerlies |
| stn007         | N 037° 02' 19" | E 001° 56' 59" | 23 sep. 2009 | Autumn | MEDI                       | Westerlies |
| stn009         | N 039° 04' 24" | E 005° 51' 35" | 28 sep. 2009 | Autumn | MEDI                       | Westerlies |
| stn018         | N 035° 45' 37" | E 014° 15' 07" | 2 nov. 2009  | Autumn | MEDI                       | Westerlies |
| stn023         | N 042° 09' 40" | E 017° 43' 52" | 18 nov. 2009 | Autumn | MEDI                       | Westerlies |
| stn025         | N 039° 24' 30" | E 019° 22' 50" | 23 nov. 2009 | Autumn | MEDI                       | Westerlies |
| stn030         | N 033° 55' 31" | E 032° 46' 26" | 15 dec. 2009 | Autumn | MEDI                       | Westerlies |
| stn031         | N 027° 08' 59" | E 034° 49' 05" | 09 jan. 2010 | Winter | REDS                       | Coastal    |
| stn032         | N 023° 22' 59" | E 037° 15' 08" | 11 jan. 2010 | Winter | REDS                       | Coastal    |
| stn033         | N 022° 03' 06" | E 038° 13' 00" | 13 jan. 2010 | Winter | REDS                       | Coastal    |
| stn034         | N 018° 23' 53" | E 039° 51' 42" | 20 jan. 2010 | Winter | REDS                       | Coastal    |
| stn036         | N 020° 49' 03" | E 063° 30' 43" | 12 mar. 2010 | Winter | ARAB                       | Coastal    |
| stn038         | N 019° 01' 32" | E 064° 37' 21" | 15 mar. 2010 | Winter | ARAB                       | Coastal    |
| stn041         | N 014° 33' 48" | E 070° 02' 23" | 30 mar. 2010 | Spring | ARAB                       | Coastal    |
| stn042         | N 005° 59' 33" | E 073° 54' 16" | 04 apr. 2010 | Spring | MONS                       | Trades     |
| stn045         | N 001° 47' 13" | E 071° 29' 40" | 12 apr. 2010 | Spring | MONS                       | Trades     |
| stn052         | S 016° 57' 24" | E 054° 00' 37" | 17 may. 2010 | Autumn | EAFR                       | Coastal    |
| stn064         | S 029° 32' 36" | E 037° 56' 08" | 08 jul. 2010 | Winter | EAFR                       | Coastal    |
| stn065         | S 035° 14' 48" | E 026° 18' 34" | 12 jul. 2010 | Winter | EAFR                       | Coastal    |
| stn066         | S 034° 53' 35" | E 018° 04' 22" | 15 jul. 2010 | Winter | EAFR                       | Coastal    |
| stn067         | S 032° 13' 53" | E 017° 42' 29" | 07 sep. 2010 | Winter | SATL                       | Trades     |
| stn068         | S 031° 01' 48" | E 004° 41' 16" | 14 sep. 2010 | Winter | SATL                       | Trades     |
| stn070         | S 020° 26' 12" | W 003° 11' 09" | 21 sep. 2010 | Winter | SATL                       | Trades     |
| stn072         | S 008° 42' 09" | W 017° 56' 23" | 5 oct. 2010  | Spring | SATL                       | Trades     |
| stn076         | S 021° 02' 44" | W 035° 22' 07" | 16 oct. 2010 | Spring | SATL                       | Trades     |
| stn078         | S 030° 08' 47" | W 043° 15' 13" | 4 nov. 2010  | Spring | SATL                       | Trades     |
| stn082         | S 047° 09' 55" | W 057° 53' 44" | 06 dec. 2010 | Spring | FKLD                       | Coastal    |
| stn084         | S 060° 13' 56" | W 060° 38' 42" | 03 jan. 2011 | Summer | ANTA                       | Polar      |
| stn085         | S 062° 14' 37" | W 049° 10' 57" | 06 jan. 2011 | Summer | APLR                       | Polar      |
| stn098         | S 025° 51' 03" | W 111° 46' 21" | 04 apr. 2011 | Autumn | SPSG                       | Westerlies |
| stn100         | S 012° 59' 40" | W 095° 59' 07" | 15 apr. 2011 | Autumn | SPSG                       | Westerlies |
| stn102         | S 005° 16' 10" | W 085° 13' 43" | 22 apr. 2011 | Autumn | PEQD                       | Trades     |
| stn109         | N 002° 04' 36" | W 084° 31' 13" | 13 may. 2011 | Spring | PEQD                       | Trades     |

**Table 1:** Geographical coordinates and Longhurst provinces (Longhurst, 2007) of sampling stations. NASE : North Atlantic Subtropical Gyral, MEDI : Mediterranean Sea, REDS : Red Sea, ARAB : NW Arabian Upwelling Province, MONS : Indian Monsoon Gyres Province, EAFR: East Africa Coastal Province, SATL : South Atlantic Gyral Province, FKLD : SW Atlantic Shelves Province : Antarctic Province, APLR: Austral Polar Province, SPSG: South Pacific Subtropical Gyre, PEQD : Pacific Equatorial Divergence.

*Database of reference dinoflagellate 18S rDNA and metabarcoding annotation*

All 18S rDNA dinoflagellate sequences available in NCBI (<http://www.ncbi.nlm.nih.gov/>) (Geer *et al.* 2009) on June 30th, 2012, were used to construct an *ad-hoc* reference database needed for the annotation of *Tara-Oceans* metabarcodes. Reference dinoflagellate sequences were aligned using MAFFT (Kato & Toh 2010; Kato & Standley 2013) and only those containing both the V4 and V9 rDNA fragments were kept for detailed phylogenetic analyses. Key DNA Tool (<http://keydnatools.com>) (Guillou *et al.* 2013) was used to detect the presence of chimeras in all remaining sequences, resulting in a total final number of 1,191 dinoflagellate reference sequences. The taxonomic status of each sequence was manually checked using phylogenetics (Neighbor Joining reconstructions as implemented in *SeaView*, (Gouy *et al.* 2010) and comparison to public taxonomic reference databases (AlgaeBase (Guiry & M 2012), and WoRMS (WoRMS Editorial Board 2012). We imposed a hierarchical nomenclature (order, family, genus and species) on all clades generated by our phylogeny. The polyphyletic clades typically found in dinoflagellate phylogenies were given the same name, but were kept distinct through the addition of successive, clade-specific numbers. All reference sequences used herein were integrated into the *PR2* - Protist Ribosomal Reference - database (Guillou *et al.* 2013), and our dinoflagellate-centered database (*DinR2*, Dinoflagellate ribosomal reference database) is available upon request. The  $\pm 4,4$  million V9 rDNA metabarcodes obtained from the world surface oceans were individually compared to each of the 28,547 *PR2* reference sequences (including the 1,191 *DinR2* sequences) using an exact global pairwise alignment algorithm (Needleman & Wunsch 1970), as implemented in GGSearch (Pearson & Lipman 1988). Each *Tara-Oceans* metabarcoding assigned to dinoflagellates received the taxonomic name of its nearest *DinR2* neighbour, or of the last common ancestor in case of a tie. Results were stored in a SQLite database.

*Assessment of molecular biodiversity*

The entire taxonomically assigned raw dataset was reduced by two subsequent cut-offs based first on the number of copies obtained per barcode, then on their percentage of identity (%ID) to reference sequences. In order to minimize the sequencing error and taxonomic assignment biases, only sequences present in 10 or more copies and with a %ID to any dinoflagellate reference sequence equal to or greater than 90% were retained in our final dataset. These criteria did not affect the overall distribution of raw dinoflagellate metabarcodes obtained without any cut-offs (Supp Fig. S1). We ended up with 83,860 different dinoflagellate metabarcodes or OTU-100% (we will call here OTU), which constituted our raw material for further biodiversity and community analyses. The choice to work with OTU was motivated by the fact that rDNA evolves particularly slowly in dinoflagellates and single substitutions in V9 rDNA can often distinguish different genera or families. Clustering at lower %ID would result in significantly lower taxonomic resolution, blurring the biodiversity patterns. To



characterize the relative importance of dinoflagellates over the whole protist community, all known protist classes (72), except the Dinophyceae, were pooled together. Only 14 Tara-Oceans stations with a complete set of data for all the 4 size fractions were used for this analysis. Further biodiversity analyses for each individual size fraction were carried out essentially at the taxonomic level of order. OTUs were assigned to 11 the eleven dinoflagellate orders (Dinophysiales, Gonyaulacales, Gymnodiniales, Lophodiniales, Noctilucales, Peridinales, Phytodiniales, Prorocentrales, Pyrocystales, Suessiales, Thoracosphaerales), plus 2 undefined groups (*Unknown* and *Undetermined*). The group '*Unknown*' includes metabarcodes that do not match any specific order but are phylogenetically clearly classified as dinoflagellate. The group '*Undetermined*' contains metabarcodes that are genetically strictly equally distant to two or more taxonomically unrelated reference sequence at the order level. For comparison, data were normalized over the total number of sequences obtained per sample. Dinoflagellate biodiversity was calculated using both OTU richness and abundance. Detailed  $\beta$ -diversity analyses of the same dataset will be presented elsewhere.

## Results

### *Dinoflagellates versus protists*

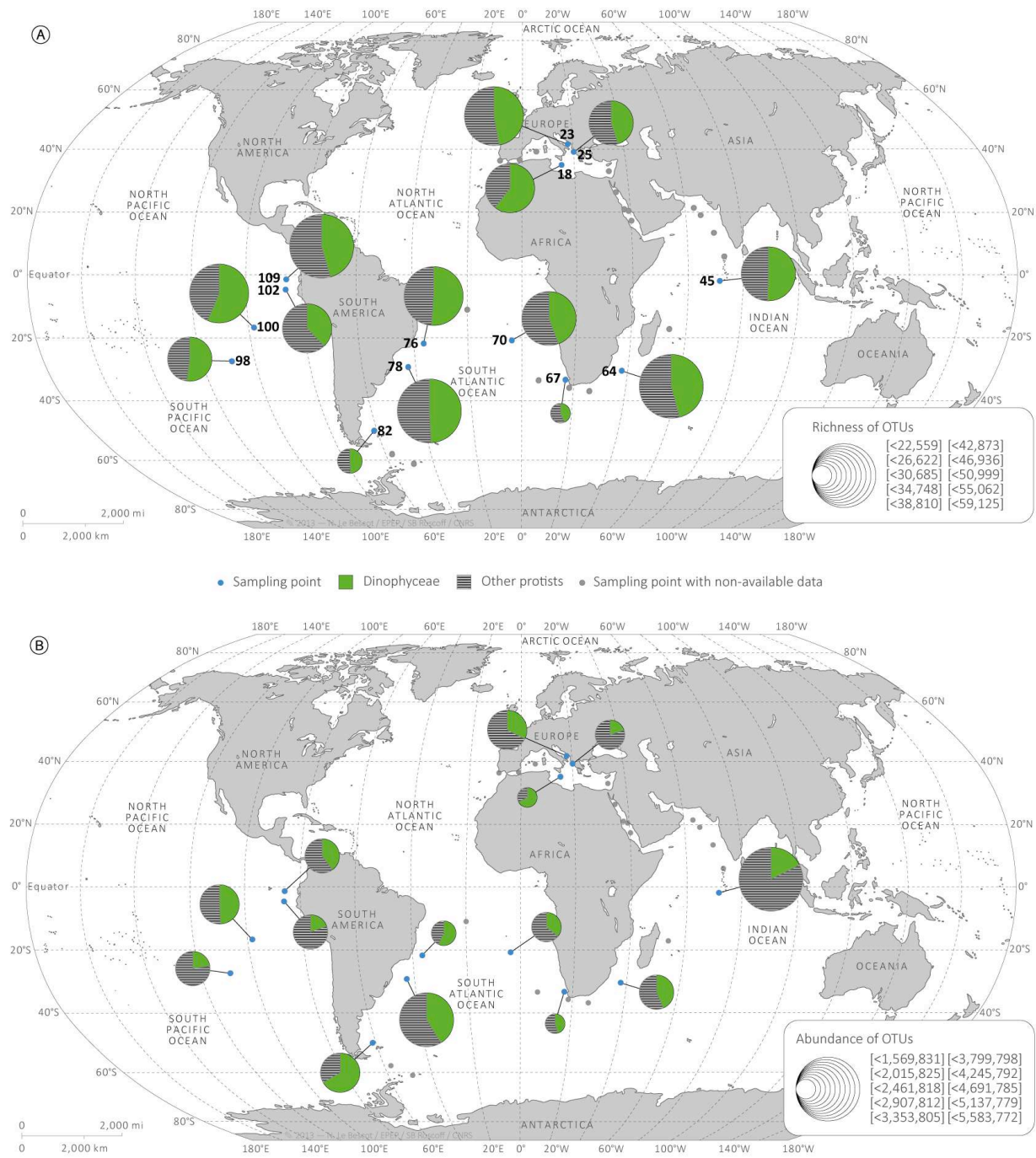
A total number of 83,860 V9 rDNA dinoflagellate OTUs (OTUs = dinoflagellate barcodes present in more than 10 copies with %ID  $\geq$  90% with dinoflagellate reference sequences) were retrieved from world surface oceans with an average sequence length of  $130 \pm 1\text{bp}$  (range: 79-143bp; mode = 130bp). Overall dinoflagellate OTU richness accounts for about 49% (range: 38-61%) of total protistan richness, dinoflagellate barcodes sometimes representing >50% of the number of total protistan barcodes (stations: 18, 45, 76, 98 and 100) (Fig. 1A). In terms of abundance, dinoflagellate OTUs account on average for 40% (range: 18-67%) of the total protistan OTU abundance, sometimes outnumbering total protistan barcode richness, even if they represent <25% of the total barcode abundance (stations: 25, 45, 98 and 102) (Fig. 1B). In terms of global biogeography, the relative richness of dinoflagellate OTUs is strikingly stable across the oceans, despite significant variation in overall dinoflagellate OTUs relative abundance. For instance in the Mediterranean Sea stations, OTU richness ranges from 45 to 61%, while OTU abundance ranges from 20 to 67%; in the Pacific Ocean OTU richness ranges from 38 to 57% while OTU abundance ranges from 20 to 49% (Fig. 1).

Dinoflagellate OTU abundance and richness varied over the whole protistan community among size fractions (Fig. 2). Generally higher absolute OTU richnesses and abundances are observed in the pico (0.8-5 $\mu\text{m}$ ) and then nano (5-20 $\mu\text{m}$ ) size fractions, with the exception of stations 64, 70, 78, 82, 100 and 109 where abundances are higher in the nanoplancton. Higher relative richness values are observed in the intermediate sizes fractions (respectively average: 64% (5-20 $\mu\text{m}$ ) and 59% (20-180 $\mu\text{m}$ ) than in the smallest (47% — 0.8-5 $\mu\text{m}$ ) and largest (24% — 180-2,000 $\mu\text{m}$ ) ones. Some exceptions, however, exist

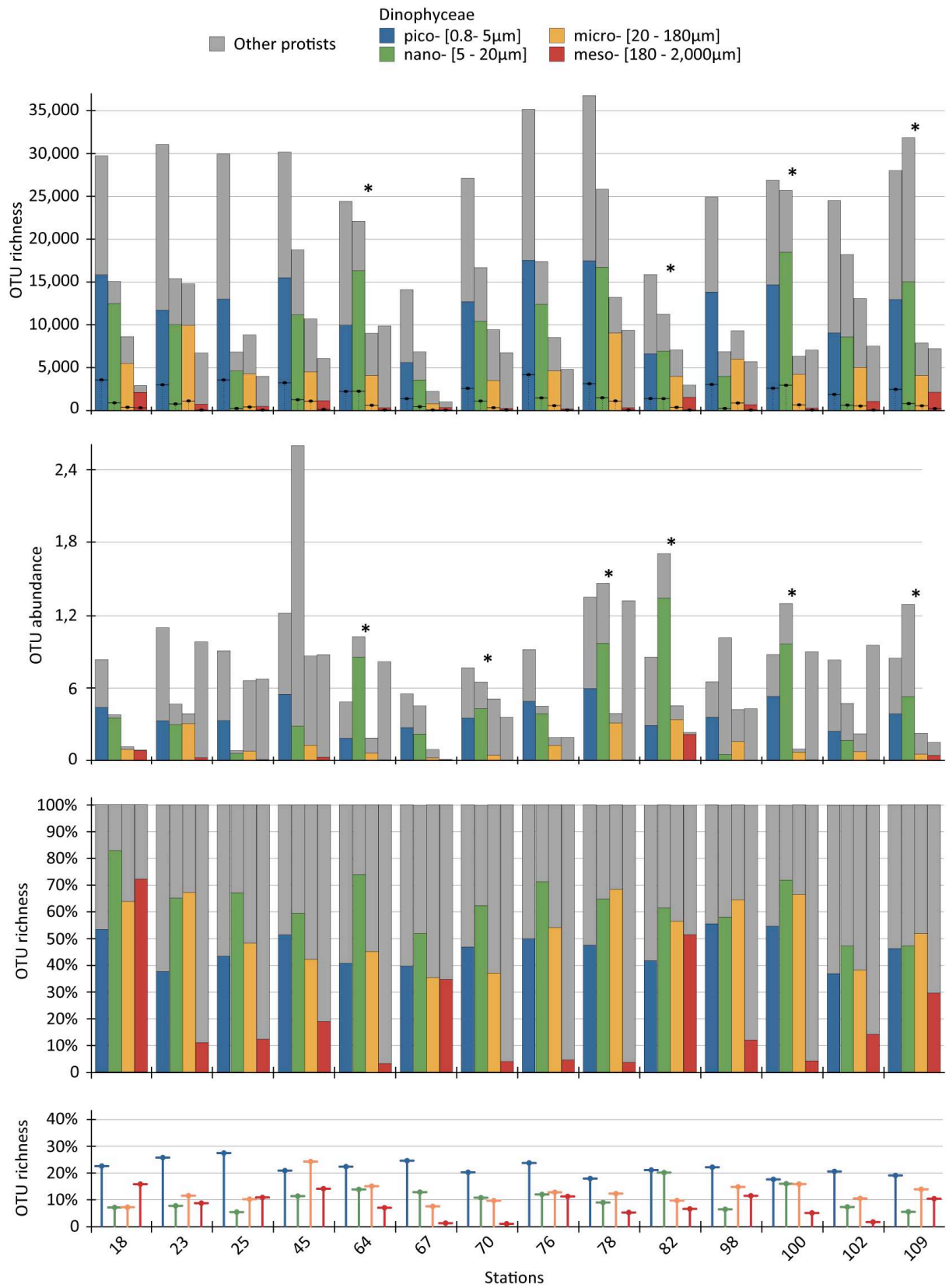
(stations 18, 45, 67, 70). Interestingly, the abundance of OTUs characteristic for a particular size-fraction (i.e. present exclusively in that fraction) is higher in the smallest size-fraction (average: 22%), although it is relatively stable amongst size fractions in terms of relative number. Station 45 is the only one characterized by a higher number of OTUs in the microplankton (25%).

#### *Dinoflagellate community characterization*

Overall *Undetermined* OTUs (sequences that cannot be unambiguously assigned to any dinoflagellate order on the basis of V9 rDNA annotation) span from 7% to 49% of the dinoflagellate communities, with an increasing ratio toward smaller size fractions. (Supp Fig. S2). On the other hand, the abundance and richness of dinoflagellate OTUs assigned to Orders vary significantly across size fractions. Some orders are represented at most stations and typically dominate in a given size fraction (Gymnodiniales in pico-/nano-plankton, Gonyaulacales in micro-plankton, and Peridinales in meso-plankton), while others (Lophodinales, Phytodinales, Pyrocystales, Thoracosphaerales) contribute much less to overall dinoflagellate communities.



**Figure 1:** OTUs dinoflagellates richness (A) and abundance (B) (green) over total protist community (grey with stripes). Pie-chart sizes are proportional to total OTUs<sub>100</sub> number analyzed per sample.



**Figure 2:** Abundances and richness of dinoflagellate OTUs over pico- (0.8-5µm), nano- (5-20µm), micro- (20-180µm) and meso- (180-2,000µm) plankton. The asterisk indicate exceptional stations where OTUs abundances are higher in nano-plankton. In the last picture dashed lines represent number of characteristic (present only in that fraction) barcodes per size classes.

### *Pico-dinoflagellates*

In the picoplankton (0.8-5 $\mu$ m size fraction) *Undetermined* OTUs represent on average 45% (range: 26-59%) of the total dinoflagellate OTUs abundance and 41% (range: 36-49%) of their richness (data available for 30 stations) (Fig. 3A). Assignment conflicts leading to undetermined dinoflagellates occur mainly amongst two super-clusters, the Gymnodiniales/Peridinales/Prorocentrales and Gymnodiniales/Peridinales groups (Suppl Fig. S3). Based on OTUs taxonomically assigned to orders, we observed that Gymnodiniales are more important than other orders in terms of both abundance (average 30%, range: 8-41%) and richness (average: 31%, range: 19-36%). Across stations, dominant Gymnodiniales OTUs are genetically related to the genera *Karlodinium* (average: 15%) and *Gyrodinium* (average: 11%). The large majority of the hundred most abundant dinoflagellate OTUs (accounting for 70% of total dinoflagellate barcodes abundance in the global dataset) are significantly more abundant in the pico-size fraction than in others, only 11 of them being top-100 OTUs are inversely more abundant in larger size fraction (Fig. S4).

At stations sampled around South Africa with a relatively short time lag (2 weeks between stations 65 and 66, and stations 67 and 68), Gymnodiniales OTU richness progressively increases from Indian Ocean (station 65) to southern Atlantic (stn66, stn68) waters, probably reflecting a community composition change.

In the same area, station 67 stands as an exception, with only 8% of Gymnodiniales OTUs, and with Prorocentrales and Peridinales representing respectively 21% and 14% of total dinoflagellate diversity. Peridinales OTUs are the second most important groups amongst pico-dinoflagellates in terms of both abundance (average 11%, range: 8-19%) and richness (average 14%, range: 12-20%). *Unknown* OTUs represent >15% of the total dinoflagellate metabarcode abundance at stations 65, 68 and 78. The richness of *unknown* OTUs is very similar among these stations, but Uncultured OTUs abundance almost double at stn078 (Fig. 3A).

### *Nano-dinoflagellates*

In the nano-planktonic samples (5-20 $\mu$ m size-fraction), the relative proportion of dinoflagellate *Undetermined* metabarcodes is still very high (average of 35%). On the other hand order-level assigned OTUs are relatively variable in terms of both richness and abundance, and a unique biodiversity pattern does not emerge across the 15 analyzed stations. A large number of Gymnodiniales OTUs is present in 10 stations over the 15 analyzed. At station 4, Gymnodiniales OTUs dominate in terms of richness (23%) but Gonyaulacales OTUs are more abundant (32%). Station 67 is characterized by a high Peridinales diversity (24%) and very high relative abundance (39%). Station 82 is largely dominated by

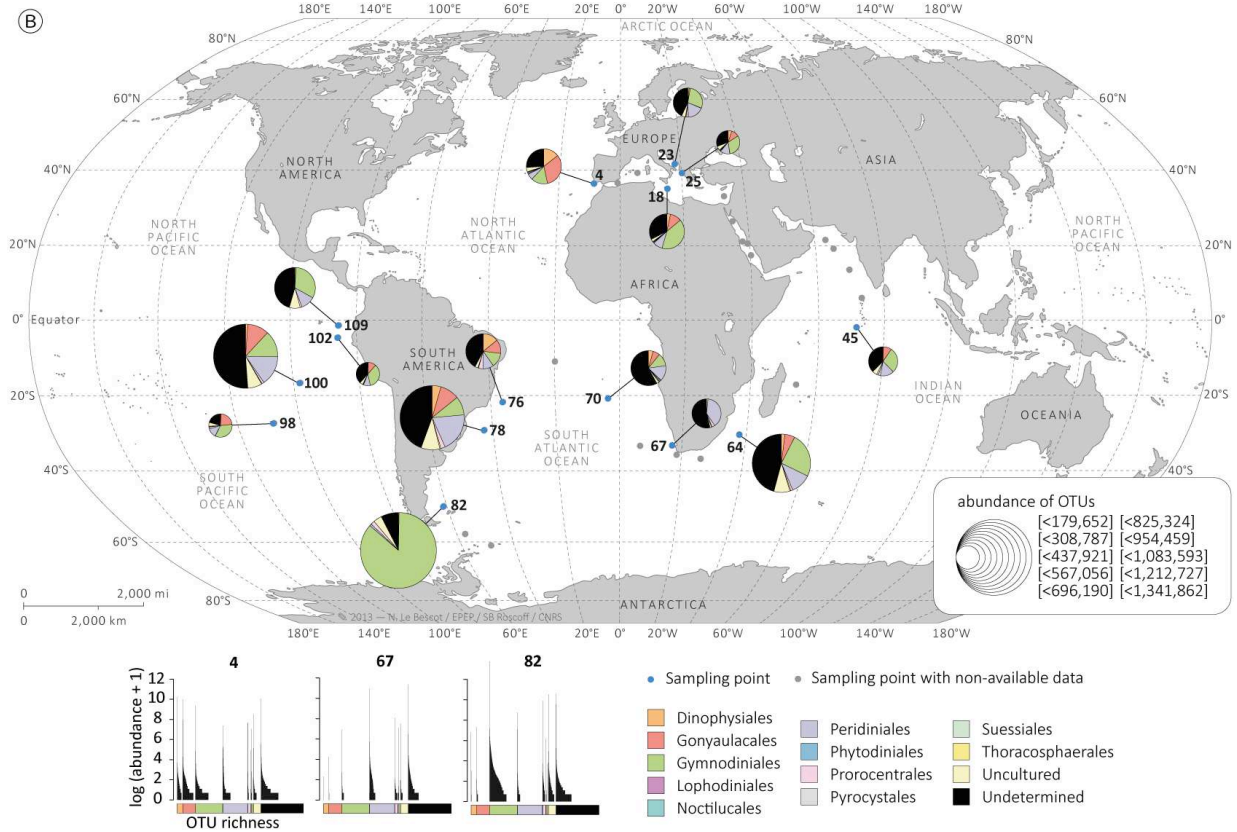
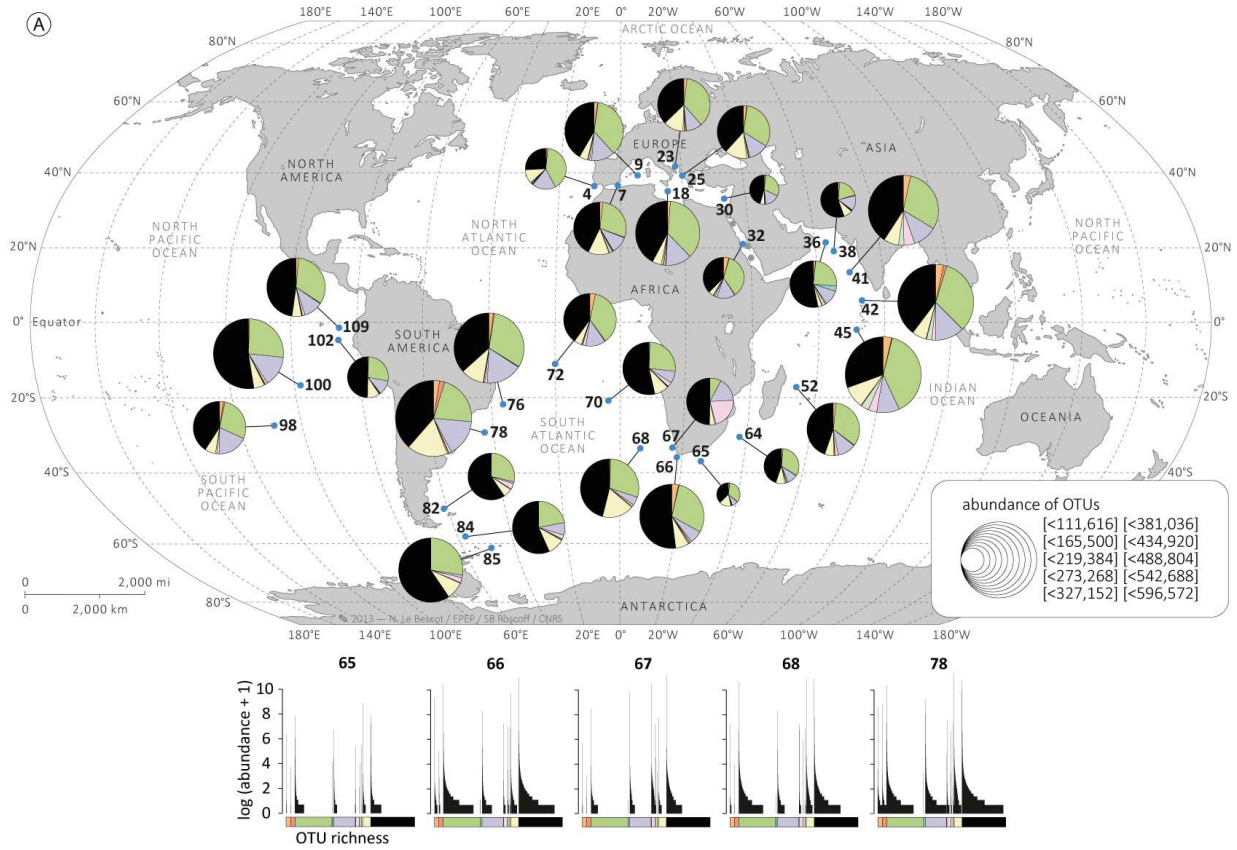
Gymnodiniales OTUs (86% and 37% of the OTUs abundance and richness, respectively), and it contains a relatively low number of *Undetermined* OTUs (Fig. 3B).

#### *Micro-dinoflagellates*

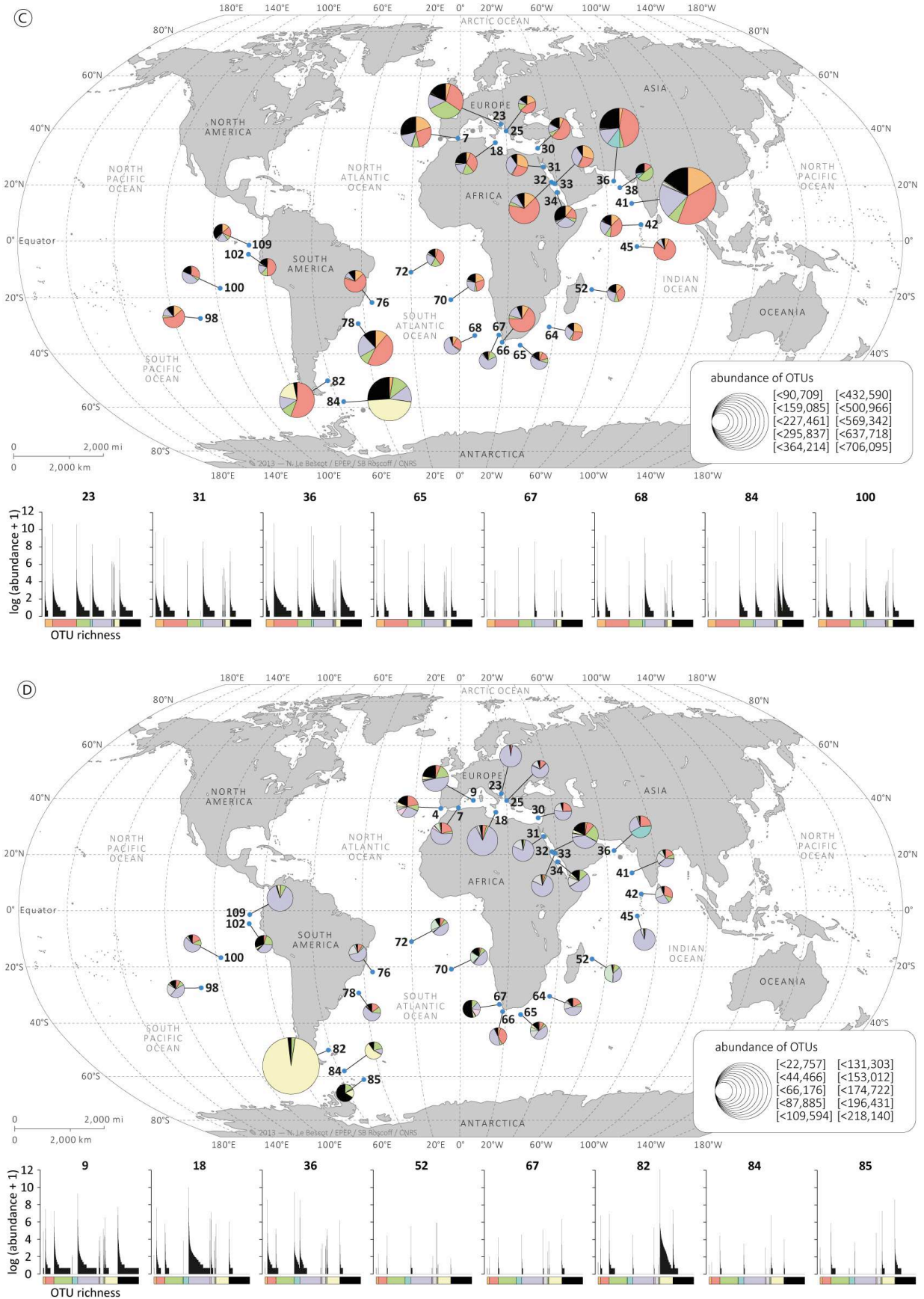
In the micro-plankton (20-180 $\mu$ m size-fraction), the relative proportion of *Undetermined* OTUs (average of 21%) is significantly less than in both smaller size fractions. Micro-dinoflagellate communities are dominated by Gonyaulacales OTU abundance and richness in respectively 23 and 19 out of the 30 analysed stations. Gonyaulacales represent on average 37% (range: 18-80%) and 26% (range: 22-48%) of total dinoflagellate OTU abundance and richness. Peridinales OTUs are the second most important dinoflagellate order in micro-plankton (abundance average: 23%, range: 21-69%, richness average: 25%, range: 21-48%). At stations 31, 33, and 64, Dinophysiales OTUs account for up to 25% and at stations 18, 23, 36, 82 and 84 Gymnodiniales are also important, contributing to up to 20%. Thus micro-dinoflagellate communities from the Atlantic Ocean and Mediterranean Sea are mainly composed of Gonyaulacales. In the Red Sea and Indian Ocean, Gonyaulacales, Peridinales and Dinophysiales tend to account for equal parts of total OTUs abundances, except for station 32 and 45. Stations around South Africa (65, 67, 68) are clearly characterized by Peridinales OTUs, except for station 66 which is instead over dominated by Gonyaulacales OTUs (67%). Note that in the Southern Ocean at station 84, *Unknown* OTUs represent  $\pm$ 47% of the total OTU abundance (Fig. 3C). At finer taxonomic resolution, >50% of the Gonyaulacales OTUs are assigned to the genus *Neoceratium* in 20 out of the 30 stations analysed. On average >75% of the Peridinales OTUs are related to *Protoperidinium*, and Dinophysiales OTUs are dominated (>60%) by *Ornithocercus* at stations 36, 45, 52 and 109 (Table 2).

#### *Dinoflagellate diversity in meso-plankton*

In the meso-planktonic size-fraction (180-2,000 $\mu$ m), Peridinales OTUs are the most abundant (average 48%, range: 15-95%) and diverse (average 37%, range from 17-66%) in, respectively, 26 and 27 stations out of the 31 analyzed. Again the Southern Ocean (stations 82 and 84) stands out, characterized by a majority of *Unknown* OTUs, while station 85 in the Southern Ocean and station 67 in coastal upwelling offshore of Cape Town are characterized by a higher number of *Undetermined* OTUs. At station 67 the taxonomically assigned part of dinoflagellate diversity is made up of relatively equal fractions of Peridinales (abundance 15% and richness 17%), Prorocentrales (abundance 14% and richness 10%) and Gymnodiniales (abundance 11% and richness 19%). Station 36 is characterized by a high abundance (42%) of Noctilucales OTUs, all strictly identical to the *Noctiluca scintillans* reference sequence (Fig. 3D).







**Figure 3:** Dinoflagellate order-based community structures assessed on OTU abundances (maps) and log +1 OTUs (selected stations, graphs) in the pico- (0.8-5µm) (A), nano- (5-20µm) (B), micro (20-180µm) (C) and meso- (180-2,000 µm) (D) plankton. In maps, pie-chart sizes are proportional the total sequence number analysed per sample.

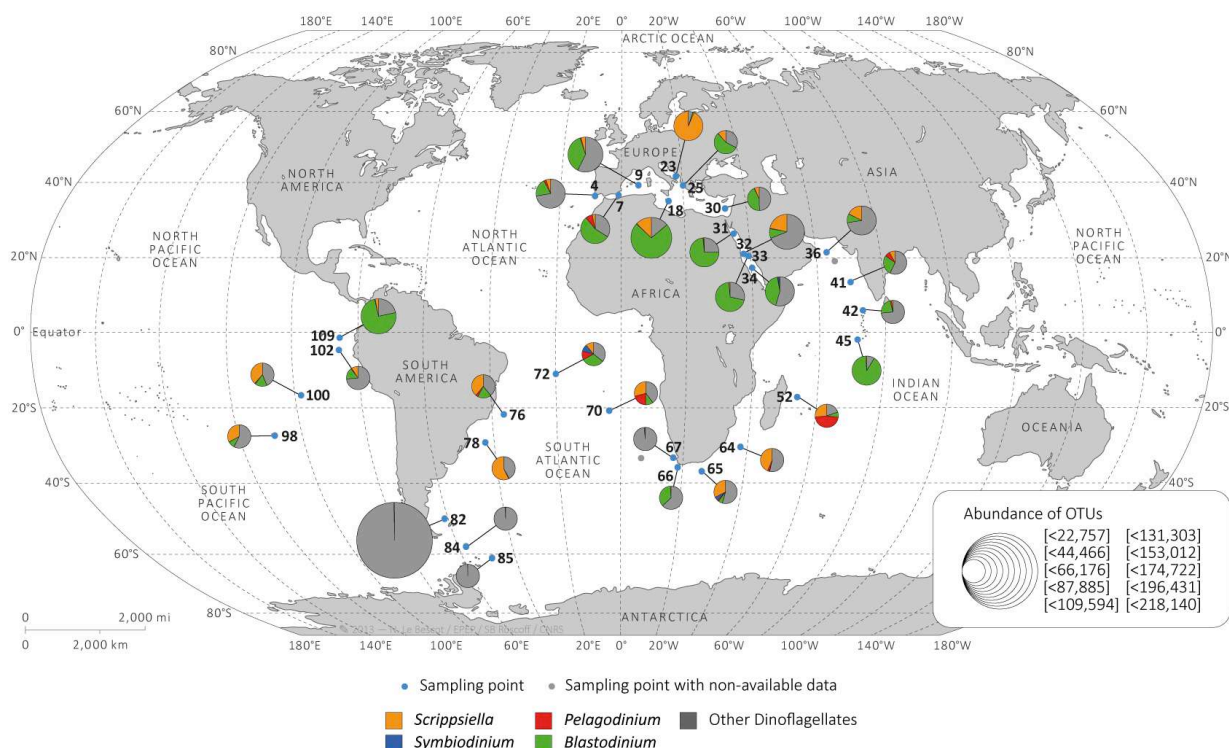


Dominant Peridinales OTUs are assigned to the genera *Blastodinium* and *Scrippsiella*, which include, respectively, parasitic and symbiotic species.

At station 52, Suessiales OTUs (46%) are all assigned (100% identity) to the *Pelagodinium béii* reference sequences, a well-known symbiotic taxon of planktonic foraminifera. Within the community of symbiotic dinoflagellate genera, *Blastodinium* OTUs represent >70% of total OTU abundance within the stations located above latitude 10°South, while *Scrippsiella* OTUs are more important below 10°S (Fig. 3, Table 2). In Southern Atlantic Ocean stations 82, 84 and 85, Peridinales contain essentially OTUs assigned to the genus *Proto-peridinium* (Table 2).

| Order          | Micro-dinoflagellates (20-180µm) (%) |           |                  |               | Size fraction 180-2,000µm (%) |              |                  |
|----------------|--------------------------------------|-----------|------------------|---------------|-------------------------------|--------------|------------------|
|                | Gonyaulacales                        |           | Peridinales      | Dinophysiales | Peridinales                   |              |                  |
|                | Neoceratium                          | Gonyaulax | Proto-peridinium | Ornithocercus | Blastodinium                  | Scrippsiella | Proto-peridinium |
| stn004         | NA                                   | NA        | NA               | NA            | 69                            | 18           | 2                |
| stn007         | 55                                   | 22        | 64               | 25            | 93                            | 5            | 1                |
| stn009         | NA                                   | NA        | NA               | NA            | 78                            | 8            | 0                |
| stn018         | 83                                   | 11        | 67               | 13            | 85                            | 15           | 0                |
| stn023         | 58                                   | 15        | 61               | 3             | 2                             | 98           | 0                |
| stn025         | 57                                   | 39        | 55               | 14            | 81                            | 16           | 0                |
| stn030         | 59                                   | 9         | 59               | 21            | 81                            | 13           | 5                |
| stn031         | 47                                   | 30        | 91               | 7             | 94                            | 1            | 0                |
| stn032         | 91                                   | 7         | 45               | 8             | 23                            | 55           | 3                |
| stn033         | 49                                   | 30        | 91               | 7             | 93                            | 1            | 3                |
| stn034         | 12                                   | 47        | 83               | 6             | 82                            | 1            | 1                |
| stn036         | 7                                    | 64        | 79               | 63            | 36                            | 63           | 1                |
| stn038         | 19                                   | 65        | 39               | 6             | NA                            | NA           | NA               |
| stn041         | 25                                   | 14        | 55               | 42            | 75                            | 21           | 2                |
| stn042         | 57                                   | 18        | 68               | 37            | 80                            | 7            | 4                |
| stn045         | 77                                   | 7         | 74               | 62            | 99                            | 1            | 0                |
| stn052         | 59                                   | 19        | 59               | 64            | 20                            | 74           | 1                |
| stn064         | 67                                   | 22        | 82               | 11            | 1                             | 95           | 3                |
| stn065         | 73                                   | 14        | 87               | 10            | 13                            | 72           | 10               |
| stn066         | 11                                   | 74        | 89               | 5             | 77                            | 2            | 17               |
| stn067         | 34                                   | 9         | 99               | 0             | 5                             | 0            | 16               |
| stn068         | 81                                   | 11        | 96               | 1             | NA                            | NA           | NA               |
| stn070         | 76                                   | 17        | 87               | 3             | 22                            | 63           | 6                |
| stn072         | 73                                   | 17        | 86               | 40            | 67                            | 24           | 6                |
| stn076         | 78                                   | 12        | 58               | 20            | 32                            | 66           | 1                |
| stn078         | 33                                   | 39        | 79               | 8             | 1                             | 96           | 1                |
| stn082         | 99                                   | 1         | 89               | 0             | 0                             | 0            | 82               |
| stn084         | 75                                   | 0         | 87               | 0             | 3                             | 0            | 67               |
| stn085         | NA                                   | NA        | NA               | NA            | 1                             | 1            | 55               |
| stn098         | 60                                   | 14        | 57               | 45            | 21                            | 75           | 1                |
| stn100         | 59                                   | 16        | 88               | 14            | 29                            | 67           | 2                |
| stn102         | 89                                   | 5         | 70               | 10            | 44                            | 26           | 10               |
| stn109         | 15                                   | 58        | 79               | 67            | 95                            | 4            | 0                |
| <b>Average</b> | 56±26                                | 24±20     | 74±16            | 20±21         | 49±36                         | 32±34        | 10±20            |

**Table 2:** Relative importance (%) of major dinoflagellate genus OTUs in fractions 20-180µm and 180-2,000µm at all sampled stations. NA: data not available.



**Figure 4:** Dinoflagellate genus-based community structures assessed on OTU abundances in fraction 180-2,000µm. Represented genera are characterized by mutualistic (*Scrippsiella*, *Pelagodinium*, *Symbiodinium*) or parasitic (*Blastodinium*) associations with other organisms. Pie-chart sizes are proportional to total sequence number analyzed per sample.

## Discussion

Initial barcoding studies on dinoflagellates aimed at assessing the best genetic marker to describe species and intraspecific diversity (Lin *et al.* 2009; Stern *et al.* 2010; 2012). Here we used broad metabarcoding to assess dinoflagellate overall biodiversity in environmental samples amongst protistan communities, testing this method as an alternative to classical phytoplankton cell counting analyses. We used a taxonomic-based approach to link classical knowledge of dinoflagellate biodiversity to next-generation sequencing-based analyses. Our ultimate goal was to identify general patterns of dinoflagellate diversity and community structure in the world ocean across protist size classes and over different oceanic regions, ecosystems and seasons. We showed the extreme importance of dinoflagellate rDNA diversity and abundance within total protist communities and we demonstrated that dinoflagellate communities display strikingly stable but distinct order-level biodiversity patterns across size-fractions, independently of space and time.

Overall, we analyzed ±84,000 distinct dinoflagellate OTUs obtained from 106 total protistan communities from 4 size-fractions and 33 world surface oceans stations to unveil global patterns of dinoflagellate biodiversity and community structure. Our objectives to compare dinoflagellate diversity to total protistan diversity and to target all dinoflagellates from a relatively large number of samples constrained the choice of molecular marker and sequencing technology. While Illumina GAllx brought the necessary sequencing power, the V9 rDNA barcode was best suited to target all eukaryotes

including dinoflagellates while being linked to the taxonomically most complete reference eukaryotic gene database available to date (Guillou *et al.* 2013). The drawback is that V9 rDNA is too short for accurate phylogenetics, and it has a particularly low-resolution power in dinoflagellates preventing biodiversity assessment at species and often-generic levels, and it occurs in multiple copies in single cells' genome. This last problem challenges the use of rDNA barcodes to assess protistan diversity and relative abundance. Indeed, a broad range of SSU rDNA copy numbers per cell has been estimated in different protist groups, such as diatoms (Galluzzi *et al.* 2004; Godhe *et al.* 2008), dinoflagellates (Godhe *et al.* 2008), ciliates (Gong *et al.* 2013), and a set of microalgal cultured strains (Zhu *et al.* 2005), varying from 1 in pico-planktonic taxa such as *Nanochloropsis salina* to thousands, tens of thousands or even hundreds of thousands in respectively dinoflagellates, diatoms, and ciliates. However, rDNA copy numbers are relatively stable within a given taxon and they have been shown to correlate positively to cell lengths and bio-volumes (Zhu *et al.* 2005; Godhe *et al.* 2008). For ecological studies bio-volumes appear to be more relevant than cell length, especially for diatoms, which reduce in size during their life cycle. Godhe (Godhe *et al.* 2008) demonstrated a linear correlation between SSU rDNA copy numbers assessed by qPCR and light measured bio-volumes for four diatom and nine dinoflagellate species in culture. Thus, the abundance data obtained in this study should obviously not be interpreted as number of individuals of a particular taxon, but as relative numbers of ribosomal genes which themselves are broadly correlated to cell biovolumes, a key ecological feature typically overlooked in classical dinoflagellate studies based on cell counting under the microscope.

### *Global dinoflagellate diversity and abundance*

Our study is the first global metabarcoding attempt to characterize pelagic dinoflagellates as a whole, across a wide variety of spatial and temporal scales and through different planktonic size fractions. Our V9 rDNA diversity data show that dinoflagellates represent about half (average 49%) of total protist richness in world surface oceans, and that proportion is strikingly stable across the >30 explored stations (range 38-61%). Even with the conservative cut-offs we applied to filter our data (>10 copies and >90%ID to reference sequence), our massive metabarcoding campaign unveiled ±84,000 distinct dinoflagellate OTUs, a much higher number than the total count of 1,555 morphological species of marine free-living dinoflagellates reported by Gómez (2012). This difference between rDNA and morphospecies diversities is obviously not trivial. However, the relative proportion of total dinoflagellate morpho-species (Gómez 2012b) to all protistan morphospecies (between 140,000 to 1.6 million (Adl *et al.* 2005), ranging from 1.4 to 16%, remains significantly lower than the 50% ratio we observed here in pelagic systems. Note that dinoflagellates are only one amongst the >100 deep-branching lineages of protists inhabiting planktonic ecosystems, some of which are known

to be species-rich. Thus our data indicate that the biodiversity of pelagic dinoflagellates has been largely underestimated. This knowledge gap will certainly increase if we further metabarcode the diversity of benthic and/or freshwater dinoflagellates, including also the rare species that were excluded from our dataset due to the strict cut-offs. This strong divergence between the molecular and morphological data needs further testing using independent measures of dinoflagellate diversity.

In contrast to the remarkable stability of dinoflagellate relative richness over space and time, the relative abundance of dinoflagellate OTUs display more variability, with values ranging from 18 to 67% of total protist OTU abundance (Fig. 2). This variability is likely due to specific environmental and seasonal parameters characterizing each sampled stations, with local selection for the growth or decline of a few dinoflagellate species. Overall, the global consistency of dinoflagellate diversity relative to its fluctuating abundance suggest strong community self-organizational rules at fundamental taxo-functional levels, with local selection of specific taxa on ecological time scales.

#### *Dinoflagellate community composition across size classes*

Significant biodiversity and ecological patterns emerge from our analyses of dinoflagellate metabarcode diversity and abundance across the pico-, nano-, micro-, and meso-planktonic size fractions. The fractionation of hundreds of liters of seawater through nets and sieves certainly biases to some extent the relative diversity of particularly fragile plankton across organismal size fractions. However, the ratio of dinoflagellate to all other protist sequences is relatively stable within each size fraction, and displays a remarkable pattern with higher relative richness in the intermediate (5-20 $\mu$ m; 20-180 $\mu$ m) size fractions as compared to the smallest (0.8-5 $\mu$ m) and largest (180-2,000 $\mu$ m) ones (Fig. 2). In addition, between 8 and 22% of barcodes are strictly specific to a single size fraction, providing further proof that our community fractionation was generally successful. In fact, the presence of identical barcodes across size fractions is expected due to natural cell debris, ontogenic and life-cycling stages involving different sizes, the presence of symbionts *sensu lato* in larger size-fractions, and the particularly slow rate of dinoflagellate rDNA evolution, meaning that a single barcode can represent several taxa of various size.

### *Picodinoflagellates (0.8-5µm)*

Previous community analyses of pico-plankton were essentially based on clone libraries of PCR-products (Lopez-Garcia *et al.* 2001; Moon-van der Staay *et al.* 2001; Diez *et al.* 2001; Cheung *et al.* 2008) and culturing methods (Massana *et al.* 2004; Balzano *et al.* 2012). Attempts to assess the ecological role of tiny dinoflagellate on the basis of quantitative estimation methods such as cells counts and carbon contribution are rare (e.g. (Siokou-Frangou *et al.* 2010; Zingone *et al.* 2010). Despite being not dominant, small (<10µm) dinoflagellates can contribute to up to 76% of total carbon production (Zingone *et al.* 2010). Pico-plankton genetic diversity studies mostly emphasize the importance of pico-green algae (Chlorophyta and Prasinophyta), estimating the relative dinoflagellate contribution on average to be 17% of total pico-eukaryotes (Massana *et al.* 2004). 454-pyrosequencing of total eukaryotic V4 18S rDNA PCR products from tropical waters further estimated that dinoflagellates represent about 19% of pico-eukaryotic communities (Cheung *et al.* 2010). Our results based on massive *Illumina* sequencing of larger volumes of extracted seawater from >30 world pico-planktonic communities suggest that dinoflagellates represent in fact a relatively stable and very important part of these communities, accounting on average for 47% and 43% of the total pico-eukaryotic community richness and abundance, respectively. Moreover, our data show that pico-dinoflagellates are not only key contributors of pico-planktonic communities, but are also significantly more diverse and *Unknown/Undetermined* than dinoflagellates from larger size-fractions (Fig. 2 and 3). The relatively higher proportion of dinoflagellate metabarcodes strictly specific to the pico-size fraction (18 to 26%) indicates that the revealed importance of pico-dinoflagellates does not come from debris of larger cell size-fractions. The potential accumulation in the smallest size fraction of debris from cells larger than 5µm is not a major bias, as we would have observed inversed abundance proportions (higher abundances in the biggest fractions than in the smallest) for most OTUs. Such inversion is found in only 11 out of the 100 most abundant OTUs (Supp Fig. S2).

The remarkable diversity of pico-dinoflagellates across the world's oceans is characterized by a large proportion of *Undetermined* OTUs (average richness and abundance of 41% and 45%, respectively) and Gymnodiniales OTUs (average richness and abundance of 31% and 30%, respectively). The assignments of most taxonomically determined OTUs to the Gymnodiniales should however be interpreted with caution, as this order is paraphyletic and inter-generic relationships are not clear (Orr *et al.* 2012). Peridinales and Prorocentrales metabarcodes are the next most important contributors to pico-dinoflagellates biodiversity. The fact that assignment conflicts leading to undetermined dinoflagellates OTUs concern mostly the Gymnodiniales/Peridinales/Prorocentrales and Gymnodiniales/Peridinales super-groups (Supp Fig. S1), reinforce our estimation of the most important orders within this size fraction. These conflicts also highlight the limit of our reference database that probably lacks many reference sequences, in particular from small feature-less cells,

within these three major orders. In fact very few pico-dinoflagellate species have been described to date. The pelagic genera *Karlodinium* and *Gymnodinium* (Gymnodiniales) (Siano *et al.* 2009; Thessen *et al.* 2012), *Azadinium* and *Heterocapsa* (Peridiniales) (Pomroy 1989; Percopo *et al.* 2013) and *Prorocentrum* (Prorocentrales) (Puigserver & Zingone 2002) and *Pelagodinium* (Suessiales) (Siano *et al.* 2010) include species of more or less 10 $\mu$ m in size. To the best of our knowledge a single dinoflagellate species, *Prorocentrum nux*, can be smaller than 5-3 $\mu$ m (Puigserver & Zingone 2002), thus truly meeting our definition of pico-plankton. The identification of *Karlodinium* OTUs amongst pico-plankton is somehow coherent with morphological information, whereas the detection of *Gyrodinium* OTUs is unexpected since described *Gyrodinium* species are generally bigger than 5 $\mu$ m. However, small naked species unidentifiable in light microscopy are typically observed in environmental samples (Siokou-Frangou *et al.* 2010), underlying the putative existence of a large biodiversity of pico-dinoflagellates overlooked in classical phytoplankton surveys. As for the great majority of pico-eukaryotic taxa from open oceans, this unveiled diversity is probably difficult to cultivate and thus not available in reference gene database. Similar pattern of extreme and novel pico-planktonic diversity in a classical group of marine phytoplankton believed to be well described morphologically has been shown in the haptophytes (Liu *et al.* 2009), and may well be common in pelagic protists.

#### *Nano-dinoflagellates (5-20 $\mu$ m)*

Metabarcoding data from nano-planktonic communities were available for fewer stations (15) than for other size fractions (30), due to the challenge in concentrating the relatively dilute biological material within the size-range 5 to 20 $\mu$ m from a large volume of seawater. Eukaryotic nano-plankton is arguably the least known component of pelagic ecosystems. However, this is the fraction where the relative richness of dinoflagellate compared to all-other protist metabarcodes is the highest (10/15 stations), suggesting a critical and underestimated importance of nano-dinoflagellates. This fraction is also characterized by a relative instability in the abundance of representatives from the different orders, when compared to the more stable proportions observed in pico- and micro-/meso-planktonic communities. Nano-dinoflagellate communities are characterized by a mixture of Gymnodiniales, Gonyaulacales, Peridiniales and Dinophysiales as well as by other orders whose relative abundance varies across spatial at temporal scales. The lack of data could partly explain the relatively non-homogenous diversity patterns within this size class, and a higher spatio-temporal sampling resolution is needed to understand the ecological and functional role of these relatively more complex communities.

### *Micro-dinoflagellates (20-180µm)*

The large majority of described, morphologically identifiable (10µm in size usually represents the lower limit of species detection by light microscopy) and genetically characterized dinoflagellates belong to this size class (Fensome *et al.* 1993; Gómez 2012b; 2013), making the micro-dinoflagellate communities relatively well known across the world's oceans in terms of both taxonomy and ecology. Our metabarcoding analysis confirms indeed that, except for some stations (7, 23, 36, 84, 9), *Undetermined* and *Unknown* dinoflagellates OTUs account for a very low percentage (<15%) of total micro-planktonic communities. Note that there is a general trend of decreasing relative importance of *Unknown/Undetermined* dinoflagellate diversity from the smallest to the largest size-fractions, matching the fact that species >20µm are relatively well described, and further supporting the remarkable and novel diversity detected in the pico- and nano-plankton. Amongst micro-dinoflagellates, the Gonyaulacales and Peridinales are the major traditional orders including genera that comprise a high number of species (e.g. *Neoceratium*, *Protoperidinium*, (Gómez 2012b; 2013). Our richness and abundance metabarcoding data confirmed that Gonyaulacales is the first, and Peridinales the second most important order within world surface oceans micro-dinoflagellate communities. The relative importance of one order over the other can vary across stations, likely depending on local environmental triggers selecting specific taxa.

### *Dinoflagellate diversity in meso-plankton (180-2,000µm)*

The largest described dinoflagellates (<2mm) are odd, fragile, and relatively rare species belonging to the order Noctilucales, whose phylogenetic position has been recently elucidated (Gómez *et al.* 2010). Apart from *Noctiluca scintillans* which is a relatively well known bloom-forming dinoflagellate occurring from temperate to sub-tropical coastal regions (Harrison *et al.* 2011), other Noctilucales (e.g. *Abedinium*, *Spatulodinium*, *Kofoidinium*, *Petalodinium*) are very poorly studied species with aberrant shapes that hampers their identification. The generally low relative abundances of Noctilucales OTUs in our metabarcoding dataset (average of <0,1%) confirms the rarity of members of this order in the world's oceans. One exception is *Tara-Oceans* station 36, where a high abundance of *Noctiluca scintillans* OTUs (42%) was observed, probably corresponding to a bloom of the species. However, this hypothesis must be confirmed by morphological analyses and cell counting of water samples.

Other dinoflagellates can be bigger than 180µm, such as members of both Gonyaulacales (i.e. *Neoceratium* spp.) and Pyrocystales (*Pyrocystis* spp.). Although some OTUs assignable to these orders were detectable in the meso-plankton (average 10% and 7% for the Gonyaulacales and Pyrocystales, respectively), Peridinales, metabarcodes were by far the most abundant (average 48%) across the

global ocean. Within Peridinales some *Proto-peridinium* spp can be larger than 180µm, however the highest number of OTUs was unexpectedly assigned to the genera *Blastodinium* and *Scrippsiella* (Table 2) which comprise only nano-/micro-dinoflagellate species. These genera also include well-known symbiotic parasitic and mutualistic species of copepods and jellyfishes, respectively (Banaszak *et al.* 1993; Skovgaard *et al.* 2007). In particular *Scrippsiella nutricula* has been recently emended as *Brandtodinium nutriculum* (the type species of the newly erected genus *Brandtodinium* within the Peridinales), the main symbiont of polycystine radiolarians (Probert *et al.* 2014). Relatively high abundances of metabarcodes assigned to *Pelagodinium* spp., the dinoflagellate genus symbiont of planktonic foraminifera (Siano *et al.* 2010), were also detected in meso-planktonic communities of stations 52, 70, and 72. Thus, the large majority of meso-planktonic dinoflagellate taxa unveiled by our metabarcoding survey correspond in fact to nano-/micro-dinoflagellate symbionts of mesoplanktonic heterotrophic protists or metazoans. The ecological role of this important community of symbiotic (parasitic, mutualistic, or commensal) dinoflagellates is still poorly understood. The symbiotic dinoflagellate metabarcodes from the meso-plankton are also found in smaller size fractions, corresponding likely to the endosymbiotic and free-living life cycle stages, respectively. Again, the putatively complex life cycles of symbiotic dinoflagellates are barely characterized. Single host organisms may harbor, as adult, several symbiotic and morphologically transformed dinoflagellate cells (Trench & Blank 1987; Caron *et al.* 2000; Siano *et al.* 2010), multi-nucleated trophont and physiologically differentiated cells (Skovgaard *et al.* 2012), or even morphologically and genetically distant species (Decelle *et al.* 2012). However the ontogeny of host-symbiont interactions (of holobionts) is usually ignored and much morpho-molecular work across size-fractions remains to be done to understand the ecological and biological features and significance of pelagic symbioses involving dinoflagellates. The biogeographic pattern of a higher prevalence of *Blastodinium* and *Scrippsiella* metabarcodes north and south of 10°North latitude, respectively (Fig. 4), may be linked to seasonal distribution of host species.



### Concluding remarks

Despite the lack of taxonomic resolution of V9 rDNA barcodes in dinoflagellates, our first deep metabarcoding study of this key phytoplanktonic lineage unveiled solid and globally coherent patterns of biodiversity at higher-taxonomic levels across protist size fractions, space, and time, supporting the value of next generation metabarcoding as an alternative tool for quantitative molecular ecology of marine protists. Our data suggest that the importance of pelagic dinoflagellates relative to all other protists has been largely underestimated and that the ribosomal diversity of the group represents about half the total protistan diversity in world surface oceans. Allowing exploration of very poorly known components of dinoflagellate biodiversity, our study unveiled rich and unique communities structures in both the largest and smallest size-fractions, with abundant Peridinales symbionts in the meso-plankton and a phenomenal diversity and abundance of Gymnodiniales assigned metabarcodes in the pico-/nano-plankton. Clearly, metabarcoding highlights the degree of our ignorance about the diversity of tiny and essentially symbiotic species escaping traditional microscopy detection.

Despite their extreme diversity, abundance, and functional complexity, dinoflagellate OTUs display strong and differential modes of community structure across organismal size fractions, suggesting that some orders have evolved their diversity within specific size ranges independently of extrinsic eco-evolutionary pressures. At finer systemic levels, the abundance of specific barcodes within a particular order at a given station is more chaotic across the sampled ecosystems, likely responding to local ecological triggers. These results illustrate the promise of using metabarcodes as quantitative data for future modeling of planktonic community dynamics and functions. However, V9 rDNA metabarcoding will clearly need to be complemented by other metabarcodes targeting finer diversity-levels and functional features (such as the diversity of group-specific chloroplasts) in order to fully assess the ecological dynamics of plankton communities. The development of morphogenetic protocols allowing linking of metabarcode diversity to phenotypic and cellular complexity will be key to anchor biological diversity to its biogeochemical context.

---

### Acknowledgements

This research was supported by a Région Bretagne PhD grant to NLB, the CNRS GDR3280 *Tara Oceans*, the ANR grant *POSEIDON*, and the "Investissement d'Avenir" program *OCEANOMICS* funded by the French government under the grant agreement ANR-11-BTBR-0008. We thank the coordinators and members of the *Tara Oceans* consortium for organizing samples and data collection, and for their continuous support and collaborative energy. We thank the commitment of the following people and sponsors who made this singular expedition possible: agnès b. and Etienne Bourgois, the Veolia Environment Foundation, Région Bretagne, World Courier, Illumina, Cap L'Orient, the EDF Foundation Diversiterre, the FRB, the Prince Albert II de Monaco Foundation, the captain, crew, and logistic team of Tara and Tara-Expeditions. *Tara Oceans* would not exist without the continuous support of the 23 participating institutes (see <http://oceans.taraexpeditions.org>). This article is contribution number 00XX of the *Tara Oceans* Expedition 2009–2012.

## References

- Adl SM, Simpson AGB, Farmer MA *et al.* (2005) The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *Journal of Eukaryotic Microbiology*, **52**, 399–451.
- Amaral-Zettler L, McCliment EA, Ducklow HW, Huse SM (2009) A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, **4**, e6372.
- Balzano S, Marie D, Gourvil P, Vaultot D (2012) Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *The ISME Journal*, **6**, 1480–1498.
- Banaszak AT, Prieto RI, Trench RK (1993) *Scrippsiella Velellae* Sp. Nov.(Peridinales) and *Gloeokinium Viscum* Sp. Nov.(Phytodiniales), Dinoflagellate Symbionts of Two Hydrozoans (Cnidaria). *Journal of Phycology*, **29**, 517–528.
- Caron DA, Dennett MR, Lonsdale DJ, Moran DM, Shalapyonok L (2000) Microzooplankton herbivory in the Ross Sea, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, **47**, 3249–3272.
- Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The ISME Journal*, **4**, 1053–1059.
- Cheung MK, Chu KH, Li CP, Kwan HS, Wong CK (2008) Genetic diversity of picoeukaryotes in a semi-enclosed harbour in the subtropical western Pacific Ocean. *Aquatic Microbial Ecology*, **53**, 295–305.
- Coffroth M-A, Santos SR (2005) Genetic Diversity of Symbiotic Dinoflagellates in the Genus *Symbiodinium*. *Protist*, **156**, 19–34.
- Decelle J, Siano R, Probert I, Poirier C, Not F (2012) Multiple microalgal partners in symbiosis with the acantharian *Acanthochiasma* sp. (Radiolaria). *Symbiosis*, **58**, 233–244.
- Diez B, Pedros-Alio C, Massana R (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Applied and Environmental Microbiology*, **67**, 2932–2941.
- Dolan JR (2005) Biogeography of aquatic microbes. *Aquatic Microbial Ecology*, **41**, 39–48.
- Edvardsen B, Jakobsen KS, Shalchian Tabrizi K *et al.* (2003) Genetic Variability and Molecular Phylogeny of *Dinophysis* Species (Dinophyceae) From Norwegian Waters Inferred From Single Cell Analyses of Rdna. *Journal of Phycology*, **39**, 395–408.
- Faust MA, Gulledge RA (2002) *Identifying harmful marine dinoflagellates*. Department of Botany, National Museum of Natural History.
- Fensome RA, Taylor FJR, Norris G *et al.* (1993) A classification of living and fossil dinoflagellates. *Micropaleontology Press*, **7**, 1–351.
- Gaines G, Elbrächter M (1987) Heterotrophic nutrition. In: *The Biology of Dinoflagellates*, pp. 224–268. Blackwell Scientific Publications London, Blackwell, Oxford, UK.
- Galluzzi L, Penna A, Bertozzini E *et al.* (2004) Development of a Real-Time PCR Assay for Rapid Detection and Quantification of *Alexandrium minutum* (a Dinoflagellate). *Applied and Environmental Microbiology*, **70**, 1199–1206.
- Geer LY, Marchler-Bauer A, Geer RC *et al.* (2009) The NCBI BioSystems database. *Nucleic acids research*, **38**, D492–D496.

- Godhe A, Asplund ME, Harnstrom K *et al.* (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Applied and Environmental Microbiology*, **74**, 7174–7182.
- Gong J, Dong J, Liu X, Massana R (2013) Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Annals of Anatomy*, **164**, 369–379.
- Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular biology and evolution*, **27**, 221–224.
- Gómez F (2012a) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *CICIMAR Océánides*, **27**, 65–140.
- Gómez F (2012b) A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity*, **10**, 267–275.
- Gómez F (2013) Problematic Biases in the Availability of Molecular Markers in Protists: The Example of the Dinoflagellates. *Acta Protozoologica*, **52**, 21–33.
- Gómez F, Moreira D, López-García P (2010) Molecular Phylogeny of Noctiluroid Dinoflagellates (Noctilucales, Dinophyceae). *Protist*, **161**, 466–478.
- Guillou L, Bachar D, Audic S *et al.* (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic acids research*, **41**, D597–D604.
- Guiry MD (2012) How many species of algae are there? *Journal of Phycology*, **48**, 1057–1063.
- Guiry MD, M GG (2012) Algaebase: Listing the World's Algae. *World-wide electronic publication, National University of Ireland, Galway. www.algaebase.org searched on August 2012.*
- Harrison PJ, Furuya K, Glibert PM *et al.* (2011) Geographical distribution of red and green Noctiluca scintillans. *Chinese Journal of Oceanology and Limnology*, **29**, 807–831.
- Jeong HJ (1999) The Ecological Roles of Heterotrophic Dinoflagellates in Marine Planktonic Community. *Journal of Eukaryotic Microbiology*, **46**, 390–396.
- Karsenti E, Sullivan MB, Acinas SG *et al.* (2011) A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biology*, **9**, e1001177.
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution*, **30**, 772–780.
- Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, **26**, 1899–1900.
- Ki J-S (2011) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. *Journal of Applied Phycology*, **24**, 1035–1043.
- Kohli GS, Neilan BA, Brown MV, Hoppenrath M, Murray SA (2013) Cob gene pyrosequencing enables characterization of benthic dinoflagellate diversity and biogeography. *Environmental Microbiology*, **16**, 467–485.
- Lin S, Zhang H, Hou Y, Zhuang Y, Miranda L (2009) High-Level Diversity of Dinoflagellates in the Natural Environment, Revealed by Assessment of Mitochondrial *cox1* and *cob* Genes for Dinoflagellate DNA Barcoding. *Applied and Environmental Microbiology*, **75**, 1279–1290.
- Litaker RW, Vandersea MW, Kibler SR *et al.* (2007) Recognizing dinoflagellate species using ITS rDNA sequences. *Journal of Phycology*, **43**, 344–355.

- Liu H, Aris-Brosou S, Probert I, de Vargas C (2009) A timeline of the environmental genetics of the haptophytes. *Molecular biology and evolution*, **27**, 161–176.
- Longhurst AR (2007) *Ecological geography of the sea*. New York.
- Lopez-Garcia P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**, 603–607.
- Lundholm N, Moestrup Ø (2006) The biogeography of harmful algae. In: *Ecology of harmful algae*, pp. 23–35. Springer.
- Massana R, Balagué V, Guillou L, Pedros-Alio C (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiology Ecology*, **50**, 231–243.
- Moon-van der Staay SY, De Wachter R, Vaultot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**, 607–610.
- Moreira D, Lopez-Garcia P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology*, **10**, 31–38.
- Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, **48**, 443–453.
- Not F, del Campo J, Balague V, de Vargas C, Massana R (2009) New Insights into the Diversity of Marine Picoeukaryotes. *PLoS ONE*, **4**, e7143.
- Not F, Siano R, Kooistra WHCF *et al.* (2012) Diversity and Ecology of Eukaryotic Marine Phytoplankton. In: *Genomic insights into the biology of algae*, pp. 1–53. Elsevier.
- Orr RJS, Murray SA, Stüken A, Rhodes L, Jakobsen KS (2012) When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE*, **7**, e50004.
- Pawlowski J, Simpson AGB, Audic S *et al.* (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biology*, **10**, e1001419.
- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 2444–2448.
- Percopo I, Siano R, Rossi R *et al.* (2013) A new potentially toxic Azadinium species (Dinophyceae) from the Mediterranean Sea, *A. dexteroporum* sp. nov. *Journal of Phycology*, **49**, 950–966.
- Pomroy AJ (1989) Scanning electron microscopy of *Heterocapsa minima* sp. nov. (Dinophyceae) and its seasonal distribution in the Celtic Sea. *British Phycological Journal*, **24**, 131–135.
- Probert I, Siano R, Poirier C *et al.* (2014) *Brandtodinium* Gen. Nov. and *B. Nutriculum* Comb. Nov. (Dinophyceae), a Dinoflagellate Commonly Found in Symbiosis with Polycystine (Radiolaria). *Journal of Phycology*, na.
- Puigserver M, Zingone A (2002) *Prorocentrum nux* sp. nov. (Dinophyceae), a small planktonic dinoflagellate from the Mediterranean Sea, and discussion of *P. nanum* and *P. pusillum*. *Phycologia*, **41**, 29–38.
- Rowan R, Powers DA (1992) Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 3639–3643.
- Schnepf E, Elbrächter M (1992) Nutritional strategies in dinoflagellates: A review with emphasis on cell biological aspects. *European Journal of Protistology*, **28**, 3–24.
- Siano R, Kooistra WHCF, Montresor M, Zingone A (2009) Unarmoured and Thin-Walled Dinoflagellates from The Gulf of Naples, With The Description of *Woloszynskia Cincta* sp. nov. (Dinophyceae, Suessiales). *Phycologia*, **48**, 44–65.

- Siano R, Montresor M, Probert I, Not F, de Vargas C (2010) Pelagodinium gen. nov. and P. béii comb. nov., a Dinoflagellate Symbiont of Planktonic Foraminifera. *Protist*, **161**, 385–399.
- Siokou-Frangou I, Christaki U, Mazzocchi MG *et al.* (2010) Plankton in the open Mediterranean Sea: a review. *Biogeosciences*, **7**, 1543–1586.
- Skovgaard A, Karpov SA, Guillou L (2012) The Parasitic Dinoflagellates Blastodinium spp. Inhabiting the Gut of Marine, Planktonic Copepods: Morphology, Ecology, and Unrecognized Species Diversity. *Frontiers in Microbiology*, **3**, 1–22.
- Skovgaard A, Massana R, Saiz E (2007) Parasitic Species of the Genus Blastodinium (Blastodiniophyceae) Are Peridinioid Dinoflagellates. *Journal of Phycology*, **43**, 553–560.
- Smayda TJ, Reynolds CS (2003) Strategies of marine dinoflagellate survival and some rules of assembly. *Journal of Sea Research*, **49**, 95–106.
- Stern RF, Andersen RA, Jameson I *et al.* (2012) Evaluating the Ribosomal Internal Transcribed Spacer (ITS) as a Candidate Dinoflagellate Barcode Marker. *PLoS ONE*, **7**, e42780.
- Stern RF, Horak A, Andrew RL *et al.* (2010) Environmental Barcoding Reveals Massive Dinoflagellate Diversity in Marine Environments. *PLoS ONE*, **5**, e13991.
- Taberlet P, Coissac E, Pompanon F (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.
- Taylor FJR (2001) Harmful Algal Blooms 2000 : Proceedings of the Ninth International Conference, Hobart, Australia, 7-11 February 2000-2001. *Harmful Algal Blooms*, 1–5.
- Taylor FJR (2004) Extraordinary Dinoflagellates: Past and Present. 61–66.
- Taylor FJR, Hoppenrath M, Saldarriaga JF (2007) Dinoflagellate diversity and distribution. *Biodiversity and Conservation*, **17**, 407–418.
- Thessen AE, Patterson DJ, Murray SA (2012) The Taxonomic Significance of Species That Have Only Been Observed Once: The Genus Gymnodinium (Dinoflagellata) as an Example. *PLoS ONE*, **7**, e44015.
- Trench RK, Blank RJ (1987) Symbiodinium Microadriaticum Freudenthal, S. Goreauii Sp. Nov., S. Kawagutii Sp. Nov. and S. Pilosum Sp. Nov.: Gymnodinioid Dinoflagellate Symbionts of Marine Invertebrates. *Journal of Phycology*, **23**, 469–481.
- Worden AZ (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecology*, **43**, 165–175.
- WoRMS Editorial Board (2012) WoRMS - World Register of Marine Species. Available from [www.marinespecies.org](http://www.marinespecies.org) at VLIZ. Accessed August 2012.
- Zhu F, Massana R, Not F, Marie D, Vaulot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, **52**, 79–92.
- Zingone A, Sarno D, Siano R, Marino D (2010) The importance and distinctiveness of small-sized phytoplankton in the Magellan Straits. *Polar Biology*, **34**, 1269–1284.

## Chapitre 5

**Évaluation du rôle éco-fonctionnel des communautés de  
dinoflagellés à partir de metabarcoding massif dans  
l'ensemble des océans**



## Sommaire

|   |            |
|---|------------|
| <b>Chapitre 5 : Évaluation du rôle éco-fonctionnel des communautés de dinoflagellés à partir de metabarcoding massif dans l'ensemble des océans</b> | <b>137</b> |
| <b>Eco-functional assessment of the world's oceans dinoflagellate communities based on massive metabarcoding</b>                                    | <b>139</b> |
| Introduction  | 140        |
| Matériel et méthode   | 144        |
| Échantillonnage   | 144        |
| Traitement statistique  | 144        |
| Résultat et discussion  | 145        |
| <b>Références</b>   | <b>150</b> |

In preparation for *Journal of Phycology*

## **Eco-functional assessment of the world's oceans dinoflagellate communities based on massive metabarcoding**

Noan Le Bescot<sup>1,2</sup>, Stéphane Audic<sup>1,2</sup>, Julie Poulain<sup>3</sup>, Samuel Chaffron<sup>4</sup>, Lionel Guidi<sup>5,2</sup>, Raffaele Siano<sup>6</sup>, Colomban de Vargas<sup>1,2</sup>

<sup>1</sup> CNRS, UMR 7144, Adaptation et Diversité en Milieu Marin, 29682 Roscoff, France

<sup>2</sup> UPMC Université Paris 06, UMR 7144, Station Biologique de Roscoff, 29682 Roscoff, France

<sup>3</sup> Génoscope, Centre National de Séquençage, 2 rue Gaston Crémieux CP5706 91057 Evry cedex, France

<sup>4</sup> VIB institute, KULeuven Campus Gasthuisberg Herestraat 49, 3000 Leuven, Belgium

<sup>5</sup> Observatoire Océanographique de Villefranche-sur-mer, 181 Chemin du Lazaret, 06234 Villefranche-sur-mer, France

<sup>6</sup> Ifremer, Centre de Brest, DYNECO/Pelagos BP70, 29280 Plouzané, France

---



### Introduction

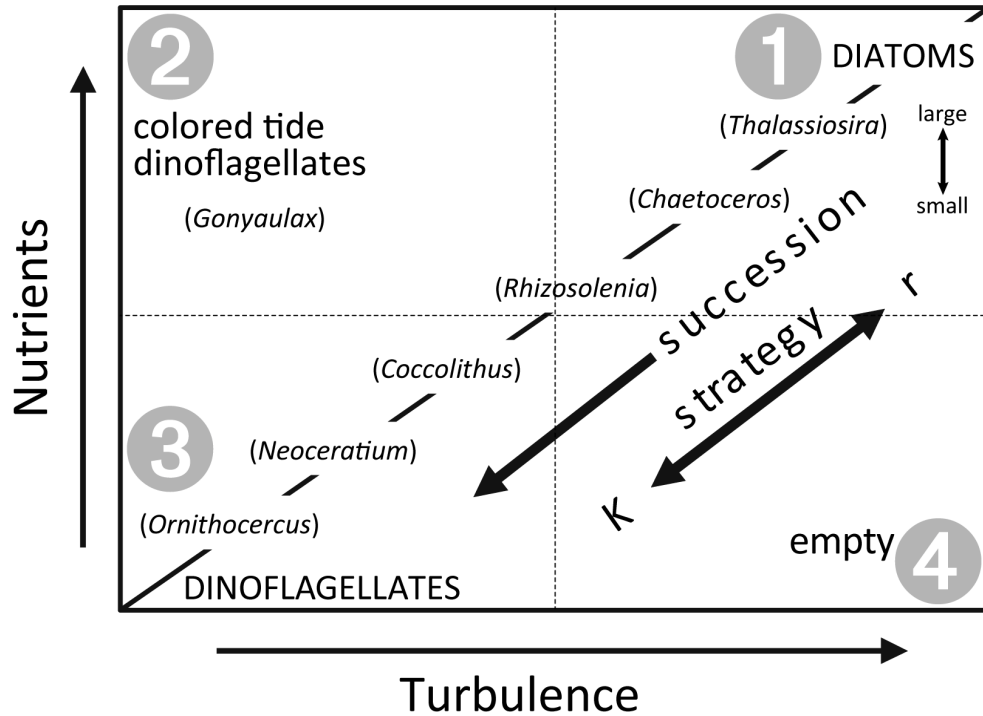
Les dinoflagellés forment le groupe dominant des protistes et sont des producteurs primaires majeurs dans les océans tropicaux et subtropicaux. Ils sont abondants dans le plancton de la fin du printemps au début de l'été dans les mers tempérées et subarctiques et présentent une très grande diversité (Taylor *et al.* 2007). Cette forte richesse spécifique ainsi que la diversité de leur morphologie et de leur physiologie témoignent en réalité de leur habilité à s'adapter et à coloniser une large gamme d'habitats. En plus de leur mode de vie pélagique, ils colonisent les mares résiduelles issues du processus des marées, les sédiments et les banquises et sont capables d'établir différentes associations endophytiques (Taylor *et al.* 2007).

On définit les capacités d'une espèce à se développer par son pouvoir d'adaptation à différentes caractéristiques du milieu aquatique. Ramon Margalef (Margalef 1978) fut le premier à proposer une vision globale de la distribution du plancton en fonction de la turbulence et de la concentration en nutriments, les considérant comme des facteurs clés qui contrôlent la morphologie et la physiologie.

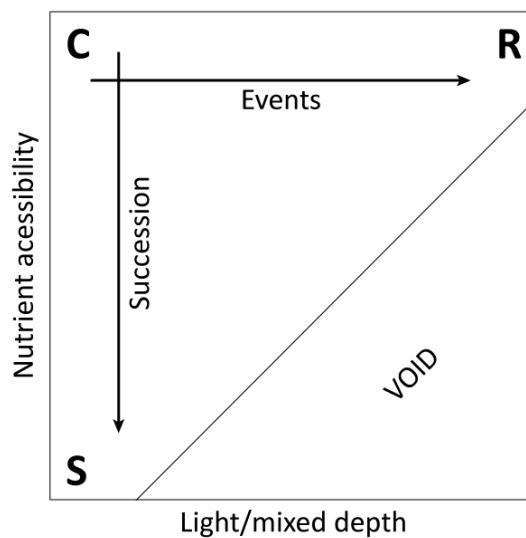
Dans son *Mandala* (Fig. 1) Margalef définit, arbitrairement, 4 domaines selon l'intensité de deux critères : la turbulence et la teneur en nutriments (Fig. 1). La répartition des communautés planctoniques (*Life-forms*) représentées par un genre au sein de ces différents domaines est principalement basée sur l'interaction entre les cycles de vie des espèces et les propriétés sélectives de l'environnement. Le domaine ❶ est défini par les efflorescences printanières des hautes latitudes, typiques des mers tempérées et boréales. Il se caractérise par une forte teneur en nutriments et de fortes turbulences. Le domaine ❷ est défini comme le domaine regroupant l'ensemble des dinoflagellés produisant des marées colorées. Il se caractérise par une forte concentration en nutriments et un faible taux de turbulence. Le domaine ❸ correspond aux eaux pélagiques chaudes et stratifiées de l'été. Il se caractérise par une faible turbulence et de faibles teneurs en nutriments. Enfin, le domaine ❹ se caractérise par une faible teneur en nutriments et de fortes turbulences. On y retrouve un grand nombre de cellules du pico- et nano-plancton (Cullen *et al.* 2002). Dans ce modèle, la succession des espèces est exprimée au travers d'un axe représentant un gradient de stratégies adaptatives avec pour extrémités les stratégies r et K.

En 2006, Reynolds (Reynolds 2006) a complexifié cette représentation en définissant trois stratégies adaptatives caractéristiques à l'écologie du plancton (et des dinoflagellés marins (Smayda & Reynolds 2001; 2003)) (Fig. 2). La stratégie C (Colon), dit du compétiteur ou du colonisateur opportuniste, est une stratégie basée sur un fort investissement physiologique dans la croissance et la reproduction. La stratégie S est basée sur la tolérance aux stress et la capacité des espèces à survivre à une raréfaction en nutriments. La stratégie R, dite rudérale, se caractérise par la capacité des espèces

à tolérer ou à s'acclimater à des altérations physiques fréquentes ou continues de leur environnement (turbulences, stress photique).

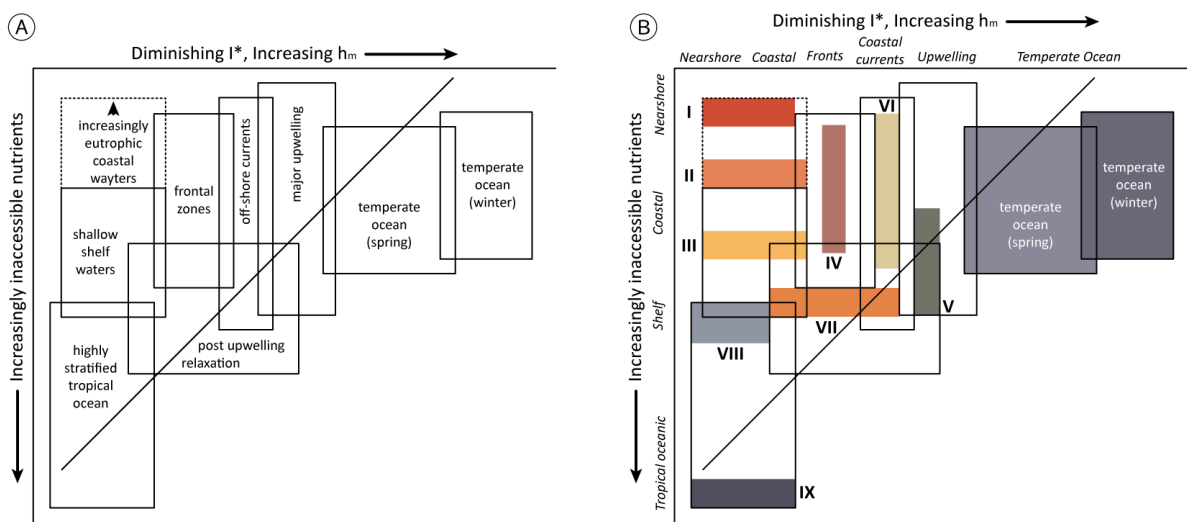


**Figure 1 :** Le *Mandala* de Margalef (adapté de (Margalef 1978; Cullen *et al.* 2002)) représente la succession des communautés planctonique dans un espace écologique défini par le degré de turbulence et la concentration en éléments nutritifs (0,02-3  $\mu\text{M}$  pour les phosphates, et 0,2-0,5  $\mu\text{M}$  pour les nitrates).



**Figure 2 :** Représentation de Reynolds illustrant le mécanisme de la sélection des espèces en fonction de 3 paramètres abiotiques: l'irradiance, la turbulence et la teneur en nutriments. (adaptée de (Smayda & Reynolds 2001)).

En 2001, Smayda et Reynolds en appliquant ces modèles aux dinoflagellés ont défini neuf types d'habitats pélagiques distincts propices aux dinoflagellés (Fig. 3) (Smayda & Reynolds 2001). Ces habitats sont ordonnés le long d'un gradient côte/océan-ouvert qui illustre la diminution des éléments nutritifs et des turbulences vers des zones euphotiques profondes. Chaque habitat est caractérisé par une communauté spécifique de dinoflagellés.



**Figure 3 :** (A) Matrice schématique des habitats pélagiques marins le long d'un gradient côte/océan-ouvert séparant les systèmes turbulents des systèmes fortement stratifiés ainsi que les systèmes déficients en nutriments. I\* réfère au niveau d'irradiance reçue par les cellules à l'intérieur de la colonne d'eau; H<sub>m</sub> représente la profondeur de la couche de mélange. (B) Association des types biologiques prédominants de dinoflagellés à la matrice turbulence-nutriments (A) le long d'un continuum côte/océan-ouvert caractérisant les habitats pélagiques. Type I = Gymnodiniales ; Type II = Peridinales et Prorocentrales ; Type III = *Neoceratium* ; Type IV = espèces des zones frontalières ; Type V = upwelling taxa relaxés ; Type VI = taxa entraînés par les courants côtiers ; Type VII = Dinophysis ; Type VIII = tropical ocean flora ; Type IX = tropical shade flora. (d'après (Smayda & Reynolds 2001)).

Les habitats, relativement peu profonds, riches en nutriments, et souvent mésosalins avec peu d'échanges avec l'eau du large, sont dominés par des Gymnodiniales de petite taille ou de taille intermédiaire (Type I). Dans les habitats où la concentration en nutriments est moindre mais demeure cependant élevée, les Peridinales et les Prorocentrales sont dominants (Type II). Les communautés de type III, correspondant au genre *Neoceratium*, se développent principalement à la fin de l'été et en automne dans les habitats côtiers des régions tempérées et boréales. En allant d'avantage vers le large, on retrouve une zone de « mélange » composée de trois types d'habitats : le premier où les espèces sont adaptées à des courants plus intenses associés à la zone frontale (Type IV) ; un second où elles sont tolérantes à la dérive au sein d'un courant côtier (Type VI) et un troisième, dit exposé (Type V), où on observe une accentuation des turbulences pendant la relaxation des courants d'upwelling. L'habitat de type VII correspond à des formes de vie dites « transitoires » correspondant à de

nombreuses espèces appartenant au genre *Dinophysis*. Pour finir, les types VIII et IX sont des types associés des habitats très oligotrophes et fortement stratifiés comme dans les régions tropicales.

En associant le concept de type biologique aux trois stratégies évolutives, Smayda et Reynolds (Smayda & Reynolds 2003) ont pu définir cinq règles permettant d'appréhender l'évolution de la communauté des dinoflagellés et le processus de sélection de ces différentes communautés :

- ⇒ **Règle n°1** : La composition et l'association des communautés de dinoflagellés dépendent de l'habitat et des conditions spécifiques de sélection des types biologiques. Cette règle définit que les dinoflagellés ne sont pas des organismes cosmopolites tolérants et libres, mais qu'ils possèdent des habitats spécifiques.
- ⇒ **Règle n°2** : Les types biologiques sont sélectionnés selon un modèle d'habitat caractérisé par des conditions physico-chimiques (la turbulence, l'irradiance et l'apport en nutriments). La sélection et la succession des types biologiques sont donc contrôlées par le rapport opposant les turbulences et l'irradiance à l'accessibilité aux nutriments.
- ⇒ **Règle n°3** : Les dinoflagellés suivent trois types de stratégies (C, S ou R) pour exploiter au mieux les changements de leur habitat. Ainsi, un habitat côtier de type I ou II, fortement enrichi en nutriments, qui se caractérise par une profondeur relativement faible et une forte stratification saisonnière de la masse d'eau, est dominé par des communautés suivant une stratégie de type C. Un habitat oligotrophe et fortement stratifié, de types VII, VIII ou XI, se caractérise par une communauté tolérante au stress (stratégie S).
- ⇒ **Règle n°4** : Le concept de « niche ouverte » est introduit et illustre le fait que dans un écosystème les niches écologiques de chaque espèce présente forment une niche globale dite « ouverte » dans laquelle chacune des espèces peut se développer mais dans un ordre d'apparition imprévisible.
- ⇒ **Règle n°5** : La sélection des espèces à l'intérieur d'un type biologique donné est stochastique. En d'autres termes, une espèce est sélectionnée si elle est au bon endroit au bon moment et dans des concentrations adaptées.

La description massive de la communauté des dinoflagellés par l'utilisation du metabarcoding offre aujourd'hui la possibilité de tester l'hypothèse d'association entre une communauté de dinoflagellés et un habitat donné à très grande échelle et avec une caractérisation plus précise de la biodiversité. Les échantillons de la mission Tara Océans, qui rassemblent des données génétiques dans leur contexte physico-chimique à l'échelle de tous les océans, sont particulièrement adaptés pour tester ces associations. L'objectif de ce chapitre est donc de rechercher une correspondance entre les communautés de dinoflagellés identifiées par metabarcoding et les caractéristiques physico-chimiques

des écosystèmes traversés pendant l'expédition Tara Océans. L'hypothèse de travail, qui confirmerait le paradigme de Smayda et Reynolds (Smayda & Reynolds 2001), consiste à vérifier l'existence d'une correspondance univoque entre certaines communautés de dinoflagellés et des conditions physico-chimiques données.

### Matériel et méthode

#### *Échantillonnage*

Le protocole d'échantillonnage et les méthodes d'acquisition et d'analyse des données génétiques (metabarcodes) utilisées dans ce chapitre sont les mêmes que ceux décrits dans le chapitre 4.

Les données physico-chimiques ont été recueillies par un grand nombre de capteurs fixés sur la Rosette : une Seabird 911 + une CTD avec deux paires de capteurs mesurant la conductivité et la température, la ISUS (*In Situ* spectrophotomètre UV) un capteur de nitrate, un capteur d'oxygène, un fluoromètre pour la chlorophylle et la matière organique dissoute colorée. Un vaste ensemble de paramètres a également été enregistré en routine à partir d'instruments déployés ponctuellement ou de dispositifs immergés en continu et munis de capteurs autonomes (Not et al., Soumis). Une unité SeaBird SBE45 MicroTSG utilisée en mode continu a fourni des données de température et de salinité pendant le transit entre les stations. Pour chaque station, des échantillons ont été prélevés à partir des bouteilles Niskin pour la chimie des carbonates, des nutriments et des pigments photosynthétiques. Les protocoles détaillés sont disponibles dans Not et al. (Soumis).

#### *Traitement statistique*

L'abondance génétique des séquences environnementales pour chaque fraction de tailles ([0,8-5µm], [5-20µm], [20-180µm] et [180-2000µm]) ont été comparées à l'aide des fonctions de la librairie « Vegan » du logiciel « R ».

Pour chaque fraction de taille, une matrice de dissimilarité a été calculée. Plusieurs méthodes de calculs ont été testés impliquant plusieurs types de normalisation : standardisation double de Wisconsin suivie du calcul de dissimilarité de Bray-Curtis (BC) (méthode 1), simple normalisation à un des effectifs de chaque échantillon suivi encore de BC (méthode 2), lissage des variations des abondances en prenant la racine carrée des abondances avant de normaliser à un, suivi de BC (méthode 3), et simple calcul de dissimilarité entre échantillons à partir de la mesure de Jaccard (présence, absence) (méthode 4).

Une analyse de similarité (ANOSIM) été réalisée à partir de ces matrices de dissimilarité, afin d'évaluer la méthode la plus à même de séparer les échantillons. La comparaison des différents résultats a permis de choisir la méthode 3, basée sur la racine carrée des abondances, pour la suite des analyses. Cette matrice de dissimilarité a alors été soumise à un positionnement multidimensionnel non linéaire (NMDS) représenté sur les figures 5 à 8, en utilisant la fonction « metaNMDS » du package « Vegan ».

Les échantillons ont ensuite été identifiés en fonction de leur bassin géographique ou d'une gamme de température, en utilisant soit l'ensemble des stations disponible, soit en enlevant les stations 82, 84, 85 et 67.

## Résultat et discussion

Les analyses NMDS montrent que dans toutes les fractions de tailles les communautés de dinoflagellés des stations situées à l'extrême Sud de l'Amérique du Sud (stations 82, 84 et 85 que nous désignerons ici comme stations Australes), ainsi que celle de la station 67 sont différentes du reste des stations analysées (Fig. 5). Les stations Australes présentent de manière générale des communautés relativement similaires. Ces communautés se développent dans des environnements aux températures extrêmement basses (de -1 à 7°C) contraignant probablement très fortement la physiologie des organismes (Fig. 7). La station 67 ne se distingue pas par une température particulièrement basse (15,2°C), mais est située dans une zone d'upwelling, l'upwelling du Benguela. Des conditions environnementales très particulières, typiques des zones d'upwelling, pourraient donc correspondre à une communauté très caractéristique. Des analyses détaillées des concentrations en nutriments, de la stratification de la colonne d'eau, de la salinité et d'autres paramètres biotique et abiotique seront nécessaires pour clarifier ce qui caractérise cet écosystème et la communauté des dinoflagellés qui lui est associée.

Pour observer de manière plus claire la répartition des autres stations échantillonnées, la même analyse a été menée en excluant de la matrice d'analyse les quatre stations Australes (Fig. 6). Dans les fractions du pico- et micro-plancton (0,8-5 et 20-180µm), pour lesquelles l'analyse NMDS est significative ( $p < 0.05$ ), on remarque une séparation des communautés de dinoflagellés pour certaines stations en fonction de leur appartenance à un bassin océanique. Cependant, cette analyse ne nous permet pas d'affirmer de manière définitive qu'il existe bien une nette distinction des communautés de pico- et micro-dinoflagellés au niveau géographique.

La communauté du nano-plancton (fraction 5-20µm) est difficilement interprétable, le nombre d'échantillons disponible y est certainement insuffisant. De même, malgré un nombre d'échantillons

plus important dans la fraction 180-2000 $\mu$ m (macro-plancton), la dispersion des échantillons observée dans l'analyse NMDS ne permet pas d'identifier des associations géographiques claires entre les différentes communautés de dinoflagellés. Cette absence d'associations claires pourrait par contre cette fois être expliquée par la constitution même de cette communauté majoritairement dominée par des genres symbiotiques (cf. description des communautés, chapitre 4). La fraction 180-2000 $\mu$ m contient en effet plusieurs hôtes comme les foraminifères, les radiolaires, ou encore les méduses qui ont des exigences écologiques très diverses. Dès lors, la relation entre les communautés de dinoflagellés et l'environnement devrait être considérée en fonction des facteurs influents sur les différents holobiot (hôte + symbiotes) et non exclusivement en considérant ceux influant potentiellement directement sur les communautés de dinoflagellés.

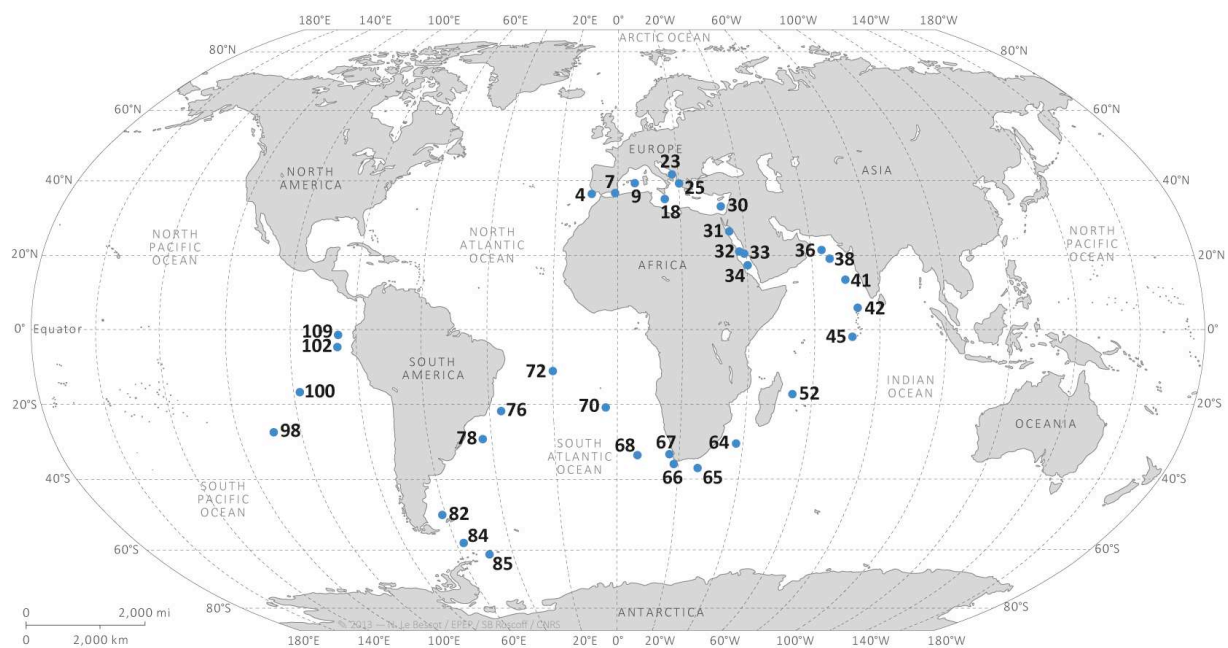


Figure 4 : Carte des stations.

| Station Number | Latitude        | Longitude       | Sampling day | Season | Temperature |
|----------------|-----------------|-----------------|--------------|--------|-------------|
| stn004         | N 036° 34' 22'' | W 006° 32' 19'' | 15 sep. 2009 | Summer | 15,1°C      |
| stn007         | N 037° 02' 19'' | E 001° 56' 59'' | 23 sep. 2009 | Autumn | 17,4°C      |
| stn009         | N 039° 04' 24'' | E 005° 51' 35'' | 28 sep. 2009 | Autumn | 16,2°C      |
| stn018         | N 035° 45' 37'' | E 014° 15' 07'' | 2 nov. 2009  | Autumn | 18,4°C      |
| stn023         | N 042° 09' 40'' | E 017° 43' 52'' | 18 nov. 2009 | Autumn | 15,2°C      |
| stn025         | N 039° 24' 30'' | E 019° 22' 50'' | 23 nov. 2009 | Autumn | 15,2°C      |
| stn030         | N 033° 55' 31'' | E 032° 46' 26'' | 15 dec. 2009 | Autumn | 19°C        |
| stn031         | N 027° 08' 59'' | E 034° 49' 05'' | 09 jan. 2010 | Winter | 26°C        |
| stn032         | N 023° 22' 59'' | E 037° 15' 08'' | 11 jan. 2010 | Winter | 26,1°C      |
| stn033         | N 022° 03' 06'' | E 038° 13' 00'' | 13 jan. 2010 | Winter | 26,5°C      |
| stn034         | N 018° 23' 53'' | E 039° 51' 42'' | 20 jan. 2010 | Winter | 27,6°C      |
| stn036         | N 020° 49' 03'' | E 063° 30' 43'' | 12 mar. 2010 | Winter | 25,4°C      |
| stn038         | N 019° 01' 32'' | E 064° 37' 21'' | 15 mar. 2010 | Winter | 25,5°C      |
| stn041         | N 014° 33' 48'' | E 070° 02' 23'' | 30 mar. 2010 | Spring | 27,2°C      |
| stn042         | N 005° 59' 33'' | E 073° 54' 16'' | 04 apr. 2010 | Spring | 27,8°C      |
| stn045         | N 001° 47' 13'' | E 071° 29' 40'' | 12 apr. 2010 | Spring | 14,5°C      |
| stn052         | S 016° 57' 24'' | E 054° 00' 37'' | 17 may. 2010 | Autumn | 24,9°C      |
| stn064         | S 029° 32' 36'' | E 037° 56' 08'' | 08 jul. 2010 | Winter | 22,3°C      |
| stn065         | S 035° 14' 48'' | E 026° 18' 34'' | 12 jul. 2010 | Winter | 21,8°C      |
| stn066         | S 034° 53' 35'' | E 018° 04' 22'' | 15 jul. 2010 | Winter | 15°C        |
| stn067         | S 032° 13' 53'' | E 017° 42' 29'' | 07 sep. 2010 | Winter | 15,2°C      |
| stn068         | S 031° 01' 48'' | E 004° 41' 16'' | 14 sep. 2010 | Winter | 16,8°C      |
| stn070         | S 020° 26' 12'' | W 003° 11' 09'' | 21 sep. 2010 | Winter | 17,1°C      |
| stn072         | S 008° 42' 09'' | W 017° 56' 23'' | 5 oct. 2010  | Spring | 24,1°C      |
| stn076         | S 021° 02' 44'' | W 035° 22' 07'' | 16 oct. 2010 | Spring | 21,6°C      |
| stn078         | S 030° 08' 47'' | W 043° 15' 13'' | 4 nov. 2010  | Spring | 19,3°C      |
| stn082         | S 047° 09' 55'' | W 057° 53' 44'' | 06 dec. 2010 | Spring | 7°C         |
| stn084         | S 060° 13' 56'' | W 060° 38' 42'' | 03 jan. 2011 | Summer | 2°C         |
| stn085         | S 062° 14' 37'' | W 049° 10' 57'' | 06 jan. 2011 | Summer | -0,8°C      |
| stn098         | S 025° 51' 03'' | W 111° 46' 21'' | 04 apr. 2011 | Autumn | 20,2°C      |
| stn100         | S 012° 59' 40'' | W 095° 59' 07'' | 15 apr. 2011 | Autumn | 20,6°C      |
| stn102         | S 005° 16' 10'' | W 085° 13' 43'' | 22 apr. 2011 | Autumn | 19,6°C      |
| stn109         | N 002° 04' 36'' | W 084° 31' 13'' | 13 may. 2011 | Spring | 26,5°C      |

**Tableau 1** : Liste des températures et saisons correspondantes pour chaque station échantillonnée.

La température ne semble pas être un élément significatif permettant de justifier la répartition biogéographique des communautés de dinoflagellés dans les fractions 0,8-5µm et 20-180µm (Fig. 8). L'analyse NMDS est par contre significative ( $p < 0,05$ ) pour la fraction 20-180µm. Il semblerait donc que dans le micro-plancton les communautés de dinoflagellés issues d'environnements aux températures très différentes puissent être similaires. C'est le cas pour les stations 23 et 72 qui présentent des communautés très similaires dans des bassins géographiques très distants (respectivement la mer Méditerranée et le milieu de l'océan Atlantique) échantillonnées à des saisons différentes (automne et printemps) et présentant des températures d'eaux superficielles très distantes 15,2 et 24,1°C (Tab. 1). Le même type d'observation est réalisé pour les stations 32, 78 et 98 respectivement situées en mer Rouge, en océan Atlantique et au milieu de l'océan Pacifique avec des températures respectives de 26,1, 19,3 et 20,2°C.



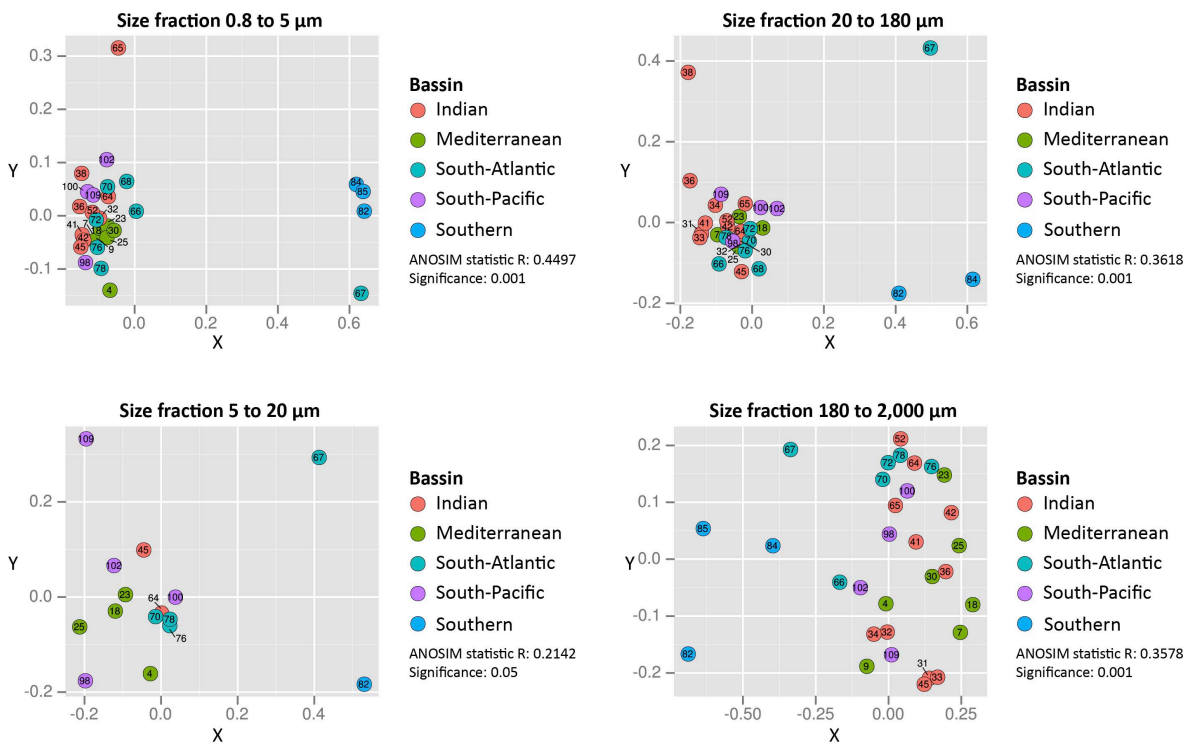


Figure 5 : NMDS illustrant la similarité des stations par fraction de tailles et leurs bassins océaniques.

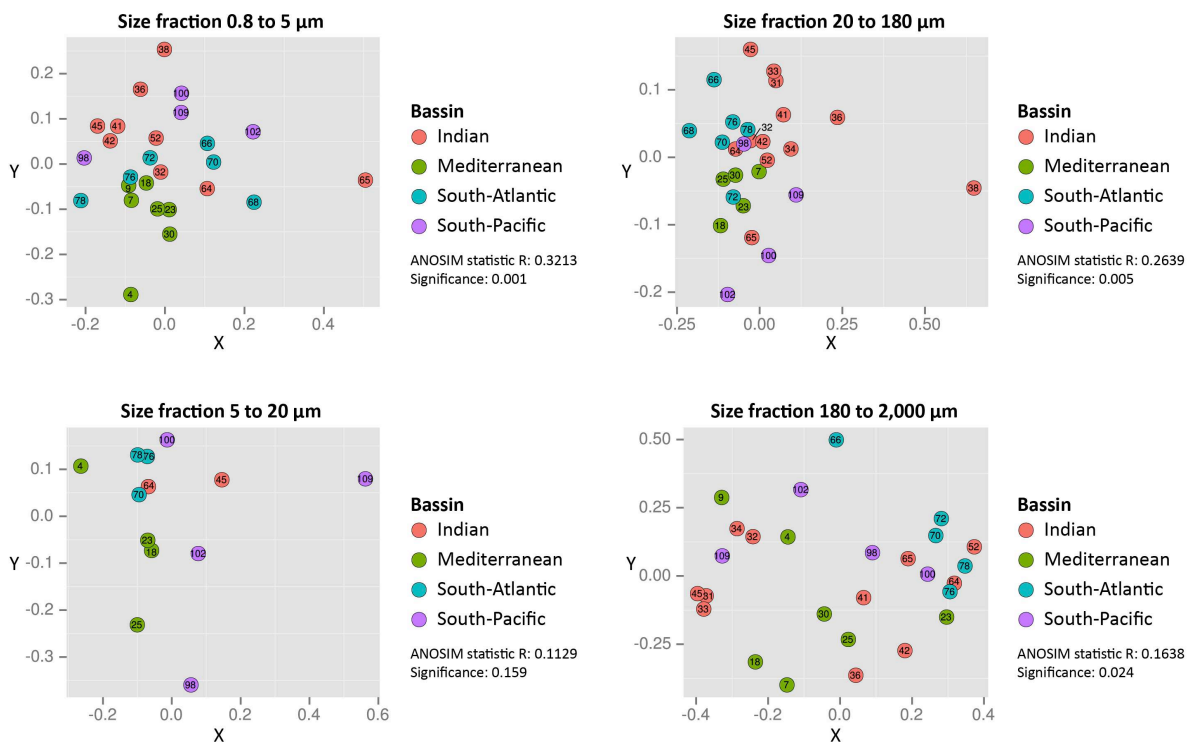


Figure 6 : NMDS illustrant la similarité des stations par fraction de tailles et leurs bassins océaniques. Les stations les plus dissimilaires ont été ici retirées.

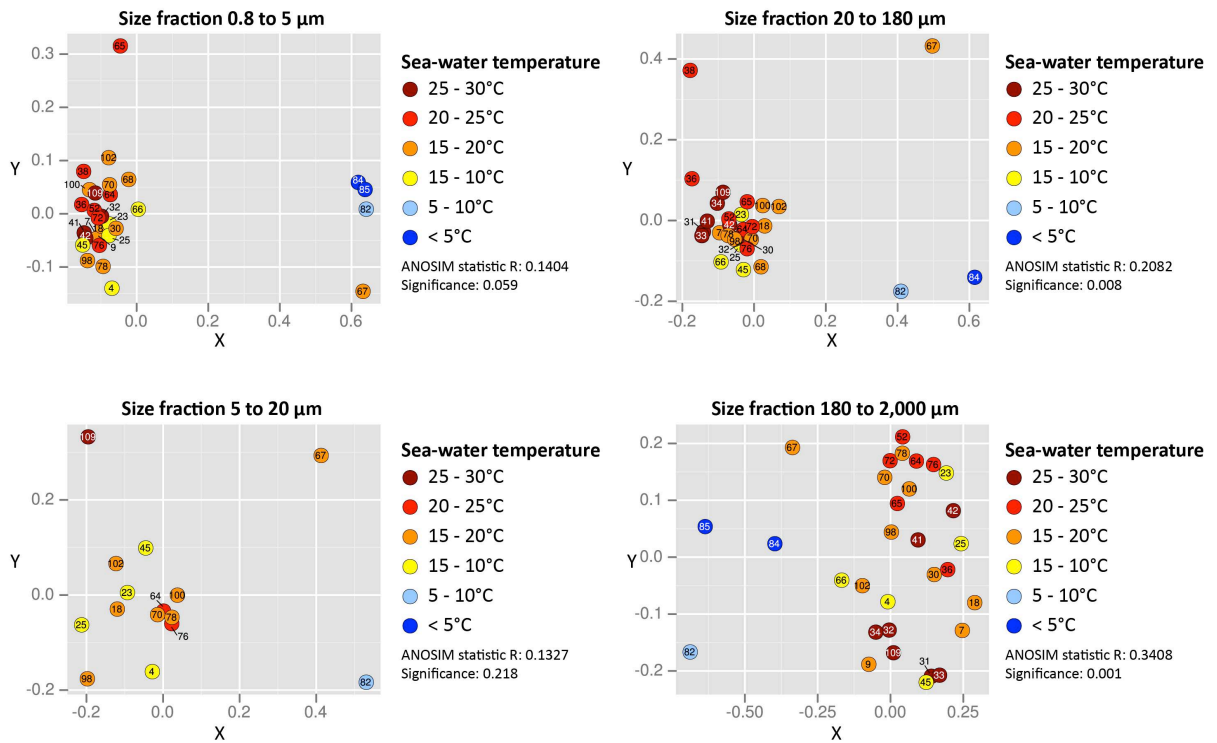


Figure 7 : NMDS illustrant la similarité des stations par fraction de tailles et leur gamme de température.

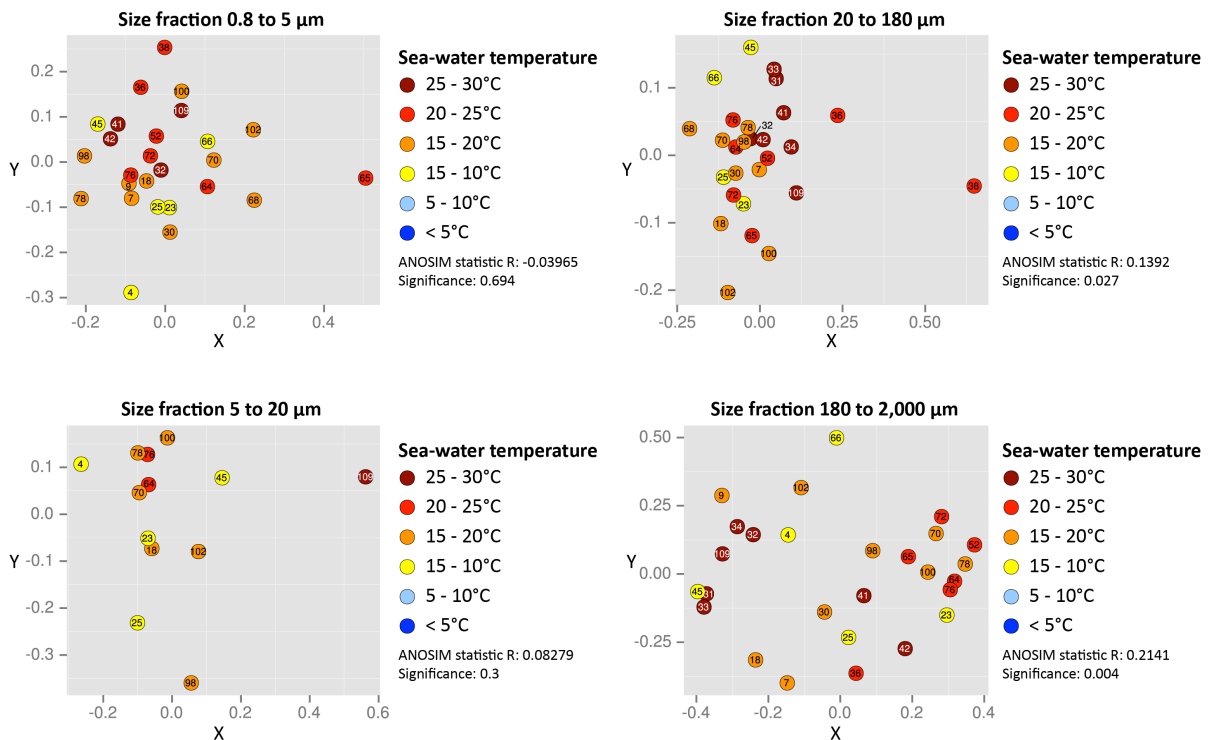


Figure 8 : NMDS illustrant la similarité des stations par fraction de tailles et leur gamme de température. Ici les stations les plus dissimilaires ont été retirées.

Nous avons considéré dans cette analyse que la température pouvait être l'unique facteur physico-chimique capable d'expliquer la structuration des communautés de dinoflagellés. Il sera bien entendu nécessaire de démontrer que ces stations sont également similaires ou distantes pour les autres facteurs environnementaux (teneurs en sels nutritifs, salinité, structure de la colonne d'eau, etc.). Cependant, nous pouvons déjà supposer qu'il existe une différence significative entre les écosystèmes côtiers de la mer Méditerranée (station 23) et l'écosystème océanique du milieu de l'océan Atlantique (station 72).

Comment, dans ces conditions, pouvons-nous expliquer la similarité de certaines communautés de dinoflagellés ? Si cette observation est confirmée, elle contredirait les paradigmes de Smayda et Reynolds (Smayda & Reynolds 2003; Reynolds 2006) qui associent une communauté de dinoflagellés à des conditions spécifiques de l'environnement. La question reste pour le moment encore ouverte, mais représente une piste de recherche intéressante et assez aisément accessible par l'analyse comparée des différents paramètres physico-chimiques avec la structure des communautés de dinoflagellés (toutes fractions de taille confondues) observées par metabarcoding.

---

### Remerciements :

Ce travail préliminaire a été grandement facilité par Stéphane Audic qui m'a aidé à produire dans un temps très court les premières NMDS de ce document. Merci.

---

## Références

- Cullen JJ, Franks PJ, Karl DM, Longhurst A (2002) Physical influences on marine ecosystem dynamics. In: *The Sea*, pp. 297–336. The sea, New York.
- Margalef R (1978) Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica acta*, **1**, 493–509.
- Reynolds CS (2006) *The Ecology of Phytoplankton*. Cambridge.
- Smayda TJ, Reynolds CS (2001) Community Assembly in Marine Phytoplankton: Application of Recent Models to Harmful Dinoflagellate Blooms. *Journal of Plankton Research*, **23**, 447–461.
- Smayda TJ, Reynolds CS (2003) Strategies of marine dinoflagellate survival and some rules of assembly. *Journal of Sea Research*, **49**, 95–106.
- Taylor FJR, Hoppenrath M, Saldarriaga JF (2007) Dinoflagellate diversity and distribution. *Biodiversity and Conservation*, **17**, 407–418.

## Conclusion et perspectives



Cette illustration de la diversité de l'univers planctonique est composée à partir d'une image faisant la couverture du livre de Christian Sardet « Plancton : aux origines du vivant » aux éditions Ulmer France parut en 2013.



## Conclusion et perspectives

Au cours de ce travail de thèse, j'ai cherché à déterminer des patrons globaux de biodiversité et de structuration des communautés de dinoflagellés planctoniques à l'échelle de l'océan mondial. Pour y parvenir, j'ai activement participé à l'élaboration puis à l'application dans le cadre de l'expédition Tara Océans d'un protocole d'échantillonnage complet du plancton à l'échelle globale qui a permis de collecter un jeu de données unique permettant une analyse morphogénétique des communautés dans leur contexte. J'ai également exploré la pertinence des approches à haut débit, à la fois en imagerie et en génétique, pour essayer de donner des éléments de réponse à ma question initiale.

Le protocole développé tout au long de cette thèse a permis de réaliser pour la première fois un échantillonnage homogène et contrôlé à l'échelle mondiale du plancton depuis les virus jusqu'au macro-plancton. Les méthodologies rassemblées/adaptées et leurs bons déroulements sur le navire Tara ont prouvé l'utilité des petits navires d'exploration pour la recherche scientifique sur le long terme. Il est possible de faire une étude océanographique diversifiée et de qualité sur un voilier lorsque l'équipement y est correctement proportionné et organisé. La constitution de multiples échantillons pour une approche multidisciplinaire (physique, chimique, génétique ou encore morphologique) d'un même espace a offert un niveau de comparaison de nos résultats et à l'échelle planétaire. Les techniques de collecte déployées ont permis de constituer un jeu de données structuré en fractions de tailles relativement homogènes et suffisamment bien réparties stratigraphiquement et géographiquement pour analyser des groupes fonctionnels du plancton à de multiples échelles.

La collecte des cellules les plus petites est une étape complexe de l'échantillonnage, car les processus de filtration employés peuvent induire un biais en exposant les cellules de grandes tailles souvent plus fragiles à une succession de contraintes pouvant les briser. Même si nos observations ne semblent pas illustrer la présence marquée d'un tel biais, il serait intéressant d'envisager d'autres processus de filtrations pour comparaisons. Chez les différentes communautés de dinoflagellés, nos analyses par metabarcoding semblent montrer qu'il y avait un bon fractionnement par classe de taille nous permettant d'étudier leurs diversités respectives. Cependant, ce fractionnement pourrait sans doute être réévalué, pour correspondre plus précisément aux tailles et formes des dinoflagellés et non comme c'est le cas ici à l'ensemble de la communauté des protistes. L'efficacité du fractionnement de taille au sein des autres groupes de protistes reste par ailleurs à démontrer.

L'analyse d'un très grand nombre d'échantillons de plancton fixé chimiquement (lugol, formol, etc.) ou *in situ* demande beaucoup de temps et une solide expérience de la part des taxonomistes. La question de l'utilisation de méthodes complémentaires d'imageries automatiques ou semi-automatiques à haut débit (comme le FlowCAM) pour accélérer les analyses des échantillons

planctoniques est donc d'actualité et a été posée pendant ma thèse. Directement acquis sur le navire, et traités plus tard en laboratoire les échantillons utilisés pour l'analyse FlowCAM ont fourni un très grand nombre d'images de particules pour l'ensemble des régions traversées. La classification de ces particules a nécessité le développement d'un outil de reconnaissance. Cet outil s'est malheureusement révélé pour certaines formes cellulaires très simples comme des sphères inefficaces pour une identification au niveau de l'espèce ou du genre. Il est donc actuellement inadapté pour une description globale de la communauté des dinoflagellés. Cependant, pour une analyse détaillée de certains genres possédant une morphologie caractéristique comme *Ceratium*, cet outil s'est révélé efficace. En effet, la principale raison des échecs de classification est le manque de caractères morphologiques clairs pour faire une distinction nette des objets détectés, accentué par un biais dans la méthode d'acquisition des images qui est basée sur une imagerie en deux dimensions des particules qui circulent en trois dimensions dans un fluide. Même si l'approche par FlowCAM pour réaliser une description complète de la diversité des dinoflagellés reste peu résolutive, cette approche pourrait néanmoins servir pour le repérage (par exemple pour des analyses couplées morphologie/séquençage) d'espèces ou de genres cibles ayant une morphologie marquée au sein des nombreux échantillons de la mission Tara Océans (ou d'autres campagnes océanographiques) ou des suivis temporels des communautés planctoniques.

Le pyroséquençage d'échantillons environnementaux a permis l'acquisition d'un nombre important de séquences provenant d'une grande variété d'environnements. Le choix du fragment V9 de l'ADNr 18S pour caractériser l'ensemble des échantillons de la mission Tara Océans a été dicté par la technologie de séquençage employée (Illumina HiSeq max  $\pm 200$ pb). Le nombre de marqueurs universels des eucaryotes disponibles dans les bases de référence et accessibles par une technologie de séquençage à haut débit était très restreint. Tout au long de mes analyses sur le groupe des dinoflagellés, j'ai pu constater que le marqueur V9 de l'ADNr 18S n'était probablement pas le plus adapté pour une analyse au niveau de l'espèce pour ce groupe. Le faible taux d'évolution et la longueur relativement réduite de ce fragment ne permettent pas une discrimination aussi précise que celle permise par d'autres marqueurs génétiques comme le LSU ou l'ITS. Cependant, le marqueur V9 m'a permis de bien caractériser les communautés échantillonnées au niveau de l'ordre et du genre. Le volume de données disponible ainsi que l'excellente profondeur d'échantillonnage m'ont permis d'observer, d'une part, la communauté dans son ensemble au niveau de l'ordre voir même parfois du genre, et d'autre part de définir une diversité et une structuration des communautés (en surface) en fonction des fractions de tailles.

Cette première étude en profondeur par metabarcoding à l'échelle mondiale d'une lignée planctonique clé, intégrant l'ensemble des océans, a dévoilé l'existence de solides et cohérents modèles permettant d'expliquer la biodiversité (en surface) par fraction de taille, dans l'espace et dans

le temps. En termes de diversité- $\alpha$ , nos données suggèrent que l'importance des dinoflagellés pélagiques par rapport à tous les autres protistes a été largement sous-estimée. La diversité ribosomique des dinoflagellés représente environ la moitié de la diversité des protistes totaux connu à la surface des océans du monde. Notre analyse est en effet probablement biaisée par l'application d'un cutoff très strict (90% d'identité avec une référence) à l'ensemble du jeu de donnée des protistes. Dans ces conditions la diversité de la biosphère rare a sans doute été que partiellement prise en compte. Notre étude a permis de dévoiler des communautés riches et structurées dans les deux plus grandes (20-180 et 180-2000 $\mu\text{m}$ ) et la plus petite fraction de taille (0,8-5 $\mu\text{m}$ ), avec des Peridinales abondants probablement sous la forme de symbiotes dans le méso-plancton, et une diversité/abondance inattendue des metabarcodes attribuées à Gymnodinales dans le pico- nano-plancton. Avec cette étude nous sommes parvenus à mettre en évidence la diversité des espèces planctoniques de très petites tailles (<5 $\mu\text{m}$ ) qui ont jusqu'ici été indétectables par une approche traditionnelle de comptage en microscopie. Compte tenu du faible nombre d'espèces décrites morphologiquement en dessous de 5 $\mu\text{m}$ , ce résultat est très étonnant, et pourrait souligner le degré de notre ignorance sur la diversité des petits dinoflagellés. Une analyse morphologique détaillée (constitution de sondes FISH, microscopie électronique) des échantillons fixés serait maintenant souhaitable pour confirmer cette hypothèse, en effet d'autres biais pourraient peut-être expliquer nos observations. Un travail sur la comparaison des échantillons d'ADN et d'ARN pourrait fournir des éléments de réponse intéressants sur la constitution de la communauté vivante ou latente au sein de la colonne d'eau. L'adéquation de l'approche par metabarcodes et des nouvelles technologies de séquençage (comme outils moléculaires quantitatifs) pour étudier l'écologie des protistes marins a toute fois été démontrée.

L'utilisation de la région V9 de l'ADNr 18S comme metabarcode a cependant montré des limites. L'ADNr 18S n'est pas le meilleur marqueur pour décrire l'ensemble des dinoflagellés, et l'utilisation de sa seule région variable V9 (ne représentant que 130pb) accentue ce phénomène. Cette dernière n'est en effet pas la plus variable et taxonomiquement résolutive. Nos résultats devront donc de toute évidence être complétés par d'autres metabarcodes ciblant des niveaux plus fins de la diversité. De même l'étude des caractéristiques fonctionnelles chez les dinoflagellés (comme la diversité chloroplastique ou l'importance relative de l'hétérotrophie vis-à-vis de l'autotrophie/mixotrophie) devrait être envisagée pour évaluer pleinement les dynamiques écologiques au sein des communautés. Le développement de protocoles morphogénétiques, permettant de lier la diversité des metabarcodes à la complexité phénotypique des cellules, sera un élément clef pour identifier et ancrer la diversité biologique dans son contexte biogéochimique. Lors de la constitution de notre base de référence, nous avons clairement identifié un manque d'homogénéité entre les assignations taxonomiques et les données génétiques produites. Le manque de diversité dans les références au



sein de certains genres (comme par exemple avoir plusieurs espèces et plusieurs séquences de référence pour chacune de ces espèces) a également conduit à des pourcentages d'identifications relativement bas. L'enrichissement de la base de références reste le point clef de toute analyse approfondie d'un groupe taxonomique. Un enrichissement de la base de référence serait envisageable par une approche morpho-génétique par *single-cell* (isolement d'une cellule, production d'une image et séquençage de plusieurs marqueurs génétiques).

Enfin, bien que préliminaire, ce travail nous a permis d'esquisser une structuration des communautés de dinoflagellés sur la base des caractéristiques environnementales. J'ai supposé que la température pourrait être un facteur capable d'expliquer la structuration des communautés de dinoflagellés au sein des différentes stations (ces analyses restent encore préliminaires). À l'exception de deux grands types d'environnements clairement distincts (les zones productives d'upwelling, et les zones froides), qui semblent être caractérisés par des communautés typiques, nous n'avons pas mis en évidence de relation claire entre la température et les communautés de surface pour les différents bassins explorés. Au contraire, il semblerait même que des communautés similaires, observées dans des écosystèmes distincts, puissent être retrouvées en correspondance avec des températures très différentes. Cette observation, si elle est confirmée, viendrait contredire les paradigmes d'écologie des communautés planctoniques qui associent des communautés de dinoflagellés données à des conditions environnementales spécifiques. Il sera nécessaire d'analyser plus en détail ces communautés similaires et de les confronter à d'autres paramètres physico-chimiques afin de confirmer nos observations. L'hypothèse d'une ressemblance des communautés dans des conditions environnementales différentes reste une perspective de travail futur pour l'étude de l'écologie des dinoflagellés par metabarcoding.

# Références générales



## Références générales

- Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. *Biophotonics*.
- Acinas SG, Antón J, Rodríguez-Valera F (1999) Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Environmental Microbiology*, **65**, 514–522.
- Adl SM, Hoppenrath M, Simpson AGB et al. (2012) The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, **59**, 429–493.
- Adl SM, Leander BS, Simpson AGB et al. (2007) Diversity, nomenclature, and taxonomy of protists. *Systematic Biology*, **56**, 684.
- Adl SM, Simpson AGB, Farmer MA et al. (2005) The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *Journal of Eukaryotic Microbiology*, **52**, 399–451.
- Álvarez E, López-Urrutia Á (2011) How to effectively sample the plankton size spectrum? A case study using FlowCAM. *Journal of Plankton Research*, **33**, 1119–1133.
- Amann R, Allers E, Moraru C et al. (2013) Single - cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology*, **15**, 2306–2318.
- Amann R, Binder BJ, Olson RJ et al. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology*, **56**, 1919–1925.
- Amann R, Pernthaler A, Pernthaler J (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology*, **68**, 3094–3101.
- Amaral-Zettler L, McCliment EA, Ducklow HW, Huse SM (2009) A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, **4**, e6372.
- Amaral-Zettler L, Zettler ER, Theroux SM et al. (2011) Microbial community structure across the tree of life in the extreme Rio Tinto. *The ISME Journal*, **5**, 42–50.
- Angly FE, Felts B, Breitbart M et al. (2006) The marine viromes of four oceanic regions. *PLoS Biology*, **4**, e368.
- Arrigo KR (2004) Marine microorganisms and global nutrient cycles. *Nature*, **437**, 349–355.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, **71**, 7724–7736.
- Balzano S, Marie D, Gourvil P, Vaultot D (2012) Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *The ISME Journal*, **6**, 1480–1498.
- Banaszak AT, Prieto RI, Trench RK (1993) *Scrippsiella Velellae* Sp. Nov.(Peridinales) and *Gloeokinium Viscum* Sp. Nov.(Phytodiales), Dinoflagellate Symbionts of Two Hydrozoans (Cnidaria). *Journal of Phycology*, **29**, 517–528.
- Banse K (2013) Reflections About Chance in My Career, and on the Top-Down Regulated World. *Annual Review of Marine Science*, **5**, 1–19.
- Behnke A, Engel M, Christen R et al. (2010) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 1–10.
- Bell JL, Hopcroft RR (2008) Assessment of Zoolmage as a tool for the classification of zooplankton. *Journal of Plankton Research*, **30**, 1351–1367.
- Benfield MC, Grosjean P, Culverhouse PF et al. (2007) RAPID: research on automated plankton identification. *Oceanography*, **20**, 172–187.

- Berney C, Fahrni J, Pawlowski J (2004) How many novel eukaryotic “kingdoms?” Pitfalls and limitations of environmental DNA surveys. *BMC Biology*, **2**, 1–13.
- Boss E, Twardowski MS, Herring S (2001) Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. *Applied Optics*, **40**, 4885–4893.
- Breitbart M (2012) Marine Viruses: Truth or Dare. *Annual Review of Marine Science*, **4**, 425–448.
- Brown L (2008) Particle Image Understanding—A Primer. *Fluid Imaging Technologies*.
- Brussaard CP, Weinbauer MG, Payet JP, Winter C (2010) Quantification of aquatic viruses by flow cytometry. *American Society of Limnology and Oceanography*, **1**, 102–109.
- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA Barcoding of Marine Metazoa. *Annual Review of Marine Science*, **3**, 471–508.
- Buitenhuis ET, Li WKW (2012) Bacterial biomass distribution in the global ocean. *Earth System Science Data*, **5**, 301–305.
- Buskey EJ, Hyatt CJ (2006) Use of the FlowCAM for semi-automated recognition and enumeration of red tide cells (*Karenia brevis*) in natural plankton samples. *Harmful Algae*, **5**, 685–692.
- Carnegie RB, Meyer GR, Blackbourn J et al. (2003) Molecular detection of the oyster parasite *Mikrocytos mackini*, and a preliminary phylogenetic analysis (vol 54, pg 219, 2003). *Diseases of Aquatic Organisms*, **54**, 219–227.
- Caron DA, Dennett MR, Lonsdale DJ, Moran DM, Shalapyonok L (2000) Microzooplankton herbivory in the Ross Sea, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, **47**, 3249–3272.
- Cavalier-Smith T, Chao EE (2004) Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). *European Journal of Protistology*, **40**, 185–212.
- Cetinić I, Perry MJ, Briggs NT et al. (2012) Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research*, **117**, C06028.
- Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The ISME Journal*, **4**, 1053–1059.
- Cheung MK, Chu KH, Li CP, Kwan HS, Wong CK (2008) Genetic diversity of picoeukaryotes in a semi-enclosed harbour in the subtropical western Pacific Ocean. *Aquatic Microbial Ecology*, **53**, 295–305.
- Christen R, Behnke A, Engel M et al. (2010a) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.
- Christen R, Chevenet F, Croce O, Hebrard M, Berry V (2010b) ScripTree: scripting phylogenetic graphics. *Bioinformatics*, **26**, 1125–1126.
- Christen R, Pawlowski J, Bachar D et al. (2011) Eukaryotic Richness in the Abyss: Insights from Pyrotag Sequencing. *PLoS ONE*, **6**, e18169.
- Claustre H, Xing X, Morel A et al. (2011) Combined processing and mutual interpretation of radiometry and fluorimetry from autonomous profiling Bio-Argo floats: Chlorophyll a retrieval. *Journal of Geophysical Research*, **116**, C06020.
- Claverie J-M, Monier A, Ogata H (2008) Taxonomic distribution of large DNA viruses in the sea. *Genome Biology*, **9**, R106.
- Cleve PT (1899) Plankton: collected by the Swedish expedition to Spitzbergen in 1898. P.A. Norstedt & Söner, Stockholm.
- Coffroth M-A, Santos SR (2005) Genetic Diversity of Symbiotic Dinoflagellates in the Genus *Symbiodinium*. *Protist*, **156**, 19–34.
- Cole JR, Chai B, Farris RJ et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic acids research*, **33**, D294–D296.
- Countway PD, Caron DA (2006) Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Applied and Environmental Microbiology*, **72**, 2496.

- Cowen RK, Guigand CM (2008) In situ ichthyoplankton imaging system (ISIS): system design and preliminary results. *Limnol Oceanogr Methods*, **6**, 126–132.
- Craig G, Alexander SJ, Anderson S et al. (2000) HoloCam: a subsea holographic camera for recording marine organisms and particles. *Proceedings of SPIE*, **4076**, 111–119.
- Croce O, Christen R, Chevenet F (2010) A new web server for the rapid identification of microorganisms. *J Microbial Biochem Technol*, **2**, 84–88.
- Cullen JJ, Franks PJ, Karl DM, Longhurst A (2002) Physical influences on marine ecosystem dynamics. In: *The Sea*, pp. 297–336. The sea, New York.
- Culverhouse PF, Williams R, Benfield M et al. (2006b) Automatic image analysis of plankton: future perspectives. *Marine Ecology Progress Series*, **312**, 297–309.
- Culverhouse PF, Williams R, Reguera B, Herry V, Gonzales-Gil S (2003) Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series*, **247**, 17–25.
- Culverhouse PF, Williams R, Simpson B (2006a) HAB Buoy: a new instrument for in situ monitoring and early warning of harmful algal bloom events. *African Journal of Marine Science*, **28**, 245–250.
- Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. New York: D Appleton.
- Davis CS, Gallager SM, Berman MS, Haury LR, Strickler JR (1992) The Video Plankton Recorder (VPR): design and initial results. *Archiv für Hydrobiologie Beiheft Ergebnisse der Limnologie*, **36**, 67–81.
- Davis JI (1997) Evolution, Evidence, and the Role of Species Concepts in Phylogenetics. *Systematic Botany*, **22**, 373–403.
- Decelle J, Siano R, Probert I, Poirier C, Not F (2012) Multiple microalgal partners in symbiosis with the acantharian *Acanthochiasma* sp. (Radiolaria). *Symbiosis*, **58**, 233–244.
- Delwiche CF (1999) Tracing the Thread of Plastid Diversity through the Tapestry of Life. *The american naturalist*, **154**, S164–S177.
- DeSantis TZ, Huber T, Hugenholtz P et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.
- Diez B, Pedros-Alio C, Massana R (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Applied and Environmental Microbiology*, **67**, 2932–2941.
- Dolan JR (2005) Biogeography of aquatic microbes. *Aquatic Microbial Ecology*, **41**, 39–48.
- Doucette GJ, Cembella AD, Boyer GL (1989) Cyst formation in the red tide dinoflagellate *Alexandrium tamarense* (dinophyceae): effects of iron stress. *Journal of Phycology*, **25**, 721–731.
- Duarte CM, del Giorgio PA (2002) Respiration in the open ocean. *Nature*, **420**, 379–384.
- Dubelaar G, Gerritzen PL (2000) CytoBuoy: a step forward towards using flow cytometry in operational oceanography. *Scientia Marina*, **64**, 255–265.
- Ducklow HW (2000) Bacterial production and biomass in the oceans. In *Microbial Ecology of the Oceans*, pp. 85–120.
- Duhaime MB, Sullivan MB (2012) Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology*, **434**, 181–186.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–1797.
- Edgcomb V, Orsi W, Bunge J et al. (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME Journal*, **5**, 1344–1356.
- Edvardsen B, Jakobsen KS, Shalchian Tabrizi K et al. (2003) Genetic Variability and Molecular Phylogeny of Dinophysis Species (Dinophyceae) From Norwegian Waters Inferred From Single Cell Analyses of Rdna. *Journal of Phycology*, **39**, 395–408.

- Ehara M, Inagaki Y, Watanabe KI, Ohama T (2000) Phylogenetic analysis of diatom *cox1* genes and implications of a fluctuating GC content on mitochondrial genetic code evolution. *Current Genetics*, **37**, 29–33.
- Embleton KV, Gibson CE, Heaney SI (2003) Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *Journal of Plankton Research*, **25**, 669–681.
- Evans C, Archer SD, Jacquet S, Wilson WH (2003) Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. *Aquatic Microbial Ecology*, **30**, 207–219.
- Evans KM, Wortley AH, Mann DG (2007) An Assessment of Potential Diatom “Barcode” Genes (*cox1*, *rbcl*, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in Sellaphora (Bacillariophyta). *Protist*, **158**, 349–364.
- Falkowski PG (2012) Ocean Science: The power of plankton. *Nature*, **483**, S17–S20.
- Falkowski PG and Knoll A. (2007) Evolution of Primary Producers in the Sea. *American Press*.
- Faria JC (2013) Resources of Tinn-R GUI/Editor for R Environment. *Ilheus*.
- Faust MA, Gullidge RA (2002) Identifying harmful marine dinoflagellates. *Department of Botany, National Museum of Natural History*.
- Feinstein TN, Traslavina R, Sun MY, Lin S (2002) Effects of Light on Photosynthesis, Grazing, and Population Dynamics of the Heterotrophic Dinoflagellate *Pfiesteria piscicida* (Dinophyceae). *Journal of Phycology*, **38**, 659–669.
- Fensome RA, MacRae AR, Williams GL (1998) Évolution des dinoflagellés et variation de leur diversité dans le temps. *Marine Micropaleontology*, 1–6.
- Fensome RA, Taylor FJR, Norris G et al. (1993) A classification of living and fossil dinoflagellates. *Micropaleontology Press*, **7**, 1–351.
- Ferat J-L, Gouar M, Michel F (1994) Multiple group II self-splicing introns in mobile DNA from *Escherichia coli*. *Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie*, **317**, 141–148.
- Fernandes JA, Irigoien X, Boyra G, Lozano JA, Inza I (2008) Optimizing the number of classes in automated zooplankton classification. *Journal of Plankton Research*, **31**, 19–29.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski PG (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, **281**, 237–240.
- Figueroa RI, Bravo I, Garcés E (2006) Multiple routes of sexuality in *Alexandrium Taylori* (dinophyceae) in culture. *Journal of Phycology*, **42**, 1028–1039.
- Figueroa RI, Garcés E, Camp J (2010) Reproductive plasticity and local adaptation in the host–parasite system formed by the toxic *Alexandrium minutum* and the dinoflagellate parasite *Parvilucifera sinerae*. *Harmful Algae*, **10**, 56–63.
- Fistarol GO, Legrand C, Rengefors K, Graneli E (2004) Temporary cyst formation in phytoplankton: a response to allelopathic competitors? *Environmental Microbiology*, **6**, 791–798.
- Foissner W, Hawksworth DL (2009) Protist diversity and geographical distribution. *Biodiversity and Conservation*, **17**, 243–259.
- Fraser JH (1968) The history of plankton sampling. *UNSECO*.
- Gaines G, Elbrächter M (1987) Heterotrophic nutrition. In: *The Biology of Dinoflagellates*, pp. 224–268. Blackwell Scientific Publications London, Blackwell, Oxford, UK.
- Gallienne CP, Robins DB (2001) Is *Oithona* the most important copepod in the world's oceans? *Journal of Plankton Research*, **23**, 1421–1432.
- Galluzzi L, Penna A, Bertozzini E et al. (2004) Development of a Real-Time PCR Assay for Rapid Detection and Quantification of *Alexandrium minutum* (a Dinoflagellate). *Applied and Environmental Microbiology*, **70**, 1199–1206.
- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

- Gao YH, Chen CP, Li Y (2001) Marine nanoplanktonic diatoms from the coastal waters of Hong Kong. *Proceedings of an International Workshop Reunion Conference*.
- Garcés E, Delgado M, Masó M, Camp J (1998) Life history and in situ growth rates of *Alexandrium taylori* (dinophyceae, pyrrrophyta). *Journal of Phycology*, **34**, 880–887.
- Gascuel O, Desper R (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of computational biology*, **9**, 687–705.
- Geer LY, Marchler-Bauer A, Geer RC et al. (2009) The NCBI BioSystems database. *Nucleic acids research*, **38**, D492–D496.
- Gilabert J (2001) Short-term variability of the planktonic size structure in a Mediterranean coastal lagoon. *Journal of Plankton Research*, **23**, 219–226.
- Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, **345**, 60–63.
- Godhe A, Asplund ME, Harnstrom K et al. (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Applied and Environmental Microbiology*, **74**, 7174–7182.
- Gómez F (2012a) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *CICIMAR Océánides*, **27**, 65–140.
- Gómez F (2012b) A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity*, **10**, 267–275.
- Gómez F (2013) Problematic Biases in the Availability of Molecular Markers in Protists: The Example of the Dinoflagellates. *Acta Protozoologica*, **52**, 21–33.
- Gómez F, Moreira D, López-García P (2010) Molecular Phylogeny of Noctiluroid Dinoflagellates (Noctilucales, Dinophyceae). *Protist*, **161**, 466–478.
- Gong J, Dong J, Liu X, Massana R (2013) Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Annals of Anatomy*, **164**, 369–379.
- Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular biology and evolution*, **27**, 221–224.
- Graham LE, Wilcox LW (2000) *Algae*. Upper Saddle River: Prentice-Hall.
- Granéli E, Turner JT (2007) *Ecology of harmful algae*. Springer.
- Groben R, John U, Eller G, Lange M, Medlin LK (2004) Using fluorescently-labelled rRNA probes for hierarchical estimation of phytoplankton diversity – a mini-review. *Nova Hedwigia*, **79**, 313–320.
- Grosjean P, Picheral M, Warembourg C, Gorsky G (2004) Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES Journal of Marine Science: Journal du Conseil*, **61**, 518–525.
- Gross L (2007) Untapped Bounty: Sampling the Seas to Survey Microbial Biodiversity. *PLoS Biology*, **5**, e85.
- Grzebyk D, Berland B (1996) Influences of temperature, salinity and irradiance on growth of *Prorocentrum minimum* (Dinophyceae) from the Mediterranean Sea. *Journal of Plankton Research*, **18**, 1837–1849.
- Guidi L, Picheral M, Stemmann L et al. (2010) The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology And Oceanography*, **8**, 462–473.
- Guillou L, Bachar D, Audic S et al. (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic acids research*, **41**, D597–D604.
- Guillou L, Bachar D, Christen R (2013) Detection of Introns in Eukaryotic Small Subunit Ribosomal RNA Gene Sequences. *Dataset Papers in Biology*, **2013**, 1–12.
- Guillou L, Massana R, Viprey M et al. (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, **10**, 3349–3365.
- Guiry MD (2012) How many species of algae are there? *Journal of Phycology*, **48**, 1057–1063.



- Guiry MD, M GG (2012) Algaebase: Listing the World's Algae. World-wide electronic publication, National University of Ireland, Galway. [www.algaebase.org](http://www.algaebase.org) searched on August 2012.
- Hackett JD, Anderson DM, Erdner DL, Bhattacharya D (2004) Dinoflagellates: a remarkable evolutionary experiment. *American Journal of Botany*, **91**, 1523–1534.
- Hallegraeff GM (2003) Harmful algal blooms: a global overview. *Phycologia*, **11**, 25–49.
- Hamsher SE, Evans KM, Mann DG, Poulíčková A, Saunders GW (2011) Barcoding Diatoms: Exploring Alternatives to COI-5P. *Protist*, **162**, 405–422.
- Harrison PJ, Furuya K, Glibert PM et al. (2011) Geographical distribution of red and green Noctiluca scintillans. *Chinese Journal of Oceanology and Limnology*, **29**, 807–831.
- Hartmann M, Howes CG, VanInsberghe D et al. (2012) Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. *The ISME Journal*, **6**, 2320–2320.
- Havskum H, Hansen PJ, Berdalet E (2005) Effect of turbulence on sedimentation and net population growth of the dinoflagellate *Ceratium tripos* and interactions with its predator, *Fragilidium subglobosum*. *Limnology And Oceanography*, **50**, 1543–1551.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- Hebert PDN, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S96–S99.
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biology*, **2**, e312.
- Heger TJ, Pawlowski J, Lara E et al. (2011) Comparing potential COI and SSU rDNA barcodes for assessing the diversity and phylogenetic relationships of cyphoderiid testate amoebae (Rhizaria: Euglyphida). *Protist*, **162**, 131–141.
- Henze K, Martin W (2003) Evolutionary biology: essence of mitochondria. *Nature*, **426**, 127–128.
- Herman AW, Beanlands B, Phillips EF (2004) The next generation of optical plankton counter: the Laser-OPC. *Journal of Plankton Research*, **26**, 1135–1145.
- Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology*, **66**, 411–453.
- Hingamp P, Acinas SG, Grimsley N et al. (2013) Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal*, **7**, 1678–1695.
- Hinnebusch AG, Klotz LC, Blanken RL, Loeblich AR III (1981) An evaluation of the phylogenetic position of the dinoflagellate *Cryptothecodinium cohnii* based on 5S rRNA characterization. *Journal of molecular evolution*, **17**, 334–347.
- Holligan PM, Fernández E, Aiken J et al. (1993) A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the North Atlantic. *Geophysical Research Letters*, **7**, 879–900.
- Hollingsworth PM, Forrest LL, Spouge JL et al. (2009) A DNA barcode for land plants. *PNAS*, **106**, 12794–12797.
- Huber T, Hugenholtz P, Faulkner G (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, **20**, 2317–2319.
- Hugenholtz P, Huber T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 289–293.
- Hyun J-H, Kim K-H (2003) Bacterial abundance and production during the unique spring phytoplankton bloom in the central Yellow Sea. *Marine Ecology Progress Series*, **252**, 77–88.
- Ibañez F, Rombouts I, Beaugrand G et al. (2009) Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings. Biological sciences / The Royal Society*, **276**, 3053–3062.
- Iwamoto S, Checkley DM Jr, Trivedi MM (2001) REFLICS: Real-time flow imaging and classification system. *Machine Vision and Applications*, **13**, 1–13.

- Jeong HJ (1999) The Ecological Roles of Heterotrophic Dinoflagellates in Marine Planktonic Community. *Journal of Eukaryotic Microbiology*, **46**, 390–396.
- Karsenti E, Sullivan MB, Acinas SG et al. (2011) A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biology*, **9**, e1001177.
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution*, **30**, 772–780.
- Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, **26**, 1899–1900.
- Ki J-S (2011) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. *Journal of Applied Phycology*, **24**, 1035–1043.
- Kirchman DL, Mitchell R (Eds.) (2008) Microbial ecology of the oceans. Wiley Blackwell.
- Kohli GS, Neilan BA, Brown MV, Hoppenrath M, Murray SA (2013) Cob gene pyrosequencing enables characterization of benthic dinoflagellate diversity and biogeography. *Environmental Microbiology*, **16**, 467–485.
- Laabir M, Amzil Z, Lassus P et al. (2007) Viability, growth and toxicity of *Alexandrium catenella* and *Alexandrium minutum* (Dinophyceae) following ingestion and gut passage in the oyster *Crassostrea gigas*. *Aquatic Living Resources*, **20**, 51–57.
- Labry C, Erard Le Denn E, Chapelle A et al. (2008) Competition for phosphorus between two dinoflagellates: A toxic *Alexandrium minutum* and a non-toxic *Heterocapsa triquetra*. *Journal of Experimental Marine Biology and Ecology*, **358**, 124–135.
- LaJeunesse TC, Lambert G, Andersen RA, Galbraith DW, Coffroth M-A (2005) Symbiodinium (Pyrrhophyta) Genome Sizes (Dna Content) Are Smallest Among Dinoflagellates. *Journal of Phycology*, **41**, 880–886.
- Larkin A-M, Blackshields G, Brown NP, Chenna R (2007) Clustal W and Clustal X version 2.0. *Bioinformatics applications note*, **23**, 2947–2948.
- Laursen L (2011) Spain's ship comes in. *Nature*, **475**, 16–17.
- Leander BS, Keeling PJ (2003) Morphostasis in alveolate evolution. *Trends in Ecology & Evolution*, **18**, 395–402.
- Lecroq B, Lejzerowicz F, Bachar D et al. (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *PNAS*, **108**, 13177–13182.
- Lee RE (2007) Phycology, Fourth edition. Cambridge.
- Li WKW (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology And Oceanography*, **43**, 1746–1753.
- Lin S (2011) Genomic understanding of dinoflagellates. *Research in microbiology*, **162**, 551–569.
- Lin S, Zhang H, Hou Y, Zhuang Y, Miranda L (2009) High-Level Diversity of Dinoflagellates in the Natural Environment, Revealed by Assessment of Mitochondrial *cox1* and *cob* Genes for Dinoflagellate DNA Barcoding. *Applied and Environmental Microbiology*, **75**, 1279–1290.
- Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*, **438**, 86–89.
- Litaker RW, Vandersea MW, Kibler SR et al. (2002) Life cycle of the heterotrophic dinoflagellate *Pfiesteria piscicida* (dinophyceae). *Journal of Phycology*, **38**, 442–463.
- Litaker RW, Vandersea MW, Kibler SR et al. (2007) Recognizing dinoflagellate species using ITS rDNA sequences. *Journal of Phycology*, **43**, 344–355.
- Liu H, Aris-Brosou S, Probert I, de Vargas C (2009) A timeline of the environmental genetics of the haptophytes. *Molecular biology and evolution*, **27**, 161–176.
- Lohmann H (1911) Über das nannoplankton und die Zentrifugierung kleinster Wasserproben zur Gewinnung desselben in lebendem Zustande. *Internationale Revue gesamten Hydrobiologie*, **4**, 1–38.
- Longhurst AR (2007) Ecological geography of the sea. *Academic Press*, New York.

- Lopez-Garcia P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**, 603–607.
- Lundholm N, Moestrup Ø (2006) The biogeography of harmful algae. In: *Ecology of harmful algae*, pp. 23–35. Springer.
- Manghisi A, Morabito M, Bertuccio C et al. (2010) Is routine DNA barcoding an efficient tool to reveal introductions of alien macroalgae? A case study of *Agardhiella subulata* (Solieriaceae, Rhodophyta) in Cape Peloro lagoon (Sicily, Italy). *Cryptogamie Algologie*, **31**, 423.
- Mann DG (1999) The species concept in diatoms. *Phycologia*, **38**, 437–495.
- Margalef R (1978) Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica acta*, **1**, 493–509.
- Marko PB, Lee SC, Rice AM et al. (2004) Fisheries: Mislabelling of a depleted reef fish. *Nature*, **430**, 309–310.
- Martin R, Quigg A (2013) Tiny Plants That Once Ruled the Seas. *Scientific American*.
- Massana R, Balagué V, Guillou L, Pedros-Alio C (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiology Ecology*, **50**, 231–243.
- Massana R, Pedrós-Alió C (2008) Unveiling new microbial eukaryotes in the surface ocean. *Current opinion in microbiology*, **11**, 213–218.
- McFadden GI, Reith ME, Munholland J, Lang-Unnasch N (1996) Plastid in human parasites. *Nature*, **381**, 482.
- Metfies K, Valentin K, Medlin LK, Mehl H, Wiltshire K (2006) Picoeukaryotic plankton diversity at the Helgoland time series site as assessed by three molecular methods. *Microbial ecology*, **52**, 53–71.
- Moon-van der Staay SY, De Wachter R, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**, 607–610.
- Moran MA, Satinsky B, Gifford SM et al. (2012) Sizing up metatranscriptomics. *The ISME Journal*, **7**, 237–243.
- Moreira D, Lopez-Garcia P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology*, **10**, 31–38.
- Moriarty R, Buitenhuis ET (2012) Distribution of known macrozooplankton abundance and biomass in the global ocean. *Earth System Science Data*, **5**, 187–220.
- Nassonova E, Fahrni J, Smirnov AV, Pawlowski J (2010) Barcoding Amoebae: Comparison of SSU, ITS and COI Genes as Tools for Molecular Identification of Naked Lobose Amoebae. *Protist*, **161**, 102–115.
- Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, **48**, 443–453.
- Newton I (1972) *Philosophiae Naturalis Principia Mathematica* (mathematical principles of natural philosophy). London (1687).
- Noble RT, Fuhrman JA (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology*, **14**, 113–118.
- Not F, del Campo J, Balague V, de Vargas C, Massana R (2009) New Insights into the Diversity of Marine Picoeukaryotes. *PLoS ONE*, **4**, e7143.
- Not F, Siano R, Kooistra WHCF et al. (2012) Diversity and Ecology of Eukaryotic Marine Phytoplankton. In: *Genomic insights into the biology of algae*, pp. 1–53. Elsevier.
- Ochoa J, Maske H, Sheinbaum J, Candela J (2013) Diel and lunar cycles of vertical migration extending to below 1000 m in the ocean and the vertical connectivity of depth-tiered populations. *Limnology And Oceanography*, **58**, 1207–1214.
- Orr RJS, Murray SA, Stüken A, Rhodes L, Jakobsen KS (2012) When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE*, **7**, e50004.
- Ortner PB, Cummings SR, Afring RP, Edgerton HE (1979) Silhouette photography of oceanic zooplankton. *Nature*, **277**, 50–51.
- Pawlowski J, Simpson AGB, Audic S et al. (2012) CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biology*, **10**, e1001419.

- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 2444–2448.
- Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? *Trends in Microbiology*, **14**, 257–263.
- Pedrós-Alió C (2012) The Rare Bacterial Biosphere. *Annual Review of Marine Science*, **4**, 449–466.
- Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef–Australia. *Scientia Marina*, **70**, 7–12.
- Percopo I, Siano R, Rossi R et al. (2013) A new potentially toxic Azadinium species (Dinophyceae) from the Mediterranean Sea, *A. dexteroporum* sp. nov. *Journal of Phycology*, **49**, 950–966.
- Pernthaler A, Pernthaler J, Amann R (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology*, **68**, 3094–3101.
- Persson A, Smith BC, Wikfors GH, Quilliam M (2006) Grazing on toxic *Alexandrium fundyense* resting cysts and vegetative cells by the eastern oyster (*Crassostrea virginica*). *Harmful Algae*, **5**, 678–684.
- Pfiester LA, Anderson DM (1987) Dinoflagellate reproduction. In: *The Biology of Dinoflagellates*, pp. 611–648. The Biology of Dinoflagellates (edited by Tylor FJR).
- Pizay M-D, Lemée R, Simon N et al. (2009) Night and Day Morphologies in a Planktonic Dinoflagellate. *Protist*, **160**, 565–575.
- Pochon X, Putnam HM, Burki F, Gates RD (2012) Identifying and Characterizing Alternative Molecular Markers for the Symbiotic and Free-Living Dinoflagellate Genus *Symbiodinium*. *PLoS ONE*, **7**, e29816.
- Pomroy AJ (1989) Scanning electron microscopy of *Heterocapsa minima* sp. nov. (Dinophyceae) and its seasonal distribution in the Celtic Sea. *British Phycological Journal*, **24**, 131–135.
- Probert I, Siano R, Poirier C et al. (2014) *Brandtodinium* Gen. Nov. and *B. Nutriculum* Comb. Nov. (Dinophyceae), a Dinoflagellate Commonly Found in Symbiosis with Polycystine (Radiolaria). *Journal of Phycology*, na.
- Pruesse E, Quast C, Knittel K et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, **35**, 7188–7196.
- Puigserver M, Zingone A (2002) *Procoentrum nux* sp. nov. (Dinophyceae), a small planktonic dinoflagellate from the Mediterranean Sea, and discussion of *P. nanum* and *P. pusillum*. *Phycologia*, **41**, 29–38.
- Raghukumar S (2002) Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *European Journal of Protistology*, **38**, 127–145.
- Reynolds CS (2006) The Ecology of Phytoplankton. *Cambridge*.
- Rombouts I, Beaugrand G, Ibanez F et al. (2009) Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings. Biological sciences / The Royal Society*, **276**, 3053–3062.
- Rowan R, Powers DA (1992) Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 3639–3643.
- Rowan R, Whitney SM, Fowler A, Yellowlees D (1996) Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. *The Plant cell*, **8**, 539–553.
- Rynearson TA, Armbrust EV (2004) Genetic differentiation among populations of the planktonic marine diatom *ditylum brightwellii* (bacillariophyceae). *Journal of Phycology*, **40**, 34–43.
- Saldarriaga JF, Taylor FJR, Cavalier-Smith T, Menden-Deuer S, Keeling PJ (2004) Molecular data and the evolutionary history of dinoflagellates. *European Journal of Protistology*, **40**, 85–111.
- Saldarriaga JF, Taylor FJR, Keeling PJ, Cavalier-Smith T (2001) Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *Journal of molecular evolution*, **53**, 204–213.
- Samson S, Hopkins T, Remsen A et al. (2001) A system for high-resolution zooplankton imaging. *IEEE Journal Of Oceanic Engineering*, **26**, 671–676.
- Santini S, Logares R, Audic S et al. (2012) Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *The ISME Journal*, **6**, 1823–1833.

- Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1879.
- Saunders GW (2009) Routine DNA barcoding of Canadian Gracilariales (Rhodophyta) reveals the invasive species *Gracilaria vermiculophylla* in British Columbia. *Molecular Ecology Resources*, **9**, 140–150.
- Saunders GW, Hill D, Sexton JP, Andersen RA (1997) Small-subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. In: *Origins of algae and their plastids*. Origins of Algae and their Plastids.
- Saunders GW, Kucera H (2010) An evaluation of *rbcl*, *tufA*, *UPA*, *LSU* and *ITS* as DNA barcode markers for the marine green macroalgae. *Cryptogamie Algologie*, **31**, 487–528.
- Scanlan DJ, Ostrowski M, Mazard S et al. (2009) Ecological Genomics of Marine Picocyanobacteria. *Microbiology and Molecular Biology Reviews*, **73**, 249–299.
- Schnepf E, Elbrächter M (1992) Nutritional strategies in dinoflagellates: A review with emphasis on cell biological aspects. *European Journal of Protistology*, **28**, 3–24.
- Seifert KA, Samson RA, deWaard JR et al. (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3901–3906.
- Siano R, Kooistra WHCF, Montresor M, Zingone A (2009) Unarmoured and Thin-Walled Dinoflagellates from The Gulf of Naples, With The Description of *Woloszynskia Cincta* sp. nov. (Dinophyceae, Suessiales). *Phycologia*, **48**, 44–65.
- Siano R, Montresor M, Probert I, Not F, de Vargas C (2010) *Pelagodinium* gen. nov. and *P. béii* comb. nov., a Dinoflagellate Symbiont of Planktonic Foraminifera. *Protist*, **161**, 385–399.
- Siddall ME, Fontanella FM, Watson SC, Kvist S, Erséus C (2009) Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. *Systematic Biology*, **58**, 445–451.
- Sieracki CK, Sieracki ME, Yentsch CS (1998) An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*, **168**, 285–296.
- Signorini SR, McClain CR (2009) Environmental factors controlling the Barents Sea spring-summer phytoplankton blooms. *Geophysical Research Letters*, **36**, L10604.
- Simon N, Campbell L, Örnólfsson E et al. (2000) Oligonucleotide Probes for the Identification of Three Algal Groups by Dot Blot and Fluorescent Whole - Cell Hybridization. *Journal of Eukaryotic Microbiology*, **47**, 76–84.
- Simpson AGB, Adl SM, Leander BS et al. (2007) Diversity, nomenclature, and taxonomy of protists. *Systematic Biology*, **56**, 684.
- Simpson AGB, Roger AJ, Silberman JD et al. (2002) Evolutionary history of “early-diverging” eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Molecular biology and evolution*, **19**, 1782–1791.
- Siokou-Frangou I, Christaki U, Mazzocchi MG et al. (2010) Plankton in the open Mediterranean Sea: a review. *Biogeosciences*, **7**, 1543–1586.
- Skovgaard A, Karpov SA, Guillou L (2012) The Parasitic Dinoflagellates *Blastodinium* spp. Inhabiting the Gut of Marine, Planktonic Copepods: Morphology, Ecology, and Unrecognized Species Diversity. *Frontiers in Microbiology*, **3**, 1–22.
- Skovgaard A, Massana R, Saiz E (2007) Parasitic Species of the Genus *Blastodinium* (Blastodiniophyceae) Are Peridinioid Dinoflagellates. *Journal of Phycology*, **43**, 553–560.
- Slade WH, Boss E, Dall'Olmo G et al. (2010) Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *American Meteorological Society*, **27**, 1733–1746.
- Smalley GW, Coats DW (2002) Ecology of the Red - Tide Dinoflagellate *Ceratium furca*: Distribution, Mixotrophy, and Grazing Impact on Ciliate Populations of Chesapeake Bay. *Journal of Eukaryotic Microbiology*, **49**, 63–73.

- Smayda TJ (1958) Biogeographical Studies of Marine Phytoplankton. *Oikos*, **9**, 158.
- Smayda TJ (1997) Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology And Oceanography*, **42**, 1137–1153.
- Smayda TJ, Reynolds CS (2001) Community Assembly in Marine Phytoplankton: Application of Recent Models to Harmful Dinoflagellate Blooms. *Journal of Plankton Research*, **23**, 447–461.
- Smayda TJ, Reynolds CS (2003) Strategies of marine dinoflagellate survival and some rules of assembly. *Journal of Sea Research*, **49**, 95–106.
- Solonenko SA, Hallam SJ, Ignacio-Espinoza JC et al. (2013) Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics*, **14**, 320.
- Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13486–13491.
- Stemmann L, Boss E (2012) Plankton and Particle Size and Packaging: From Determining Optical Properties to Driving the Biological Pump. *Annual Review of Marine Science*, **4**, 263–290.
- Stemmann L, Gorsky G, Ohman MD et al. (2010) Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research*, **32**, 285.
- Stemmann L, Youngbluth M, Robert K et al. (2008) Global zoogeography of fragile macrozooplankton in the upper 100–1000 m inferred from the underwater video profiler. *ICES Journal of Marine Science: Journal du Conseil*, **65**, 433–442.
- Stern RF, Andersen RA, Jameson I et al. (2012) Evaluating the Ribosomal Internal Transcribed Spacer (ITS) as a Candidate Dinoflagellate Barcode Marker. *PLoS ONE*, **7**, e42780.
- Stern RF, Horak A, Andrew RL et al. (2010) Environmental Barcoding Reveals Massive Dinoflagellate Diversity in Marine Environments. *PLoS ONE*, **5**, e13991.
- Subirana L, Grimsley N, Thomas R et al. (2011) Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. *Environmental Microbiology*, **13**, 1412–1420.
- Sunde M (2005) Class I integron with a group II intron detected in an Escherichia coli strain from a free-range reindeer. *Antimicrobial agents and chemotherapy*, **49**, 2512–2514.
- Suttle CA (2007) Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, **5**, 801–812.
- Taberlet P, Coissac E, Pompanon F (2012) Towards next - generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.
- Taberlet P, Coissac E, Pompanon F et al. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic acids research*, **35**, e14–e14.
- Takabayashi M, Santos SR, Cook CB (2004) Mitochondrial DNA Phylogeny of the Symbiotic Dinoflagellates (Symbiodinium, Dinophyta). *Journal of Phycology*, **40**, 160–164.
- Taylor FJR (2001) Harmful Algal Blooms 2000 : Proceedings of the Ninth International Conference, Hobart, Australia, 7-11 February 2000-2001. *Harmful Algal Blooms*, 1–5.
- Taylor FJR (2004) Extraordinary Dinoflagellates: Past and Present. 61–66.
- Taylor FJR, Hoppenrath M, Saldarriaga JF (2007) Dinoflagellate diversity and distribution. *Biodiversity and Conservation*, **17**, 407–418.
- Taylor FJR, Pahlinger U (1987) Ecology of dinoflagellates. In: Taylor FJR (ed) The biology of dinoflagellates. *Botanical Monographs, vol 21*. Blackwell Science Publications, Boston, pp 399–529.
- Thessen AE, Patterson DJ, Murray SA (2012) The Taxonomic Significance of Species That Have Only Been Observed Once: The Genus Gymnodinium (Dinoflagellata) as an Example. *PLoS ONE*, **7**, e44015.
- Tintore J, Testor P, Meyers G et al. (2010) Gliders as a component of future observing systems. *Oceanobs09net*, **2**.

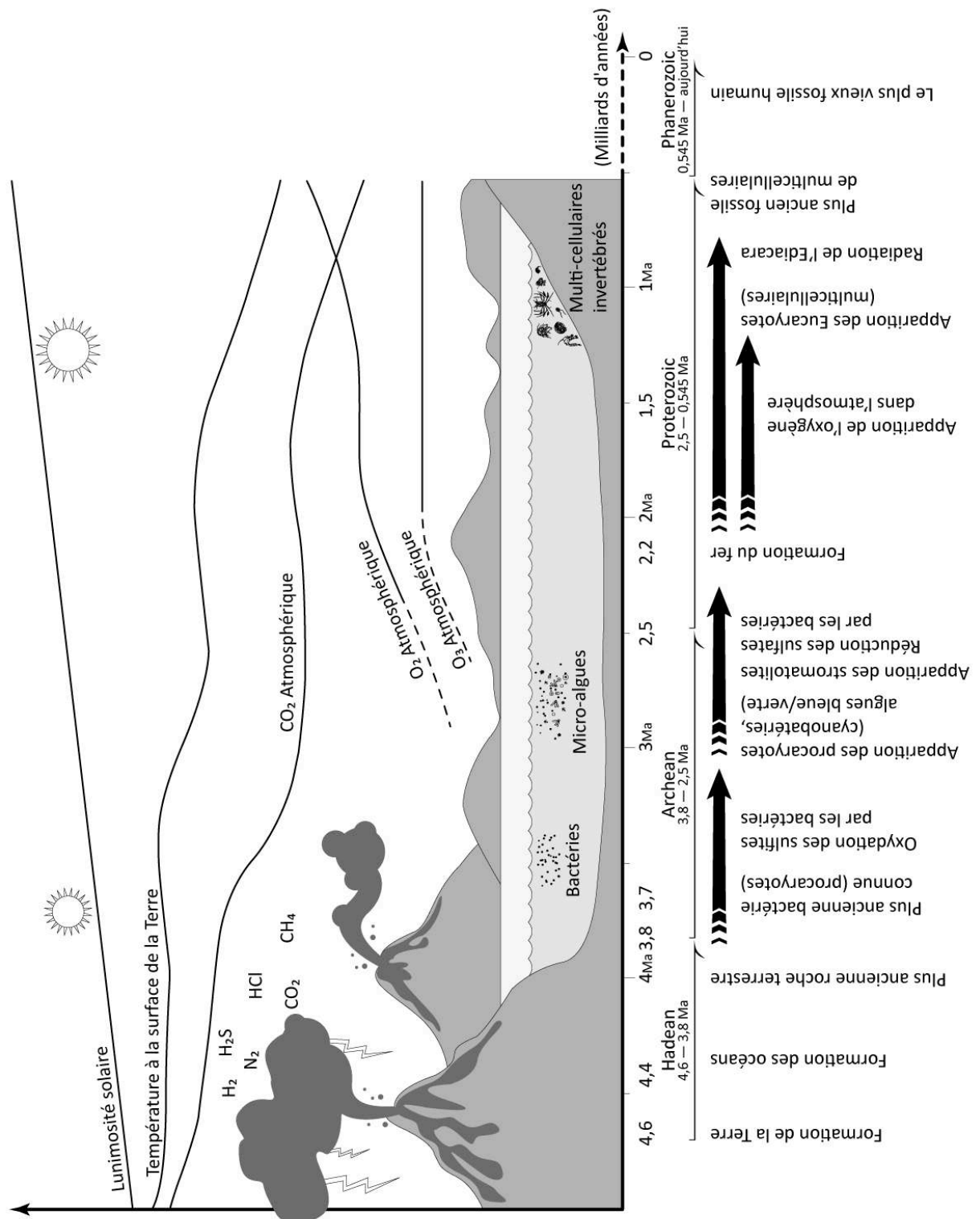
- Trench RK, Blank RJ (1987) Symbiodinium Microadriaticum Freudenthal, S. Goreauii Sp. Nov., S. Kawagutii Sp. Nov. and S. Pilosum Sp. Nov.: Gymnodinioid Dinoflagellate Symbionts of Marine Invertebrates. *Journal of Phycology*, **23**, 469–481.
- Turner JT (2004) The importance of small planktonic copepods and their roles in pelagic marine food webs. *Zool Stud*, **43**, 255–266.
- Watson EB and Harrison TM, (2005) Zircon thermometer reveals minimum melting conditions on earliest Earth. *Science*, **308**, 841–844.
- Weinbauer MG, Rowe JM, Wilhelm SW (2010) Determining rates of virus production in aquatic systems by the virus reduction approach. *American Society of Limnology and Oceanography*, **1**, 1–8.
- Wiebe PH, Ashjian C, Sutor M, Lavery A, Fratantoni D (2010) Spatial and Temporal Variability of Zooplankton Thin Layers: The Effects of Composition and Orientation on Acoustic Detection of Layers. *Onrnavymil*.
- Wiebe PH, Benfield MC (2001) Zooplankton Sampling With Nets And Trawls. In: *Encyclopedia of Ocean Sciences*, pp. 3237–3253. Elsevier.
- Wiebe PH, Benfield MC (2003) From the Hensen net toward four-dimensional biological oceanography. *Progress in Oceanography*, **56**, 7–136.
- Williams M (2012) Organization and Preliminary Analysis of Global Planktonic Community Flowcam Images From the Tara Oceans Expedition.
- Worden AZ (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecology*, **43**, 165–175.
- Worden AZ (2008) Ecology and diversity of picoeukaryotes. In: *Microbial ecology of the oceans* (eds Kirchman DL, Mitchell R), pp. 159–205. Wiley Blackwell.
- WoRMS Editorial Board (2012) WoRMS - World Register of Marine Species. Available from [www.marinespecies.org](http://www.marinespecies.org) at VLIZ. Accessed August 2012.
- Yilmaz S, Deng L, Sullivan MB et al. (2012) Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging. *mBio*, **3**, e00373–12.
- Zarauz L, Irigoien X (2008) Effects of Lugol's fixation on the size structure of natural nano-microplankton samples, analyzed by means of an automatic counting method. *Journal of Plankton Research*, **30**, 1297–1303.
- Zhu F, Massana R, Not F, Marie D, Vaultot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, **52**, 79–92.
- Zingone A, Sarno D, Siano R, Marino D (2010) The importance and distinctiveness of small-sized phytoplankton in the Magellan Straits. *Polar Biology*, **34**, 1269–1284.

# **Annexes**





## Annexe — préambule



**Annexe I — Préambule :** Nos connaissances actuelles nous permettent de situer les origines de la vie dans les océans et nous indique que la physico-chimie de notre planète est étroitement liée à l'apparition de cette vie. La composition de l'atmosphère de notre planète est encore aujourd'hui très largement influencée par les processus de vie des microorganismes marins.



# The *Tara* Oceans Expedition: A comprehensive sampling of marine plankton biota in their environmental context

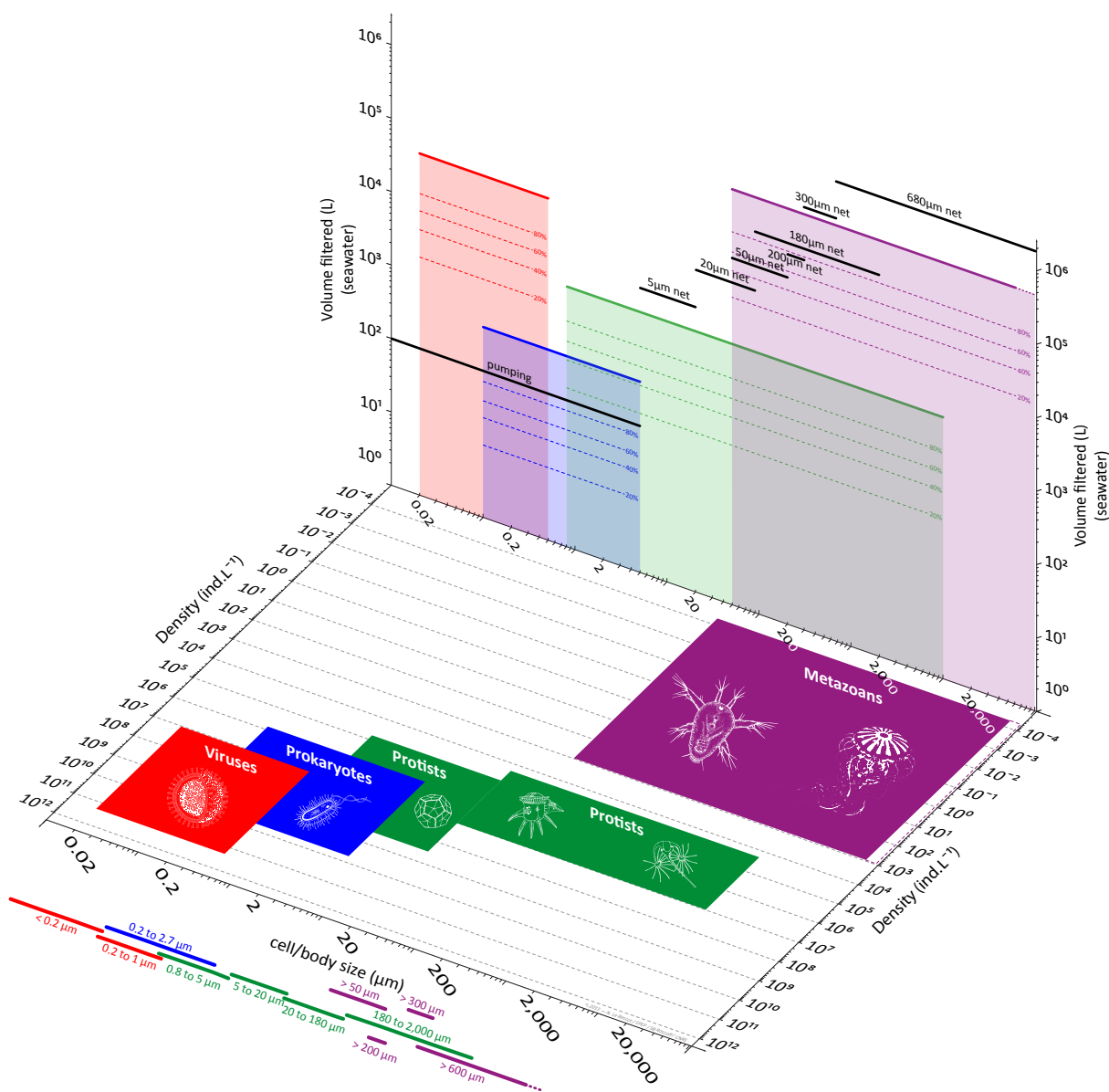
## Supplementary Information

The operational procedures described in this article are based on the February 2011 version of the protocols that had been adjusted during the first year of the expedition. Whenever modifications or adjustments of the protocols were made, these are indicated in the on-line samples list resource, along with the list of sampling legs, samplers on-board, leg reports and log sheets, available at <http://www.pangaea.de/>.

### Onboard laboratories set up

Set up of the wet laboratory (Fig. 4 and Supplementary Fig. 9) and dry laboratory (Fig. 4 and Supplementary Fig. 10) for large filtering and imaging capacities, respectively, required a year of preparation before the expedition was launched. The Dry Lab used to be a crew cabin (2 bunk beds) and it took 12 months to prepare the room (benches, plugs, continuous flow...) as a lab. The first planning phase implicated the captain, the imaging platform coordinator and several engineers alongside a representative of the company in charge of the boat refurbishing. The second phase dealt with the set up of the continuous flow system bringing water from underneath the front part of the boat directly to the Dry Lab to allow the connection of three apparatus: The MicroTSG, chlorophyll fluorometer, FRRF, and the WETLabs ac-S spectrophotometer. A seabird thermo-salinograph MicroTSG SBE-45, was installed along the in-line flowthrough system together with a chlorophyll fluorometer. With a real-time readout, it assisted the scientists with sampling decision by providing real time data. The WETLabs ac-S spectrophotometer was installed in flow through mode and provided, by means of an automated switching interval to do measurements of dissolved particles, an estimate of the particulate absorption and attenuation coefficients (See (Slade *et al.* 2010)). These measurements provided, in addition, three proxies of biogeochemical properties; Chlorophyll *a* concentration via the absorption line-height at 676 nm, Particulate organic carbon (see review regression between beam-attenuation at 650nm and POC in (Cetinić *et al.* 2012) and a small-particle size parameter (Boss *et al.* 2001). Installed in the Dry Lab the FlowCAM operation on *Tara* Oceans were aiming at: 1) quantitative characterization of the microplankton (15 to >300µm) in the pumped water samples, and 2) qualitative assessment and record of the diversity in the 20-180µm fraction of the 20µm net sample. The quantitative characterization (1) will yield phytoplankton and total particle abundance and size spectra, as well as images of the particles. The qualitative assessment of the net material (2) will give a quick look at the dominant morphotypes collected, and record these images for future

characterization of morphological diversity. The *Tara* Oceans FlowCAM was successfully set up, calibrated, and first operated by Dr Mike Sieracki at Station 18 off Malta. FlowCAM samples were run on all subsequent stations according to the protocol, with some variations due to weather conditions and station specificities (coral reef stop-overs, water masses follow up...). Typically, at each station 6 samples were run for a total of 4 hours, 600ml of sea water and 25/50,000 images.



**Supplementary Figure 1:** This plot shows the size distribution of planktonic organisms (X axis) and their range of abundance in 1L of seawater based on a literature survey (Y axis). The numerical data used to draw this plot are listed in supplementary Table 2. For each of the size fractions considered, average volumes sampled during the expedition are represented on the Z axis. Considering published estimates of species richness for each size class considered, we estimated the volume of water required to obtain a full coverage of the extant diversity assuming that all the diversity would be present everywhere, which is very likely not the case so the current volume estimate correspond to the highest one can expect (Z axis). Dotted lines represent the volume that would be required to sample a given percentage of the estimated diversity. Ultimately, this figure will be refined based on the diversity outcomes from the *Tara* Oceans expedition.

Qualitative assessment of samples were also performed using a SteREO Discovery.V20, a modular stereo microscope with motorized 20x zoom generously provide by Carl Zeiss Microscopy GmbH and an AxioVert microscope refurbished for the expedition. Both were bolted to the bench and maintained using rubber/silicon pads to ensure stability and workability. Depending on weather conditions and time, the optical engineer was able to perform extensive qualitative assessment. However, time constraints has limited the use of those two instruments during normal stations except for the coral reef stop-overs where extensive uses was possible. In parallel, we established a macrophotography studio that could be easily fitted in a Pelicase and stored in the front storage area and bunks of the central *Tara* area. This studio was scarcely used during regular stations and mainly designed for coral reef stations.

## Details of the biological sampling protocols

All the material used for sampling (e.g. containers, jars, buckets, funnels, etc.) were bleached (ca 1-5% final concentrations) and thoroughly rinsed with freshwater.

## Information associated with Supplementary Figure 2

[BGCO-CARB] BioGeoChemistry and Oceanography sampling for CARBOnate chemistry: Seawater was carefully dispensed directly from the Niskin into 500mL glass bottles with silicon tubing to avoid generating bubbles. Bottles were filled by letting seawater overflow until, at least, a volume equal to the jar volume flows out. The bottle was tightly closed with a greased glass stopper, leaving an air space in the neck of the bottle. Bottles were stored at RT. Triplicates.

[BGCO-NUT] BioGeoChemistry and Oceanography sampling for NUTrients: Seawater was dispensed directly from the Niskin into 125mL polyethylene bottles up to 2/3 of the volume and immediately stored at -20°C. Duplicates.

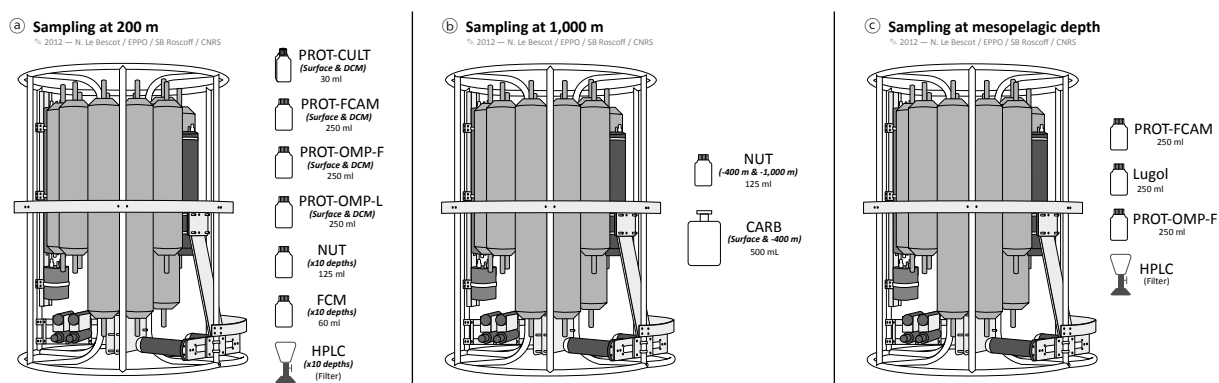
[BGCO-HPLC] BioGeoChemistry and Oceanography sampling for pigments by the HPLC method: Two liters of seawater were filtered onto 25mm GF/F filters that were stored in liquid nitrogen, using 2mL cryovials. Duplicates.

[PROT-OMP-L & PROT-OMP-F] Protist sampling for Optical Microscopy of Phytoplankton preserved in lugol or formol: Seawater was dispensed directly from the Niskin into 250mL dark glass bottles. For lugol preservation, 5mL of lugol was added to 245mL of seawater (2% final conc.) and stored at 4°C. For formol preservation, 25mL of formol was added to 225mL of seawater (3.7% final concentration) and stored at RT. Duplicates.

[**PROT-CULT**] Protist sampling for cell CULTures: Three different growth medium conditions were used. The first one had no medium and consisted of plain seawater. The second one consisted of 15mL of seawater and 15mL of K/4 culture medium. The third one consisted of 15mL of seawater and 15mL of K/4 culture medium and 2-3 drops of GeO<sub>2</sub> (Germanium oxide). Cultures were maintained in 50mL culture flasks, stored at 20°C in an incubator. Triplicates.

[**PROT-FCAM**] Protist sampling for onboard quantitative enumeration, measurement and identification of micro plankton using a FlowCAM (see Supplementary Fig.11 for details): Seawater was dispensed directly from the Niskin into 250mL plastic bottles.

[**FCM**] samples for onboard flow cytometry acquisition were collected at surface daily from the flow through system or along a water profile (10 samples within the upper 100m) using the Niskin bottles when at sampling stations. For each samples, 60ml of seawater are collected, filtered through 30µm Nitex screen. From this sample 1ml will be preserved as detailed in the [**BACT-FCM**] protocol and 1ml will be run directly on the Accuri flow cytometer installed on board.



**Supplementary Figure 2:** Summary of the sampling procedures and preservation methods for biogeochemical parameters and classical planktonic communities investigation.

### Information associated with Supplementary Figure 3

[**VIRUS-DNA-FeCl**] Viromics through iron chloride precipitation: Raw seawater was treated with FeCl<sub>3</sub> to precipitate the viruses. For 60L, 6mL of 10g Fe/L stock solution (4.83gr of FeCl<sub>3</sub>-6H<sub>2</sub>O + 100mL MilliQ water, stored at 4°C) was added (final concentration of 1mg Fe/L seawater filtrate), and shaken vigorously for 1min right after addition of FeCl<sub>3</sub> stock solution. After 1hr at room temperature, for the precipitation to occur, the FeCl<sub>3</sub>-treated viral fraction sample was filtered onto a 1.0µm, 142mm, polycarbonate (PC) membrane filter placed on top of a 0.8µm, 142mm, support filter (Pall Gelman) using a peristaltic pump at maximum pressure <15psi. Using forceps, polycarbonate filters from each sample were placed into one or two 50mL tubes (up to 3 filters per tube). Filters were stored at 4°C.

[VIRUS-qPCR] Quantitative PCR analysis: 100mL of the 0.22µm filtrate were collected into two 50mL tubes and stored at 4°C. Duplicates.

[VIRUS-SYBR] Acid nucleic staining: 0.22µm filtrate was collected in 5mL sterile cryotubes. For each, 3.920mL of filtrate was dispensed and 80µL of glutaraldehyde 25% added. After 15 minutes of incubation in the dark, samples were flash frozen and stored in liquid nitrogen. Triplicates.

[VIRUS-FCM-DMSO & VIRUS-FCM-Glut] Flow cytometry: First protocol, 1mL of 200µm pre-filtered seawater was dispensed into 2mL cryovials and 70µL of DMSO was added to each and the tube inverted to mix. Triplicates. Second protocol, 1mL of 200µm pre-filtered seawater was dispensed into 2mL cryovials and 5µL of fresh 25% glutaraldehyde added (final conc. 0.125%), mixed and incubated in the dark for 15min. Triplicates. Both DMSO and glutaraldehyde preserved samples were flash frozen and stored in liquid nitrogen.

[VIRUS-FISH] Fluorescent *in situ* hybridization: 10mL of 20µm pre-filtered seawater was fixed with 1mL of Paraformaldehyde (10% stock solution) for sub-surface and DCM samples. Triplicates. For mesopelagic samples 110mL of each (11mL of 10% PFA) were prepared. Duplicates. All samples were gently mixed and stored at 4°C between 1 and 24h before proceeding with filtrations. The fixed samples were filtered onto 0.2µm polycarbonate white filters, 25mm (Millipore, GTTP). Once filtrations of seawater samples were completed the samples were rinsed with ca 2mL of filtered MilliQ water. Upon completion, filters were dried at RT before storage in Petrislides, sealed with Parafilm and stored at -20°C.

[VIRUS-TEM] Transmission electronic microscopy: 3.920mL of 200µm pre-filtered seawater was dispensed in each 5mL sterile cryotubes and 80µL of 25% glutaraldehyde was added. Samples were incubated 15 minutes in dark and subsequently flash frozen and stored in liquid nitrogen. Duplicates.

[VIRUS-CULT], cultures attempt: Samples were collected into sterile Falcon tubes in which 50mL of 0.2µm filtrate were dispensed and stored at 4°C. Duplicates.

\*\*\*\*\*

[GIRUS-DNA-Dry] Girus metagenomics: 200L of water pre-filtered through 0.8µm (from protist filtrate). The pumping was done first through a 0.45µm Express Plus filter then through a 0.22µm Express Plus filter. Filters were changed if clogging occurred, sometimes resulting in several filters for



the same samples (typically 1-3 filters for each sample). The 142mm 0.45µm Millipore Express Plus filters and the 142mm 0.22µm Millipore Express Plus filters were placed into separate 50mL Falcon tubes and stored at -20°C. The third size fraction corresponds to 20L of 0.2µm Express Plus filtrate (of the 200L mentioned previously) filtered through a 0.1µm MF filter (until the 20L had passed, or until total filter clogging). The 0.1µm membrane was placed into a 50mL Falcon tube and stored at -20°C.

[GIRUS-CULT-Dry] Girus culture attempt: Samples were collected in duplicates and filtered onto 0.22µm membranes (47mm in diameter). One liter of 5µm pre-filtered surface seawater (or <3µm for Mesopelagic) was collected in a bottle. This water was filtered onto 0.22µm PC membranes (filter was changed if it clogged). The 0.2µm filters were placed face up in a sterile petrislide and the closed slide was sealed with parafilm. Petrislides were kept at 4°C.

\*\*\*\*\*

[BACT-DNA] Bacterial DNA was collected at 3 depths and in duplicates. For each, 100L of seawater was pre-filtered through 20µm and collected in bleached clean carboys. Then, seawater went through serial filtrations onto 3µm 142mm polycarbonate filters (Millipore) and 0.22µm 142mm (Millipore, Express Plus filters) using a peristaltic pump (flow rate between 11-13psi). The 3µm polycarbonate filter was placed in a 5mL cryotube, flash-frozen and stored in liquid nitrogen. The 0.22µm Express Plus filters was placed in a 50mL Falcon tube and stored immediately at -20°C. The filtrate was collected in 3 x 20L carboys for phage filtrations (see virus section above).

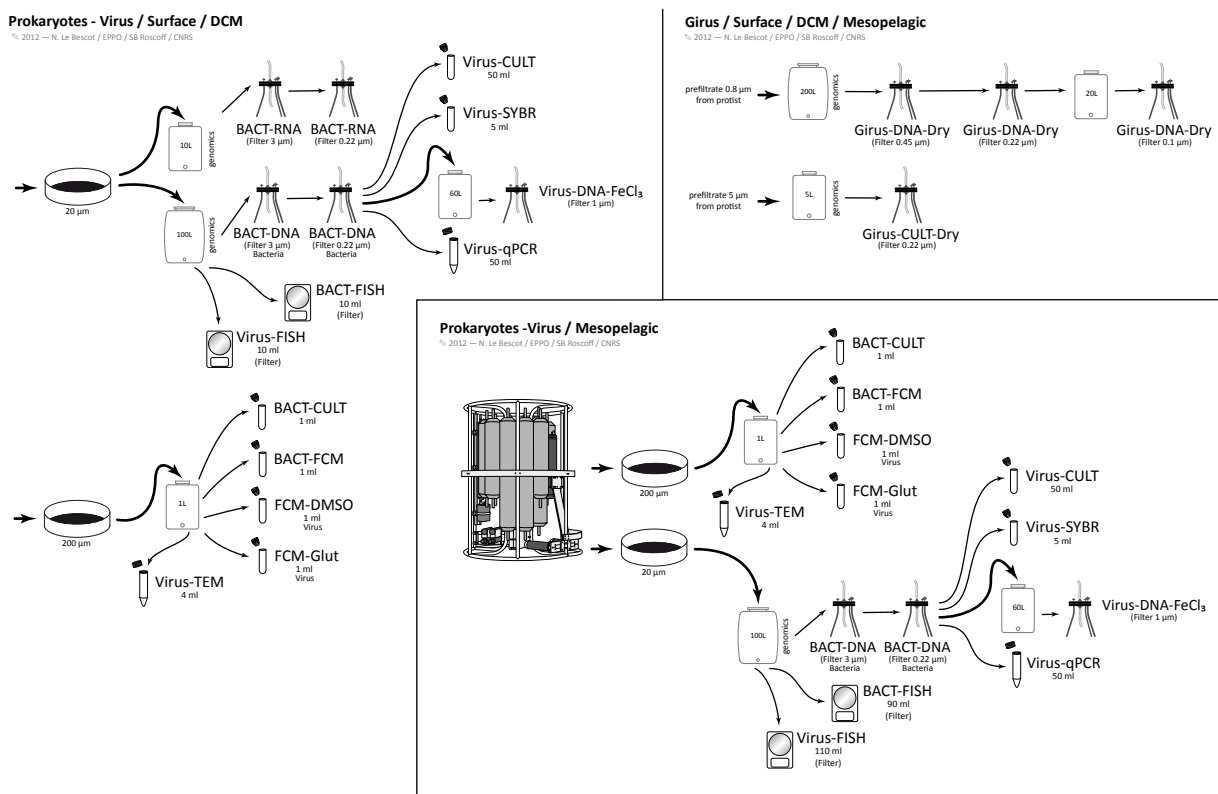
[BACT-RNA] Bacterial RNA was collected at sub-surface and DCM. Two 10L carboys were rinsed twice with 20µm pre-filtered seawater and subsequently filled up. From each carboy, the water was filtered simultaneously through serial filtrations onto 3µm 142mm polycarbonate filter (Millipore) and 0.22µm 142mm (Express PLUS Membrane, Millipore, Ref. GPWP14250) using a peristaltic pump with flow rate between 11-13psi (Masterflex, EW-77410-10). The 3µm polycarbonate filter was placed in a 5mL cryotube, flash-frozen and stored in liquid nitrogen. The 0.22µm filter was placed carefully in a 50mL tube to which 10mL of RNA later were added. This 0.22µm filter was stored at -20°C. The filtration procedure for RNA analysis was restricted in time to 10-15min maximum and during filtration the 10L carboys were kept in a cooler or on ice blocks. The exact amount of filtered water was carefully reported in the Log Sheet.

[BACT-FCM] Flow cytometry of bacteria. Four replicates of 1mL of seawater (pre-filtered by 200µm mesh) were collected from each depth. Samples were fixed immediately using cold 25%

glutaraldehyde (final concentration 0.125%), left in the dark for 10min at room temperature, flash-frozen and kept in liquid nitrogen on board and then stored at -80°C on land.

**[BACT-FISH]** Bacterial FISH analysis. For each of the triplicates, 10mL of seawater for surface and DCM and 90mL for OMZ (prefiltered by 20µm) were fixed in paraformaldehyde (1.5% final concentration), filtered onto a 0.2µm polycarbonate 25mm filter (Millipore) as described by (Amann *et al.* 2002). Samples were gently mixed and stored at 4°C between 1 and 24h before to proceed with filtrations. Once filtrations of samples were completed the samples were rinsed with ca 2mL of filtered MilliQ water and the water filtered. Upon completion the filters were dried at RT before to be stored in petrislides (Millipore) at -20°C.

**[BACT-CULT]** Bacterial culture attempt: The 1mL triplicates of 200µm pre-filtered seawater was dispensed into 2mL cryovials and 70µL of DMSO (7% final conc.) was added to each and the tube inverted to mix. Samples were flash frozen and stored in liquid nitrogen.



**Supplementary Figure 3:** Flow charts of the sampling procedures and preservation methods carried out for viruses, giruses and prokaryotes.

## Information associated with Supplementary Fig 4

Samples from **the 180-2,000µm size fraction** (Supplementary Fig.4a) were collected at sub-surface and DCM. For sub-surface (ca 5m depth), the 180µm Bongo Net equipped with a flow meter, temperature and depth recorder was towed horizontally for 15min. For DCM, the net was deployed down to the DCM layer, sat there for 5-10min, and then towed back to the surface. Depth was controlled using acoustic SCANMAR system. Nets were rinsed from the outside with filtered seawater. Water sample from the two cod-ends was sieved through 2mm and poured into an 8L bottle. The volume was adjusted to 3L with 0.1µm filtered seawater. Bottle was gently mixed and then split into three 1L fractions.

[**PROT-DNA 180-2000 & PROT-RNA 180-2000**] Duplicate samples were collected for RNA analysis. For each one, 0.5L of water was filtered through 2 x 47mm 10µm polycarbonate membranes. After filtration, the two membranes were recovered into one single 5mL cryotube and immediately stored in liquid nitrogen. The entire process did not exceed 15 minutes. The exact same procedure was repeated to prepare duplicate samples for DNA analysis.

The remaining 1L was divided to carry out fixation for morphological analysis of the communities.

[**PROT-OM-L 180-2000**] For lugol fixation, duplicates of 50mL water sample were poured into a 60mL dark-plastic bottles pre-filled with 1mL of lugol and stored at +4°C.

[**PROT-OM-F 180-2000**] Duplicates for formol fixation were carried out by pouring 45mL of sample into a 50mL Falcon tube pre-filled with 5mL of neutralized 37% formol and then stored at room temperature.

[**PROT-FISH-W 180-2000**] Duplicate samples were collected for liquid FISH by pouring 45mL of sample into a 50mL Falcon tube pre-filled with 5mL of neutralized 37% formol. The tube was incubated vertically for 1-3h in the dark in a coolbox, then the supernatant discarded and the tube filled up to 30mL with a mix of 50:50 vol:volPBS:EtOH. Samples were stored at -20°C.

[**PROT-ETOH 180-2000**] For ethanol fixation, duplicate samples were collected by filtering 250mL of sample through a 20µm sieve. The material retained was rinsed on the sieve and, recovered into a 50mL Falcon tube filled up to 40mL with 95% molecular grade EtOH. Samples were kept at +4°C for 12 hours and the whole process was repeated with fresh EtOH. Samples were finally stored at -20°C.

[PROT-LIVE 180-2000] Finally, 50mL of sample were kept in a 50mL Falcon tube for live observation onboard under a stereo-microscope and/or an inverted microscope in the dry Lab.

Samples from the **20-180µm size fraction** (Supplementary Fig.4b) were collected at sub-surface and DCM. For sub-surface (ca 5m depth), the 20µm Bongo net equipped with a flow meter, temperature and depth recorder, was towed horizontally for 15min. For DCM, the net was deployed down to the DCM layer, left sitting there for 5-10min, and then towed back to the surface. Depth was controlled using acoustic SCANMAR system. Nets were rinsed from the outside with filtered seawater. Water sample from the two cod-ends was sieved through 180µm and poured into an 8L bottle. The volume was adjusted to 3L with 0.1µm filtered seawater. Bottle was gently mixed and then split into three 1L fractions.

[PROT-DNA 20-180 & PROT-RNA 20-180] Duplicates were collected for RNA analysis. For each one, 0.5L of water was filtered through 2 x 47mm 10µm polycarbonate membranes. After filtration the two membranes were recovered into one single 5mL cryotube and immediately stored in liquid nitrogen. The entire process did not exceed 15 minutes. The exact same procedure was repeated to prepare duplicate samples for DNA analysis.

The remaining 1L was divided to carry out fixation for morphological analysis of the communities.

[PROT-OM-L 20-180] For the lugol fixation, duplicates of 50mL water sample were poured into 60mL dark-plastic bottles pre-filled with 1mL of lugol and stored at +4°C.

[PROT-OM-F 20-180] Duplicates for formol fixation were carried out by pouring 45mL of sample into a 50mL Falcon pre-filled with 5mL of neutralized 37% formol. Samples were stored at room temperature.

[PROT-FISH-W 20-180] Duplicate samples were collected for liquid FISH by pouring 45mL of sample into a 50mL Falcon tube pre-filled with 5mL of neutralized 37% formol. The tube was incubated vertically for 1-3h in the dark in a coolbox, then the supernatant discarded and the tube filled up to 30mL with a mix of 50:50 vol:volPBS:EtOH. Samples were then stored at -20°C.

[PROT-ETOH 20-180] For ethanol fixation, duplicate samples were collected by filtering 250mL of sample through a 20µm sieve. The material retained was rinsed on the sieve and recovered into a 50mL Falcon tube filled up to 40mL with 95% molecular grade EtOH. The samples were kept at +4°C for 12 hours and the whole process was repeated with fresh EtOH. Samples were then stored at -20°C.

[[PROT-TEM 20-180](#)] For TEM observation, duplicate samples were collected by pouring 48mL of water into a 50mL Falcon tube pre-filled with 2mL of EM-grade 25% glutaraldehyde (1% final conc.). Tubes were stored at +4°C.

[[PROT-HTM 20-180](#)] Duplicate samples for High Throughput Microscopy (HTM) were prepared by pouring 45mL of sample into a 50mL Falcon tube pre-filled with 5mL of 10% paraformaldehyde and 500µL of 25% glutaraldehyde EM grade. Tubes were incubated horizontally for at least 24h in the dark at +4°C and then stored vertically in the dark at +4°C.

[[PROT-LIVE 20-180](#) & [PROT-FCAM 20-180](#)] Finally, two times 50mL of sample were kept in 50mL Falcon tubes for onboard FlowCAM analysis and live observation under a stereo-microscope and/or an inverted microscope.

Samples from the **5-20µm size fraction** (Supplementary Fig.4c) were collected at sub-surface and DCM. For sub-surface (ca 5m depth), the 5µm Net equipped with a flow meter, temperature and depth recorder, was towed horizontally for 15min. For DCM, the net was deployed down to the DCM layer, left sitting there for 5-10min, and then towed back to the surface. Depth was controlled using acoustic SCANMAR system. Net was rinsed from the outside with filtered seawater. Water sample from the cod-end was sieved through 20µm and poured into an 8L bottle. The volume was adjusted to 3L with 0.1µm filtered seawater. Bottle was gently mixed and split into three 1L fractions.

The delicate 5µm nets cannot be deployed in harsh sea state conditions. Whenever necessary, the following sampling protocols were applied to collect samples of the 5-20µm size fraction. Seawater was pumped thanks to a peristaltic pump (down to 80m) and run through the Gravity Plankton Sieving System (GPSS). The GPSS consisted of a succession of 3 nets (mesh size of, successively, 20µm, 5µm, and 5µm again), equipped with small collectors and placed on top of each other to act as a sieving system. At the bottom of the sieving process, a plastic funnel was mounted to collect the <5µm sieved water (Supplementary Fig.4d). When ready to process the 5-20µm size fraction, the pump was turned off and the time recorded to calculate the volume of seawater sieved. Once empty, the 5µm GPSS nets were rinsed from inside with 0.1µm filtered seawater in order to push all the biological material into the two 650mL cod-ends. The water of both 5µm cod-ends was poured into the 8L bottle and volume adjusted to 3L with 0.1µm filtered seawater. Bottle was gently mixed and split into three 1L fractions.

In both cases, i.e. sampling with either net or GPSS, 2L were used for DNA/RNA analysis and 1L for morphological analysis.

[PROT-DNA 5-20 & PROT-RNA 5-20] Duplicate samples were collected for RNA analysis. For each one, 0.5L of water was filtered through 2 x 47mm 5µm polycarbonate membranes. After filtration the two membranes were recovered into one single 5mL cryotube and immediately stored in liquid nitrogen. The entire process did not exceed 15 minutes. The exact same procedure was repeated to prepare duplicate samples for DNA analysis.

[PROT-OM-L 5-20] For the lugol fixation, duplicates of 50mL seawater sample were poured into 60mL dark-plastic bottles pre-filled with 1mL of lugol and stored at +4°C.

[PROT-OM-F5-20] Duplicates for formol fixation were carried out by fixing 45mL of sample into 50mL Falcons pre-filled with 5mL of neutralized 37% formol and then stored at room temperature.

[PROT-FISH-W 5-20] Duplicate samples were collected for liquid FISH by pouring 45mL of sample into 50mL Falcon tubes pre-filled with 5mL of neutralized 37% formol. The tubes were incubated vertically for 1-3h in the dark in a coolbox, then the supernatant discarded and the tubes filled up to 30mL with a mix of 50:50 vol:volPBS:EtOH. The samples were then stored at -20°C.

[PROT-TEM 5-20] For TEM observation, duplicate samples were collected by pouring 48mL of water into 50mL Falcon tubes pre-filled with 2mL of EM-grade 25% glutaraldehyde (1% final conc.). Tubes were stored at +4°C.

[PROT-HTM 5-20] Duplicate samples for High Throughput Microscopy (HTM) were prepared by pouring 45mL of sample into 50mL Falcon tubes pre-filled with 5mL of 10% paraformaldehyde and 500µL of 25% glutaraldehyde EM grade. Tubes were incubated horizontally for at least 24h in the dark at +4°C and then stored vertically in the dark at +4°C.

[PROT-SCG 5-20] Duplicate samples for single cell genomics were collected by pouring 4mL of sample into 5mL cryotubes pre-filled with 600µL of Glycine Betaine. Samples were immediately stored in liquid nitrogen.

[PROT-FISH-D 5-20] Duplicate samples were prepared for dry FISH on membranes. For each replicate, 45mL of water sample were poured into 50mL Falcon tubes pre-filled with 5mL of 37% neutralized formaldehyde and incubated for 1-24h at +4°C. After fixation, the samples were filtered through 25mm 5µm polycarbonate membranes, subsequently rinsed with MilliQ water and gently vacuum dried. Membranes were air-dried for 5-10 minutes, placed in a petrislide, and stored at -20°C.

[[PROT-SEM 5-20](#)] Finally, duplicate samples for SEM observation were collected. For each one, 50mL of sampled water was filtered through a 47mm 5µm polycarbonate membrane. Membrane was vacuum dried, placed in petrislides and carefully air-dried prior to storage at room temperature.

Sampling of the **0.8-5µm size fraction** (Supplementary Fig.4d) was done at sub-surface and DCM using a large peristaltic pumping system (pump A40, Tech-pump). The hose system could reached down to 80m depth and was thoroughly rinsed during 10min by pumping seawater at maximum speed, Twice, 100L of water were then collected after pre-filtration through the GPSS (i.e. particle size <5µm, see above).

[[PROT-DNA 0.8-5](#) & [PROT-RNA 0.8-5](#)] For RNA analysis, 100L of this water were filtered through two 142mm 0.8µm polycarbonate membranes on top of a Dacron pad filter, mounted in parallel and connected to a large A20 peristaltic pump. The filtration process was restricted to 10-15min. Filters were recovered, each of them placed into a single 5mL cryotube (thus generating two pseudo-replicates) and immediately stored in liquid nitrogen. The exact same procedure was repeated to generate 2 samples for DNA analysis.

The filtrate (<0.8µm) from these procedures was collected and used for Girus filtrations (see Girus protocols above).

Two liters from GPSS pre-filtered seawater were collected for fixation for morphological analysis.

[[PROT-TEM 0.8-5](#)] For TEM observation, duplicate samples were collected by pouring 48mL of water into 50mL Falcon tubes pre-filled with 2mL of EM-grade 25% glutaraldehyde (1% final conc.). Tubes were stored at +4°C.

[[PROT-FISH-D 0.8-5](#)] Duplicate samples were prepared for dry FISH on membranes. For each replicate, 90mL of water sample were poured into a 250mL jar, pre-filled with 10mL of 37% neutralized formaldehyde and incubated for 1-24 h at +4°C. After fixation, the samples were filtered through 25mm 0.8µm polycarbonate membranes, subsequently rinsed with MilliQ water and gently vacuum dried. Membranes were air-dried for 5-10 minutes, placed in a petrislide, and stored at -20°C.

[[PROT-SEM 0.8-5](#)] Finally, duplicate samples for SEM observation were collected. For each one, 500mL of sampled water were filtered through 2 x 47mm 0.8µm polycarbonate membranes. Membranes were vacuum dried, placed in petrislides and carefully air-dried prior to storage at room temperature.

**Total seawater, i.e. not size-fractionated**, was also collected for protists at sub-surface and DCM (Supplementary Fig.4e).

[**PROT-DNA >0.8** & **PROT-RNA>0.8**] For RNA analysis, 100L of total seawater were collected with the on-deck peristaltic pump (A40) and immediately filtered through 2 x 142mm 0.8µm polycarbonate membranes on top of a Dacron pad filter, mounted in parallel and connected to a large A20 peristaltic pump. The filtration process was restricted to 10-15min. Filters were recovered, each of them placed into a single 5mL cryotube (thus generating two pseudo-replicates) and immediately stored in liquid nitrogen. The exact same procedure was repeated to generate samples for DNA analysis.

Two liters of total seawater were collected from the outlet tube of the on-deck peristaltic pump (A40) for morphological analysis.

[**PROT-FCM >0.8**] For Flow cytometry analysis, 1.5mL of water was dispensed in triplicates into 2mL cryotubes containing 150µL of 10% paraformaldehyde and 15µL EM-grade 25% glutaraldehyde. After incubation during 10-15min in the dark, samples were stored in liquid nitrogen.

[**PROT-SYRACO**] Duplicate samples were collected for Syraco analysis. For each one, 500mL of water was filtered through a 47mm 0.8µm nitrocellulose membrane. Importantly, membranes were vacuum-dried, placed in petrislides and allowed to air-dry at room temperature for hours before storage at room temperature.

[**PROT-SCG >0.8**] For both duplicate samples of single cell genomics, 4mL of water were poured into 5mL cryotubes pre-filled with 600µL of 48% Glycine Betaine and immediately stored in liquid nitrogen. Single cell genomic for bacteria were performed on cells isolated from that sample.

**For the mesopelagic depth** (Supplementary Fig.4f), deep water was sampled with the Niskin bottles mounted on the Rosette. The first cast (96L) was dedicated to RNA sampling, while the second cast was used for DNA and morphology sampling. Niskin bottles were emptied into a 100L carboy using the A20 peristaltic pump.

[**PROT-DNA >3** & **PROT-DNA 0.8-3** & **PROT-RNA >3** & **PROT-RNA >0.8-3**] For RNA analysis, 96L of water were filtered with the A20 peristaltic pump, through successively, a 142mm 3µm polycarbonate membrane, followed by a 142mm 0.8µm polycarbonate membrane, each one on top of a Dacron pad mesh. Filtration time was restricted to 10-15min. Membranes were recovered; each of them was placed into a 5mL cryotube and immediately stored in liquid nitrogen. As for epipelagic depths, the



filtrate (<0.8µm) was kept for Girus filtrations. The same procedure was used for DNA analysis, but with 80L of water.

The remaining 16L of deep water from the Niskin bottles were used for morphology analysis. Water was collected in a carboy and concentrated to 1.2L using a 0.8µm Tangential flow filtration (or by pumping extremely gently through a 0.8µm polycarbonate membrane). This concentrated seawater was used for morphological samples.

[**PROT-FISH-W >0.8**] Duplicate samples were collected for liquid FISH by pouring 45mL of sample into 50mL Falcon tubes pre-filled with 5mL of 37% neutralized formal. The tubes were incubated vertically for 1-3h in the dark in a coolbox, then the supernatant discarded and the tubes filled up to 30mL with a mix of 50:50 vol:vol PBS:EtOH. Samples were stored at -20°C.

[**PROT-TEM >0.8**] For TEM observation, duplicate samples were collected by pouring 48mL of water into 50mL Falcon tubes pre-filled with 2mL of EM-grade 25% glutaraldehyde (1% final concentration). Tubes were stored at +4°C.

[**PROT-HTM >0.8**] Duplicate samples were prepared for High Throughput Microscopy (HTM). For each sample, 45mL of sample was poured into a 50mL Falcon tube pre-filled with 5mL of 10% paraformaldehyde and 500µL of 25% glutaraldehyde EM grade. Tubes were incubated horizontally for at least 24h in the dark at +4°C and then stored vertically in the dark at +4°C.

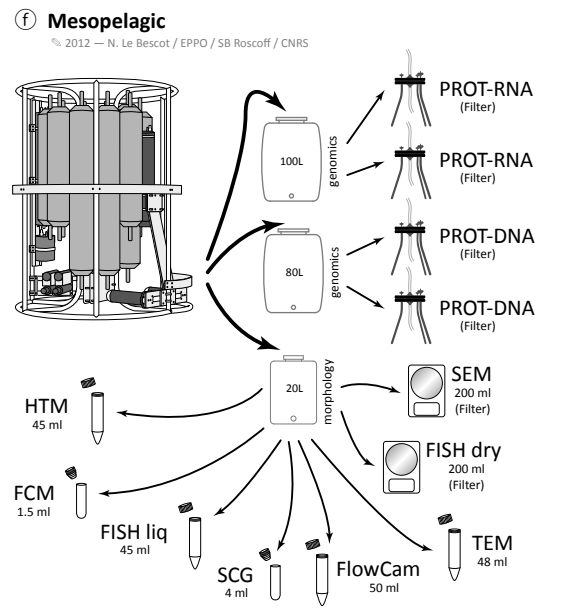
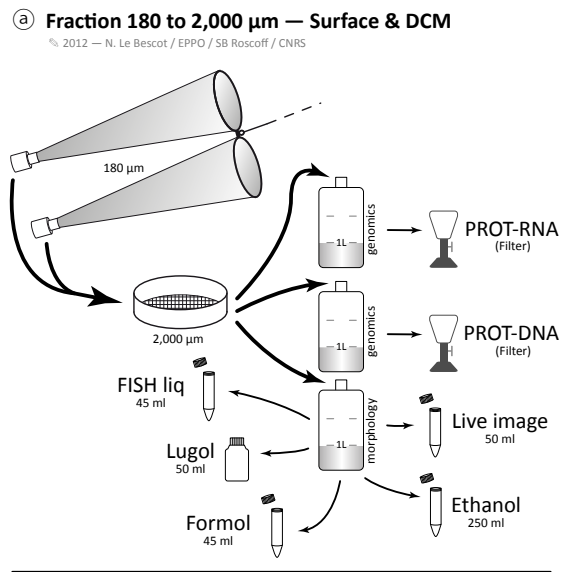
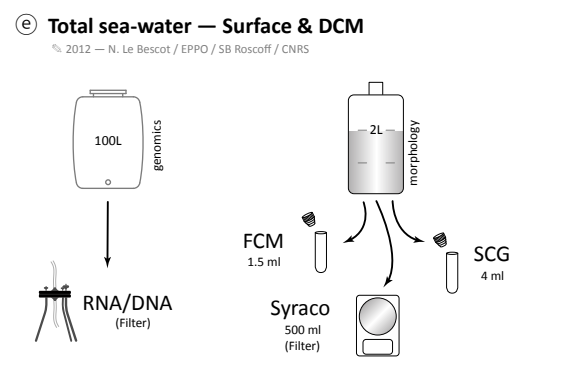
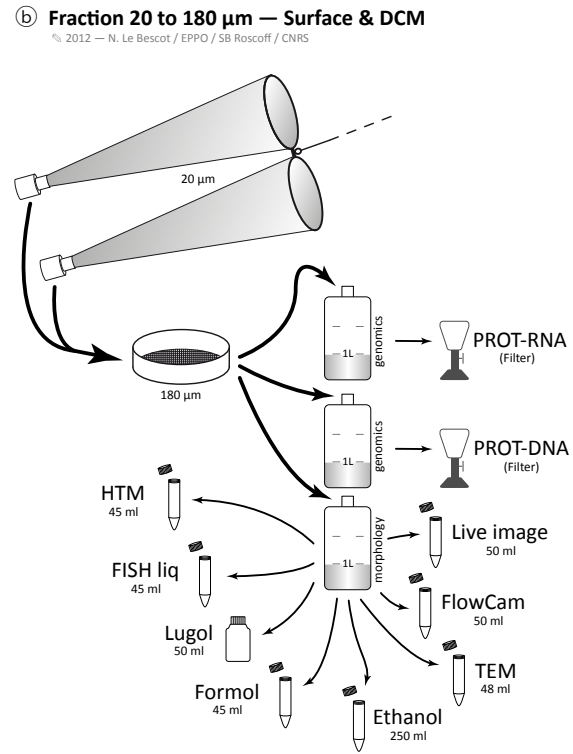
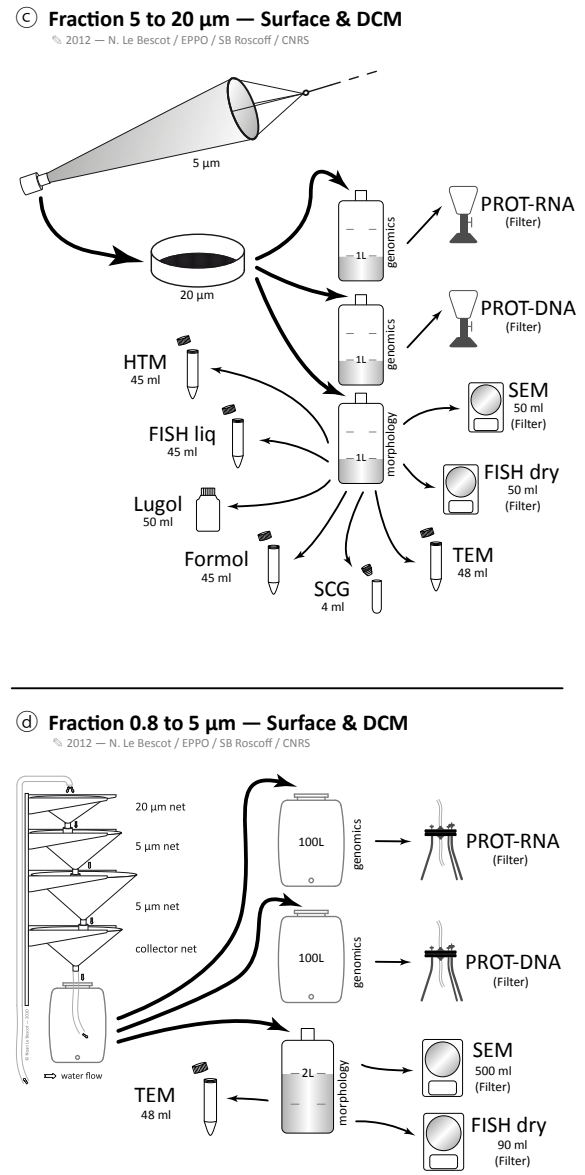
[**PROT-SCG >0.8**] Duplicate samples for single cell genomics were collected by pouring 4mL of sample into 5mL cryotubes pre-filled with 600µL of 48% Glycine Betaine. Samples were immediately stored in liquid nitrogen.

[**PROT-FISH-D >0.8**] Duplicate samples were prepared for dry FISH on membranes. For each replicate, 200mL of concentrated water was poured into a 250mL jar pre-filled with 20mL of 37% neutralized formaldehyde and incubated for 1-24h at +4°C. After fixation, water was filtered through 25mm 0.8µm polycarbonate membranes, subsequently rinsed with MilliQ water and gently vacuum dried. Membranes were air-dried for 5-10 minutes, placed in a petrislide, and stored at -20°C.

[**PROT-SEM >0.8**] Duplicate samples for SEM observation were collected. For each one, 200mL of sampled water was filtered through a 47mm 0.8µm polycarbonate membrane. Membrane was vacuum dried, placed in petrislides and air-dried prior to storage at room temperature.

[PROT-FCM >0.8] For Flow cytometry analysis, 1.5mL of water was dispensed in triplicates into 2mL cryotubes containing 150µL of 10% paraformaldehyde and 15µL EM-grade 25% glutaraldehyde. After incubation during 10-15min in the dark, samples were stored in liquid nitrogen.

[PROT FCAM >0.8] Finally, 50mL of water were collected in a Falcon tube for onboard FlowCAM analysis.



Supplementary Figure 4: Flow charts of sampling procedures and preservation methods applied for the protist size fractions.

## Information associated with Supplementary Figure 5

Large planktonic organisms were collected using a series of nets of various mesh sizes, at various depths, at day and night time, and samples were preserved following three distinct protocols. Depth was controlled using acoustic SCANMAR system. Before to proceed with sample harvest from the cod-ends, all nets were rinsed from the outside with raw seawater upon recovery. After each use, nets, cod-ends, and sieves were rinsed with fresh water and checked for holes.

[METAZ-TMOR >50 & METAZ-TGEN >50] The 50 $\mu$ m nets were towed vertically from 100m depth to the surface at a speed of 0.8m/s. At night, 2 net casts were performed; one was preserved in formol and the other in Ethanol. During day-time, one single tow was performed and the material collected preserved in formol.

[METAZ-TMOR >200] The WP11 200 $\mu$ m nets were deployed vertically, from 100m depth to the surface, one time at night and one time during the day. The material collected was preserved in formol.

[METAZ-TMOR>300 & METAZ-TGEN >300 & METAZ-MGEN >300] The 300 $\mu$ m bongo nets were deployed one time, both at night and day time. They were towed in oblique from 500m depth to the surface. The content of one cod-end was preserved in formol and the other cod-end split one part was preserved in ethanol while the other part was preserved in RNAlater buffer.

[METAZ-TMOR >680] The Regent 680 $\mu$ m net was also towed in oblique, from 500m depth to the surface, one time at both night and day time. Samples were preserved in formol.

[METAZ-TMOR >500] Whenever the sea state allowed, a multi-net equipped with 5 closing nets of 500 $\mu$ m mesh size was deployed from 1,000m deep to the surface. Deployments were performed at night and day-time and nets closed according to water column features obtained from the CTD profiles. Samples collected in each of the 5 nets were preserved in formol.

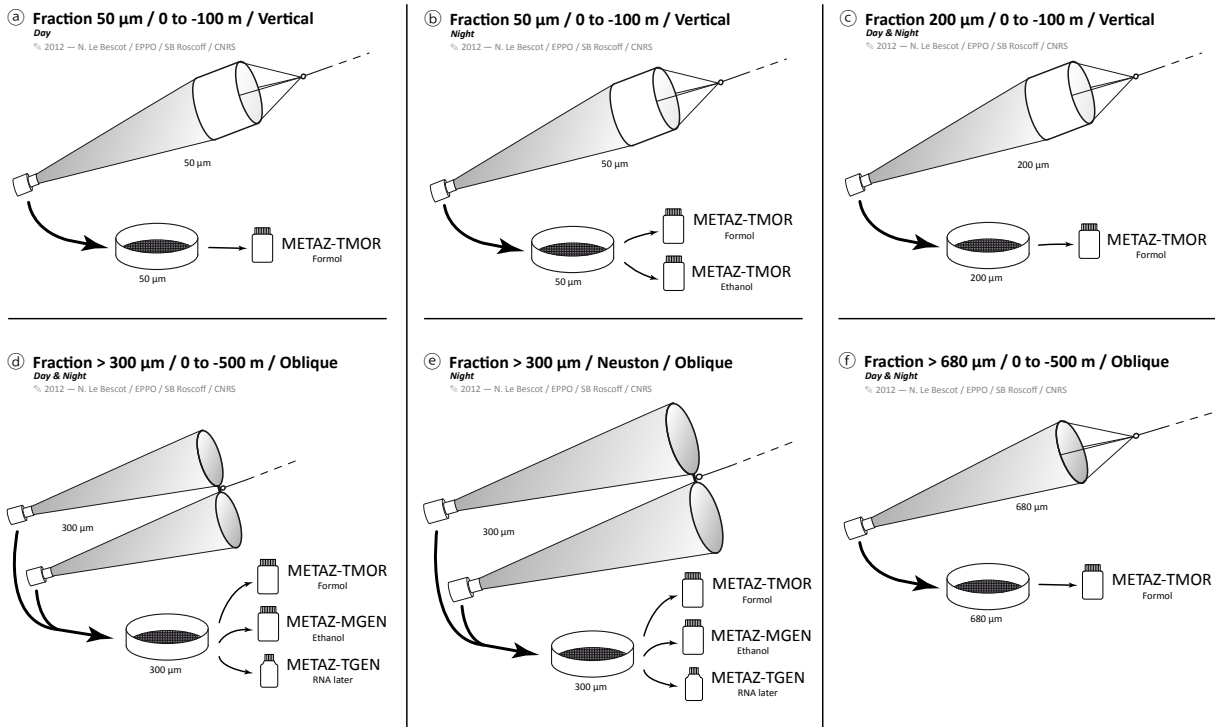
The procedure for the three preservation methods mentioned above is detailed.

[For the Ethanol fixation (-TGEN)] samples recovered from the cod end were concentrated gently by sieving onto a mesh size equal to the one of the net used and the samples were transferred into a 250mL jar using the ethanol squeeze bottle. The jar was filled 2/3 with ethanol. Samples were first stored at 4°C for 24 hours and the ethanol changed before longer-term storage at -20°C.

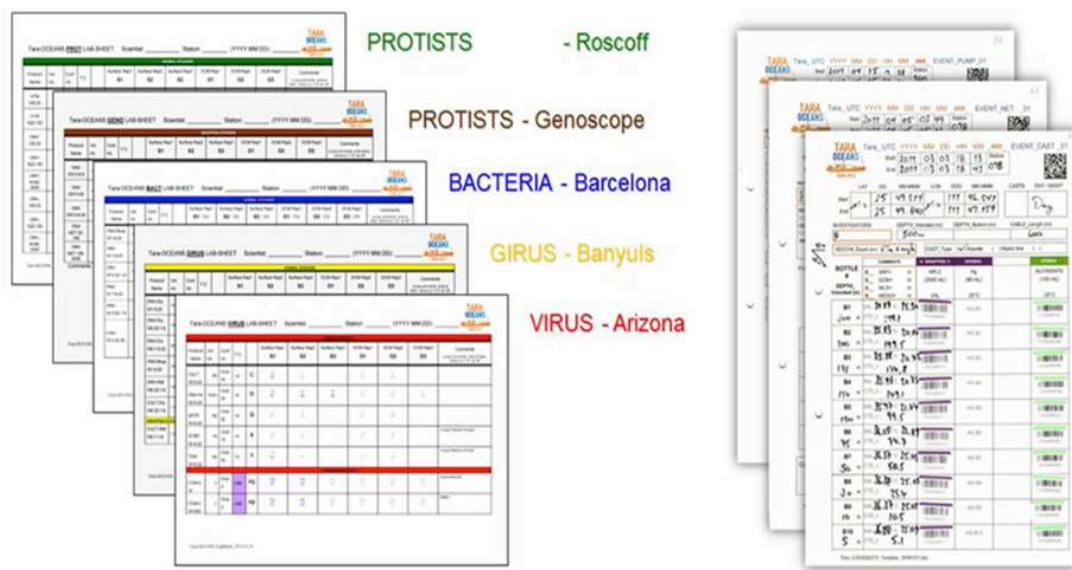
[For formol fixation (-TMOR)] samples collected were concentrated by sieving and subsequently poured into a 250mL jar pre-filled with 50mL of Borax buffer (50gr of borax into 2.5L of MilliQ water)

and into which 10mL of 37% formol was added. The jar was then filled with 0.1µm SeaWater up to 250mL and stored at RT.

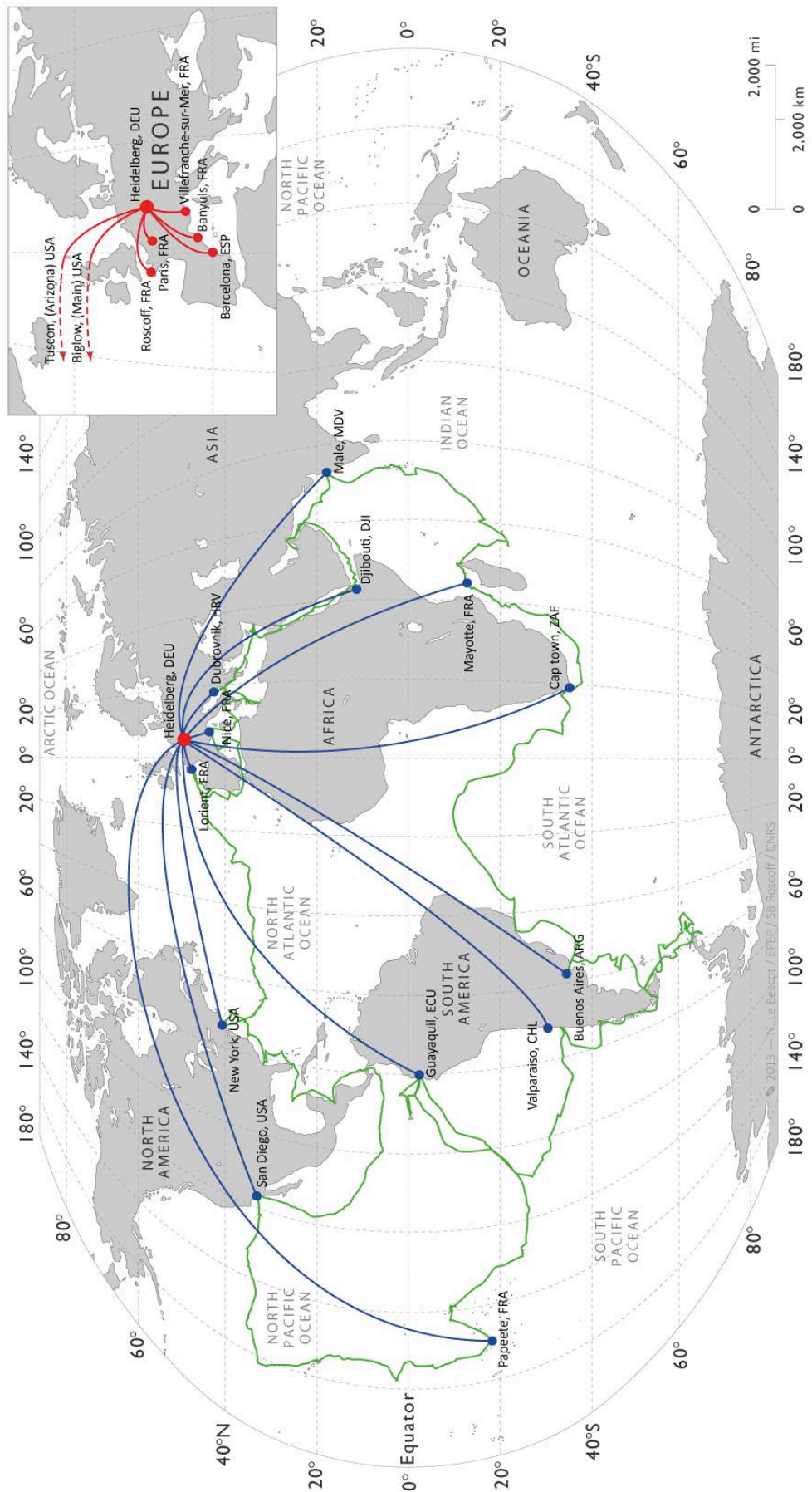
[Finally RNAlater samples (-MGEN)] were collected for metagenomics community analysis. 60mL plastic bottles were loaded with 2 spoons of concentrated sample and filled up to 2/3 of the bottle with RNAlater before storing at -20°C.



**Supplementary Figure 5:** Flow charts of the sampling procedures and preservation methods carried out with various plankton nets deployed to collect large planktonic organisms.



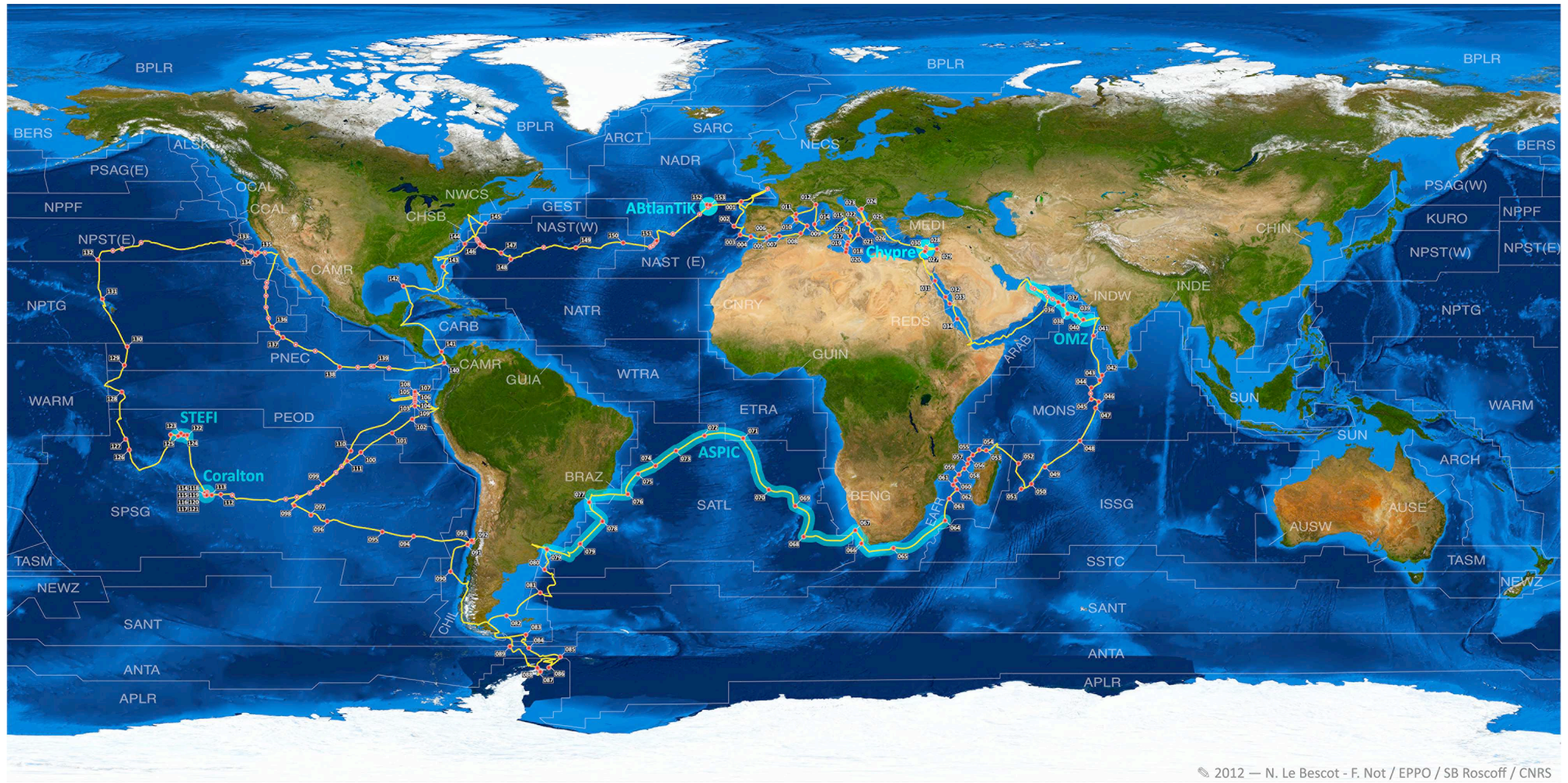
**Supplementary Figure 6:** Log sheets used onboard to record sampling events and to track individual samples. The color code corresponds to dispatch locations (laboratories) for respective samples.



**Supplementary Figure 7:** Map of shipping points and distribution of samples collected during the Tara Oceans expedition. Blue corresponds to samples from the boat to the dispatch center in Heidelberg (Germany) and red from Heidelberg to the laboratories.



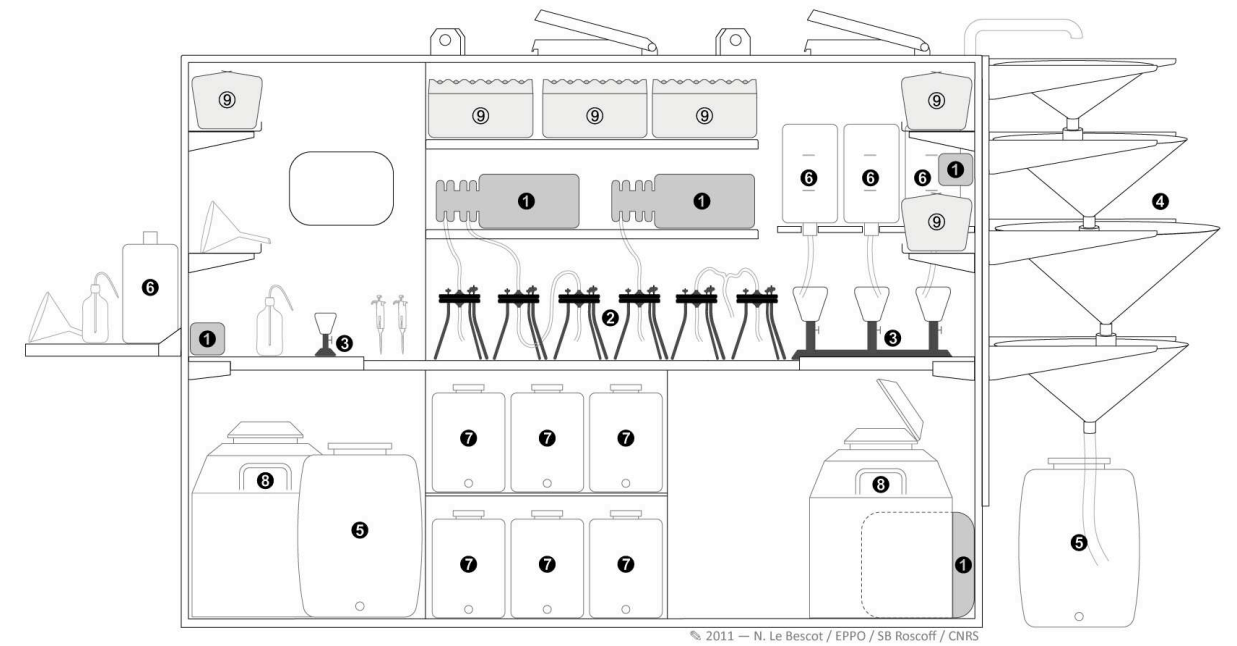




**Supplementary Figure 8:** Detailed numbering of sampling locations (zoom in to check actual station number). See also supplementary taracoceans-project2009-2012.kml file.



Ⓐ

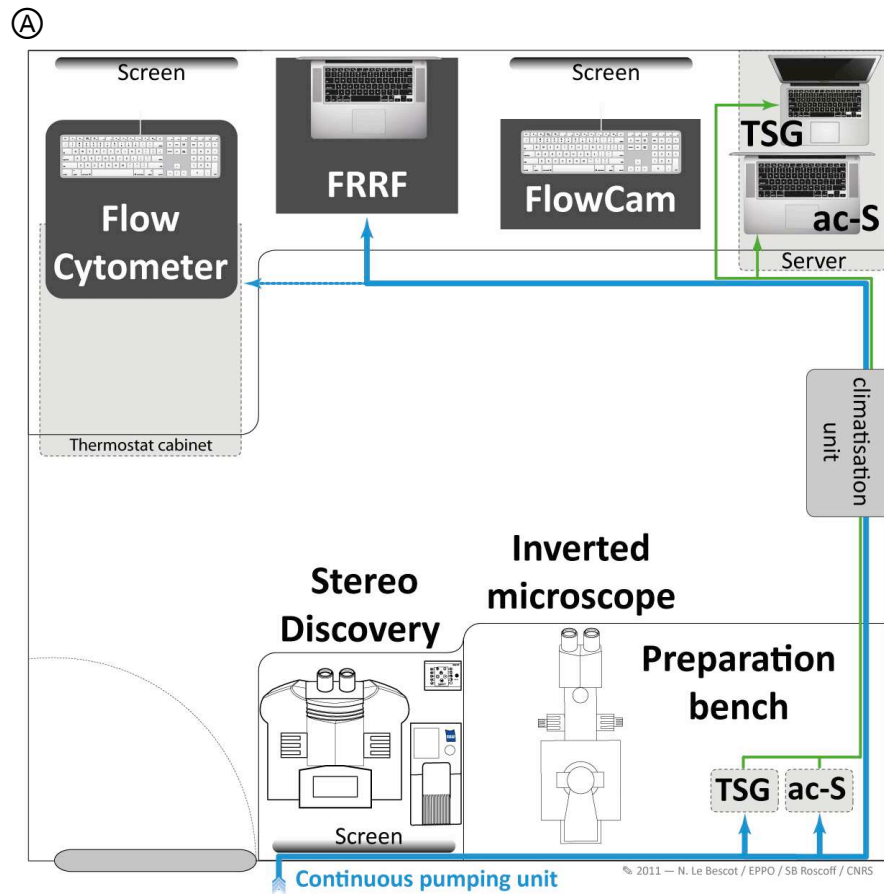


© 2011 — N. Le Bescot / EPPD / SB Roscoff / CNRS

Ⓑ



**Supplementary Figure 9:** Ⓐ Detail of the Wet Lab installed on board during the Tara Oceans expedition. (1) Pumps, (2) 142 mm filtration holder, (3) 47 and 25 mm filtration holder, (4) Gravity Plankton Sieving System nets, (5) 100 L tanks, (6) 2 L bottles for pigments, (7) 20 L tanks for prokaryotes and viruses filtrations, (8) liquid nitrogen dewars, (9) storage boxes. Ⓑ Picture showing inside the wet lab (© D. Sauveur/Fonds Tara 2010)

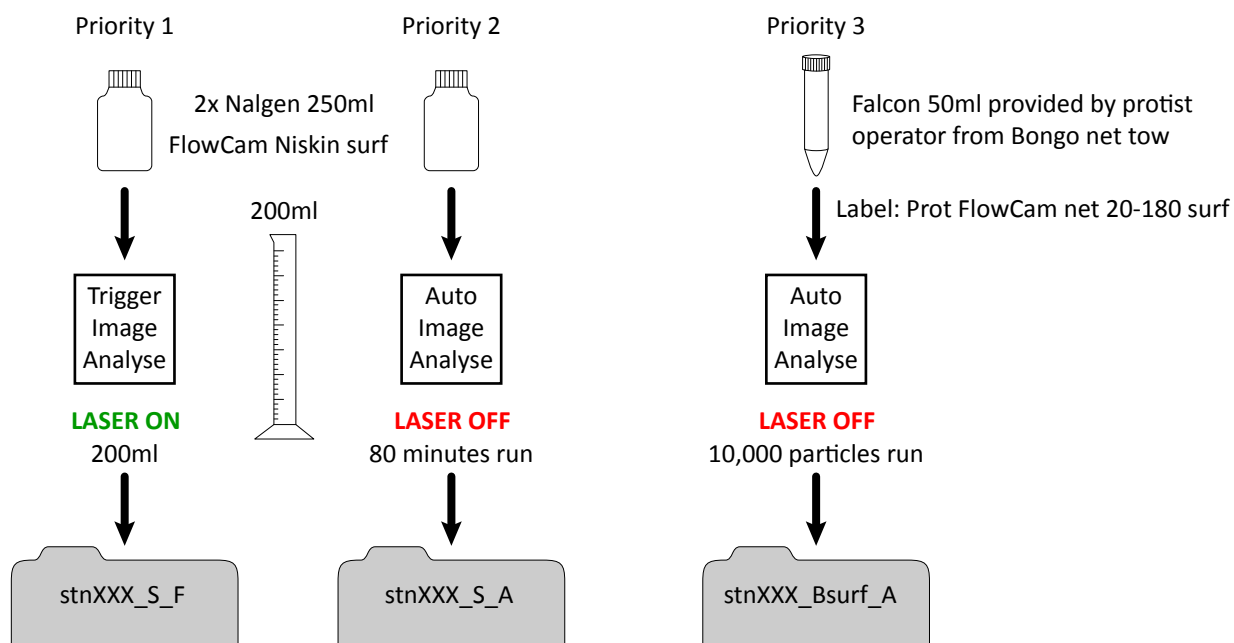


(B)

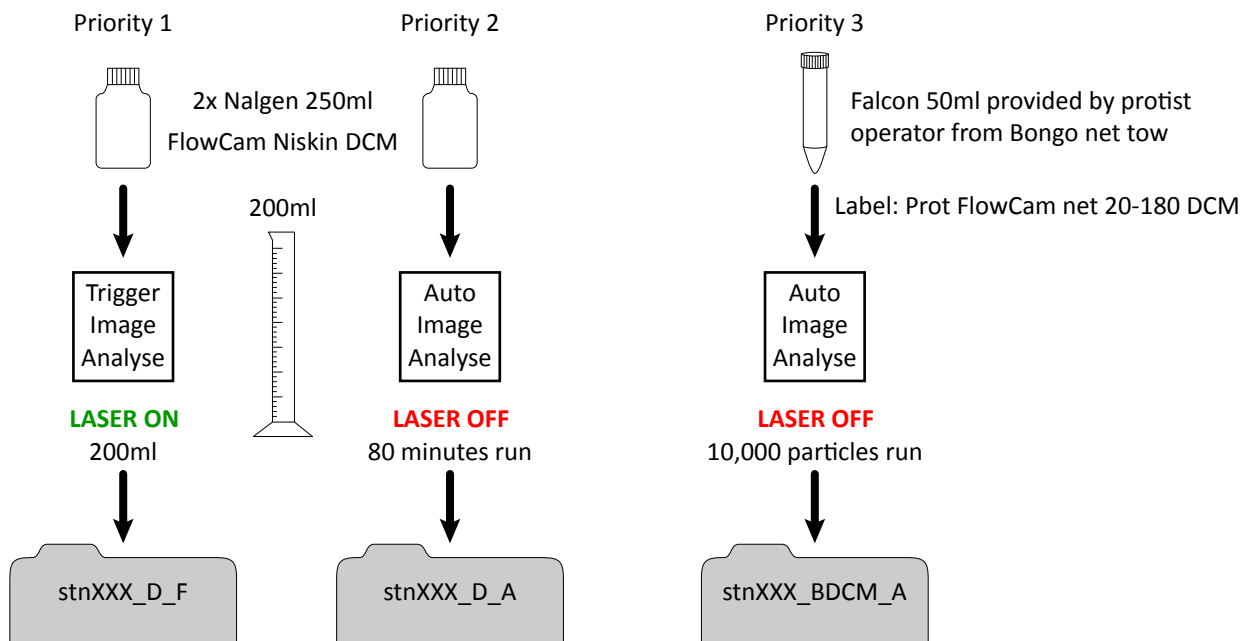


**Supplementary Figure 10:** (A) Detail of the Dry lab instruments and set up. TSG stands for Thermo-salino-graph, ac-S is the spectrophotometer (B) Picture inside the dry lab (© D. Sauveur/Fonds Tara 2010).

### Surface water sample



### DCM water sample



**Supplementary Figure 11:** Rationale for FlowCAM analytical procedures performed onboard Tara.

**Supplementary Table 1:** Tara Oceans consortium coordinators role and contact information.

| Person                                 | Role  | Institute (19)                         | Country (8)        | Email contact                |
|--|---|--|--------------------|------------------------------|
| Coordination, Logistics & Technologies |   |  |                    |                              |
| Eric Karsenti                          | Overall coordination of the expedition      | EMBL, Heidelberg & CNRS                | Germany/<br>France | karsenti@embl.de             |
| Etienne Bourgois                       | Tara-Oceans Co-director                     | Agnes b, Paris                         | France             | etienne@taraexpeditions.org  |
| Stefanie Kandels-Lewis                 | Science logistics                           | EMBL, Heidelberg                       | Germany            | kandels@embl-heidelberg.de   |
| Romain Trouble                         | Tara logistics                              | Agnes b, Paris                         | France             | romain@taraexpeditions.org   |
| Didier Velayoudon                      | Scientific consortium logistics             | dVIPC, Paris                           | France             | didier.velayoudon@dvipc.com  |
| Sabrina Speich                         | Remote sensing                              | UBO & CNRS, Brest                      | France             | speich@univ-brest.fr         |
| Gaby Gorsky                            | Operational oceanography                    | UPMC & CNRS, LOV, Villefranche/Mer     | France             | gabriel.gorsky@obs-vlfr.fr   |
| Fabrice Not                            | Operational biology                         | CNRS & UPMC, SB Roscoff                | France             | not@sb-roscoff.fr            |
| Emmanuel Reynaud                       | Onboard imaging platform                    | UC Dublin                              | Ireland            | emmanuel.reynaud@ucd.ie      |
| Patrick Wincker                        | Sequencing & Omics data                     | CEA, Genoscope, Evry                   | France             | pwincker@genoscope.cns.fr    |
| Mike Sieracki                          | FlowCam & Single Amplified Genomes          | Bigelow Oceans Science, Maine          | USA                | msieracki@bigelow.org        |
| Christian Sardet                       | Scientific outreach and media               | UPMC & CNRS, LOV, Villefranche/Mer     | France             | christian.sardet@obs-vlfr.fr |
| Eloise Fontaine                        | Tara communication and media                | Agnes b, Paris, France                 | France             | eloise@taraexpeditions.org   |
| OceansBiota                            |   |  |                    |                              |
| Lars Stemmann                          | Zooplankton                                 | UPMC & CNRS, LOV, Villefranche/Mer     | France             | stemmann@obs-vlfr.fr         |
| Colomban de Vargas                     | Protists                                    | CNRS & UPMC, SB Roscoff                | France             | vargas@sb-roscoff.fr         |
| Silvia Acinas                          | Prokaryotes                                 | CSIC Ciencas del Mar, Barcelona        | Spain              | sacinas@icm.csic.es          |
| Hiroyuki Ogata                         | Eukaryotic DNA viruses (giruses)            | Tokyo Institute of Technology, ACLS    | Japan              | ogata@acls.titech.ac.jp      |
| Matt Sullivan                          | Phages                                      | University of Arizona, Tucson          | USA                | mbsulli@email.arizona.edu    |
| Francesca Benzoni                      | Coral reefs                                 | Milano University                      | Italy              | francesca.benzoni@unimib.it  |
| Data management & analyses             |   |  |                    |                              |
| Stéphane Pesant                        | Data management and standardization         | Marum & Pangea, Bremen                 | Germany            | spesant@marum.de             |
| Pascal Hingamps                        | Bioinformatics data management and analyses | Aix Marseille University, CNRS, IGS    | France             | pascal.hingamp@univmed.fr    |
| Daniele Iudicone                       | Physical oceanography                       | Stazione Zoologica Anton Dohrn, Naples | Italy              | iudicone@szn.it              |
| Chris Bowler                           | Functional Omics & diatoms                  | CNRS IBENS, Paris                      | France             | cbowler@biologie.ens.fr      |
| Jeroen Raes                            | Systems biology and bioinformatics          | VIB, Brussels                          | Belgium            | jeroen.raes@vib-vub.be       |
| Peer Bork                              | Systems biology and bioinformatics          | EMBL, Heidelberg                       | Germany            | jeroen.raes@vib-vub.be       |
| Mick Follows                           | Global ocean biogeochemical modelling       | MIT, Boston                            | USA                | mick@ocean.mit.edu           |

**Supplementary Table 2:** Recorded density and estimated diversity for organisms of each planktonic size fraction. Data for both coastal and open ocean ecosystems are considered.

| Organisms          | Cell size    | Density<br># ind. / volume              | Diversity<br># sp / volume   | References                          |
|--------------------|--------------|---|--|-------------------------------------|
| <b>Viruses</b>     |              |   |  |                                     |
|                    | 0,02 - 0,2µm | $10^9 - 200 \times 10^9 / L$            |  | (Brussaard <i>et al.</i> 2010)      |
|                    | 0,02 - 0,2µm | $10 \times 10^9 / L$                    |  | (Suttle 2007)                       |
|                    | 0,02 - 0,2µm | $60 \times 10^6 - 260 \times 10^6 / L$  |  | (Weinbauer <i>et al.</i> 2010)      |
|                    | 0,02 - 0,2µm | $12 \times 10^6 - 350 \times 10^6 / L$  |  | (Weinbauer <i>et al.</i> 2010)      |
|                    |              |   | $10^5 - 10^6 / L$  | (Angly <i>et al.</i> 2006)          |
| <b>Prokaryotes</b> |              |   |  |                                     |
|                    |              | $3 \times 10^9 / L$                     |  | (Gilbert 2001)                      |
|                    | < 1µm        | $4 \times 10^9 / L$                     |  | (Buitenhuis & Li 2012)              |
|                    | < 2µm        | $10^9 - 5 \times 10^9 / L$              |  | (Ducklow 2000)                      |
|                    | < 2µm        | $500 \times 10^6 / L$                   |  | (Ducklow 2000)                      |
|                    | < 2µm        | $200 \times 10^6 - 900 \times 10^6 / L$ |  | (Hyun & Kim 2003)                   |
|                    | < 2µm        | $200 \times 10^6 - 600 \times 10^6 / L$ |  | (Hyun & Kim 2003)                   |
|                    | < 2µm        | $500 \times 10^6 - 7.5 \times 10^9 / L$ |  | (Li 1998)                           |
|                    | < 2µm        | $225 \times 10^6 - 300 \times 10^6 / L$ |  | (Kirchman & Mitchell 2008)          |
|                    | < 2µm        | $200 \times 10^6 - 80 \times 10^6 / L$  |  | (Kirchman & Mitchell 2008)          |
|                    |              | $10^9 / L$                              |  | (Scanlan <i>et al.</i> 2009)        |
|                    |              |   | $10^7 / L$   | (Pedrós-Alió 2006)                  |
|                    |              |   | $2 \times 10^4 / L$  | (Pedrós-Alió 2012)                  |
|                    |              |   | $2 \times 10^4 / L$  | (Amaral-Zettler <i>et al.</i> 2011) |
| <b>Protists</b>    |              |   |  |                                     |
|                    |              | $3 \times 10^4 / L$                     |  | (Gilbert 2001)                      |
|                    | 0,8 - 2µm    | $183 \times 10^3 / L$                   |  | (Raghukumar 2002)                   |
|                    | 0,6 - 10µm   | $10^5 - 4 \times 10^6 / L$              |  | (Signorini & McClain 2009)          |
|                    | 0,2 - 2µm    | $20 \times 10^6 - 10^9 / L$             |  | (Evans <i>et al.</i> 2003)          |
|                    | 0,8 - 2µm    | $5 \times 10^5 - 3 \times 10^9 / L$     |  | (Countway & Caron 2006)             |
|                    | 0,2 - 3µm    | $10^6 - 20 \times 10^6 / L$             |  | (Worden 2008)                       |
|                    | 8 - 10µm     | $10 \times 10^6 - 300 \times 10^6 / L$  |  | (Holligan <i>et al.</i> 1993)       |
|                    | < 20µm       | $108 \times 10^3 / L$                   |  | (Gao <i>et al.</i> 2001)            |
|                    | > 20µm       | $260 \times 10^3 / L$                   |  | (Gao <i>et al.</i> 2001)            |
|                    | 10 - 100µm   | $10 \times 10^6 / L$                    |  | (Taylor FJR, Pahlinger U., 1987)    |
|                    | 5 - 2,000µm  | $10 \times 10^9 / L$                    |  | (Taylor FJR, Pahlinger U., 1987)    |
|                    | 18 - 500µm   | $20 \times 10^3 - 500 \times 10^3 / L$  |  | (Smalley & Coats 2002)              |
|                    |              |   | $3 \times 10^3 / L$  | (Amaral-Zettler <i>et al.</i> 2011) |
|                    |              |   | $1.4 \times 10^6 - 16 \times 10^6$ (all protists, not only marine)   | (Simpson <i>et al.</i> 2007)        |
|                    | 1 - 10,000µm |   |  | (Christen <i>et al.</i> 2010)       |
|                    |              |   | have suggested that the number of protist species may easily exceed one million  | (Edgcomb <i>et al.</i> 2011)        |
| <b>Metazoans</b>   |              |   |  |                                     |
|                    |              | 0,02 - 80/L                             | Coastal Mediterranean sea  | (Gilbert 2001)                      |
|                    | > 100µm      | Few – 250/L                             | Copepods, North Atlantic, coastal sea  | (Turner 2004)                       |
|                    | > 200µm      |   | Have reported using 8 datasets covering a wide range of latitude that copepod taxonomic diversity (genus level) range from 3 to 40 | (Ibañez <i>et al.</i> 2009)         |
|                    | > 500µm      | Few-80/L                                | Copepods, North Atlantic, open sea   | (Gallienne & Robins 2001)           |
|                    | > 100µm      | Few-150/L                               | Copepods, North Atlantic, open sea   | (Gallienne & Robins 2001)           |
|                    | > 10,000µm   | 0.00005-0.0007/L                        |  | (Stemmann <i>et al.</i> 2008)       |
|                    | > 2,000µm    | 0,0006 - 5/L                            |  | (Moriarty & Buitenhuis 2012)        |
|                    |              |   |  | (Stemmann <i>et al.</i> 2008)       |
|                    |              |   | $7 \times 10^3$ described species  | (Bucklin <i>et al.</i> 2011)        |



## References

- Adl SM, Leander BS, Simpson AGB *et al.* (2007) Diversity, nomenclature, and taxonomy of protists. *Systematic Biology*, **56**, 684.
- Amann R, Pernthaler A, Pernthaler J (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology*, **68**, 3094–3101.
- Amaral-Zettler L, Zettler ER, Theroux SM *et al.* (2011) Microbial community structure across the tree of life in the extreme Rio Tinto. **5**, 42–50.
- Angly FE, Felts B, Breitbart M *et al.* (2006) The marine viromes of four oceanic regions. *PLoS Biology*, **4**, e368.
- Behnke A, Engel M, Christen R *et al.* (2010) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 1–10.
- Boss E, Twardowski MS, Herring S (2001) Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. *Applied Optics*, **40**, 4885–4893.
- Brussaard CP, Weinbauer MG, Payet JP, Winter C (2010) Quantification of aquatic viruses by flow cytometry. *American Society of Limnology and Oceanography*, **1**, 102–109.
- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA Barcoding of Marine Metazoa. *Annual Review of Marine Science*, **3**, 471–508.
- Buitenhuis ET, Li WKW (2012) Bacterial biomass distribution in the global ocean. *Earth System Science Data*, **5**, 301–305.
- Cetinić I, Perry MJ, Briggs NT *et al.* (2012) Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research*, **117**, C06028.
- Christen R, Behnke A, Engel M *et al.* (2010) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.
- Countway PD, Caron DA (2006) Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Applied and Environmental Microbiology*, **72**, 2496.
- Ducklow HW (2000) Bacterial production and biomass in the oceans. In *Microbial Ecology of the Oceans*, pp. 85–120.
- Edgcomb V, Orsi W, Bunge J *et al.* (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME Journal*, **5**, 1344–1356.
- Evans C, Archer SD, Jacquet S, Wilson WH (2003) Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a *Micromonas* spp. population. *Aquatic Microbial Ecology*, **30**, 207–219.
- Gallienne CP, Robins DB (2001) Is *Oithona* the most important copepod in the world's oceans? *Journal of Plankton Research*, **23**, 1421–1432.
- Gao YH, Chen CP, Li Y (2001) Marine nanoplanktonic diatoms from the coastal waters of Hong Kong. *Proceedings of an International Workshop Reunion Conference*.
- Gilabert J (2001) Short-term variability of the planktonic size structure in a Mediterranean coastal lagoon. *Journal of Plankton Research*, **23**, 219–226.
- Holligan PM, Fernández E, Aiken J *et al.* (1993) A biogeochemical study of the coccolithophore, *Emiliana huxleyi*, in the North Atlantic. *Geophysical Research Letters*, **7**, 879–900.
- Hyun J-H, Kim K-H (2003) Bacterial abundance and production during the unique spring phytoplankton

- bloom in the central Yellow Sea. *Marine Ecology Progress Series*, **252**, 77–88.
- Ibañez F, Rombouts I, Beaugrand G *et al.* (2009) Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings. Biological sciences / The Royal Society*, **276**, 3053–3062.
- Kirchman DL, Mitchell R (Eds.) (2008) *Microbial ecology of the oceans*. Wiley Blackwell.
- Li WKW (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology And Oceanography*, **43**, 1746–1753.
- Moriarty R, Buitenhuis ET (2012) Distribution of known macrozooplankton abundance and biomass in the global ocean. *Earth System Science Data*, **5**, 187–220.
- Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? *Trends in Microbiology*, **14**, 257–263.
- Pedrós-Alió C (2012) The Rare Bacterial Biosphere. *Annual Review of Marine Science*, **4**, 449–466.
- Pernthaler A, Pernthaler J, Amann R (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology*, **68**, 3094–3101.
- Raghukumar S (2002) Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *European Journal of Protistology*, **38**, 127–145.
- Rombouts I, Beaugrand G, Ibanez F *et al.* (2009) Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings. Biological sciences / The Royal Society*, **276**, 3053–3062.
- Scanlan DJ, Ostrowski M, Mazard S *et al.* (2009) Ecological Genomics of Marine Picocyanobacteria. *Microbiology and Molecular Biology Reviews*, **73**, 249–299.
- Signorini SR, McClain CR (2009) Environmental factors controlling the Barents Sea spring-summer phytoplankton blooms. *Geophysical Research Letters*, **36**, L10604.
- Simpson AGB, Adl SM, Leander BS *et al.* (2007) Diversity, nomenclature, and taxonomy of protists. *Systematic Biology*, **56**, 684.
- Slade WH, Boss E, Dall'Olmo G *et al.* (2010) Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *American Meteorological Society*, **27**, 1733–1746.
- Smalley GW, Coats DW (2002) Ecology of the Red-Tide Dinoflagellate *Ceratium furca*: Distribution, Mixotrophy, and Grazing Impact on Ciliate Populations of Chesapeake Bay. *Journal of Eukaryotic Microbiology*, **49**, 63–73.
- Stemmann L, Youngbluth M, Robert K *et al.* (2008) Global zoogeography of fragile macrozooplankton in the upper 100–1000 m inferred from the underwater video profiler. *ICES Journal of Marine Science: Journal du Conseil*, **65**, 433–442.
- Suttle CA (2007) Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, **5**, 801–812.
- Turner JT (2004) The importance of small planktonic copepods and their roles in pelagic marine food webs. *Zool Stud*, **43**, 255–266.
- Weinbauer MG, Rowe JM, Wilhelm SW (2010) Determining rates of virus production in aquatic systems by the virus reduction approach. *American Society of Limnology and Oceanography*, **1**, 1–8.
- Worden AZ (2008) Ecology and diversity of picoeukaryotes. In: *Microbial ecology of the oceans* (eds Kirchman DL, Mitchell R), pp. 159–205. Wiley Blackwell.
- Taylor FJR, Pahlinger U (1987) *Ecology of dinoflagellates*. In: Taylor FJR (ed) *The biology of dinoflagellates*. *Botanical Monographs*, vol **21**. Blackwell Science Publications, Boston, pp 399–529.

Published in *Nucleic Acids Research*, 2013, Vol. 41

## The Protist Ribosomal Reference database (PR<sup>2</sup>): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy

Laure Guillou<sup>1,2</sup>, Dipankar Bachar<sup>3,4</sup>, Stéphane Audic<sup>1,2</sup>, David Bass<sup>5</sup>, Cédric Berney<sup>5</sup>, Lucie Bittner<sup>1,2</sup>, Christophe Boutte<sup>1,2</sup>, Gaétan Burgaud<sup>6</sup>, Colombar de Vargas<sup>1,2</sup>, Johan Decelle<sup>1,2</sup>, Javier del Campo<sup>7</sup>, John R. Dolan<sup>8</sup>, Micah Dunthorn<sup>9</sup>, Bente Edvardsen<sup>10</sup>, Maria Holzmann<sup>11</sup>, Wiebe H.C.F. Kooistra<sup>12</sup>, Enrique Lara<sup>13</sup>, **Noan Le Bescot**<sup>1,2</sup>, Ramiro Logares<sup>7</sup>, Frédéric Mahé<sup>1,2</sup>, Ramon Massana<sup>7</sup>, Marina Montresor<sup>12</sup>, Raphael Morard<sup>1,2</sup>, Fabrice Not<sup>1,2</sup>, Jan Pawlowski<sup>11</sup>, Ian Probert<sup>14,15</sup>, Anne-Laure Sauvadet<sup>1,2</sup>, Raffaele Siano<sup>16</sup>, Thorsten Stoeck<sup>9</sup>, Daniel Vaultot<sup>1,2</sup>, Pascal Zimmermann<sup>17</sup> and Richard Christen<sup>3,4</sup>

<sup>1</sup>CNRS, UMR 7144, Adaptation et Diversité en Milieu Marin, 29682 Roscoff, France, <sup>2</sup>UPMC Université Paris 06, UMR 7144, Station Biologique de Roscoff, 29682 Roscoff, France, <sup>3</sup>CNRS, UMR 7138, Systématique Adaptation Evolution, Parc Valrose, BP71. F06108 Nice cedex 02, France, <sup>4</sup>UMR 7138, Université de Nice-Sophia Antipolis, Systématique Adaptation Evolution, Parc Valrose, BP71. F06108 Nice cedex 02, France, <sup>5</sup>Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK, <sup>6</sup>Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (EA3882), ESMISAB, Technopôle Brest-Iroise, 29280 Plouzané, France, <sup>7</sup>Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Barcelona, Catalonia, Spain, <sup>8</sup>Laboratoire d'Océanographie de Villefranche, Marine Microbial Ecology, UPMC Université Paris 06 et CNRS, UMR7093, Station Zoologique, BP28, 06230 Villefranche-sur-Mer, France, <sup>9</sup>Department of Ecology, University of Kaiserslautern, 67663 Kaiserslautern, Germany, <sup>10</sup>Department of Biology, University of Oslo, Marine Biology, NO-0316 Oslo, Norway, <sup>11</sup>Department of Genetics and Evolution, University of Geneva, Switzerland, <sup>12</sup>Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy, <sup>13</sup>Laboratory of Soil Biology, University of Neuchâtel, Rue Emile Argand 11, CH-2000, Neuchâtel, Switzerland, <sup>14</sup>UPMC Université Paris 06, 29682 Roscoff, France, <sup>15</sup>CNRS, FR2424, Station Biologique de Roscoff, 29682 Roscoff, France, <sup>16</sup>Ifremer, Centre de Brest, DYNECO/Pelagos BP70, 29280 Plouzané, France and <sup>17</sup>Point Compétence Informatique, Rue Jean-Baptiste Say, 56850 Caudan, France

### Abstract

The interrogation of genetic markers in environmental meta-barcoding studies is currently seriously hindered by the lack of taxonomically curated reference data sets for the targeted genes. The Protist Ribosomal Reference database (PR<sup>2</sup>, <http://ssu-rrna.org/>) provides a unique access to eukaryotic small sub-unit (SSU) ribosomal RNA and DNA sequences, with curated taxonomy. The database mainly consists of nuclear-encoded protistan sequences. However, metazoans, land plants, macrosporidic fungi and eukaryotic organelles (mitochondrion, plastid and others) are also included because they are useful for the analysis of highthroughput sequencing data sets. Introns and putative chimeric sequences have been also carefully checked. Taxonomic assignation of sequences consists of eight unique taxonomic fields. In total, 136 866 sequences are nuclear encoded, 45 708 (36501 mitochondrial and 9657 chloroplastic) are from organelles, the remaining being putative chimeric sequences. The website allows the users to download sequences from the entire and partial databases (including representative sequences after clustering at a given level of similarity). Different web tools also allow searches by sequence similarity. The presence of both rRNA and rDNA sequences, taking into account introns (crucial for eukaryotic sequences), a normalized eight terms ranked-taxonomy and updates of new GenBank releases were made possible by a long-term collaboration between experts in taxonomy and computer scientists.

**Keywords:** SSU, Ribosomal, Reference, Database, Next Generation Sequencing



## Introduction

The modern definition of the term “protist” refers to unicellular eukaryotes that are either free-living or parasitic, sometimes forming colonies, but without clear differentiation into tissues. This includes all eukaryotes other than land plants (and macro-algae), animals and fungi with differentiated tissues. Protists are notoriously paraphyletic and include a wide range of microorganisms using a huge variety of reproductive, nutritional and life-history strategies. Nevertheless, the term protist has pragmatic uses and has recently gained in popularity. Large-scale analysis of protistan diversity is complicated by their heterogeneity, which reflects their extremely broad distribution and implication in multiple ecological and functional processes. This difficulty is exacerbated by the following facts: (i) species delineation is often obscure owing to lack of clear morphological criteria and paucity of knowledge concerning processes of sexual recombination; (ii) the taxonomy of protists has been radically modified in recent decades in light of new phylogenetic data; and (iii) a large proportion of protists are probably still not cultivable or yet unknown. Molecular barcoding using SSU rRNA (Small Sub-Unit Ribosomal) gene sequences consequently has become extremely popular among protistologists. Environmental barcoding has unveiled an extensive genetic diversity of protists in a wide range of ecosystems (Lopez-Garcia *et al.* 2001; Moon-van der Staay *et al.* 2001), including lineages only known by their genetic signatures (orphan environmental sequences). Recently, the use of next generation sequencing (NGS) technologies targeting selected domains of the SSU rRNA gene has permitted ecological studies of complex assemblages at ever increasing scales (Christen *et al.* 2010a; 2011; Edgcomb *et al.* 2011; Lecroq *et al.* 2011; Hartmann *et al.* 2012). However, interpretation of such data is currently seriously hindered by the lack of taxonomically curated reference data sets. Unassigned and incorrectly assigned sequences are accumulating at an increasing and alarming rate in public databases, to the extent that in early 2012, almost 20% of submitted SSU rRNA eukaryotic gene sequences had no or a very poor taxonomic assignment (see the website for more details: <http://ssu-rna.org>). Undetected chimeric sequences (Berney *et al.* 2004), as well as the presence of introns in gene sequences (Guillou *et al.* 2013), are also problematic.

To facilitate and increase the efficiency and accuracy of NGS data sets analyses, we here present the first comprehensive-curated database that places eukaryotic SSU rRNA gene sequences within a coherent ranked taxonomic framework covering eukaryotic diversity. Every sequence was quality checked and annotated using a multi-level taxonomic assignment. As a lot of protists are still only known by their environmental sequences, cluster names were retained when the formal taxonomy was missing [such as Syndiniales (Guillou *et al.* 2008) and MArine STramenopiles, MAST (Santini *et al.* 2012)]. Although curated in less detail, sequences from metazoa, land plants and macrosporic fungi, as well as eukaryotic organelles (mitochondria, plastids, etc.), are also included in the database for their ecological interests. For example, protists may live in close association with

metazoan (commensalisms, symbioses, etc.), and very small metazoan exists, inhabiting similar ecological niches. For example, copepods and polychaetes, as well as benthic animal larvae coexist with planktonic protists in aquatic systems. They may also have a great interest in ecological studies (as predators for example), even for protistologists. Even if this database is dedicated to protists, such outgroup sequences are of high relevance for extracting these groups in further analyses of NGS data sets when 'universal' eukaryotic primers are used for polymerase chain reaction (PCR) amplifications. Metazoan sequences in PR2 allow not identifying them wrongly as new deep lineages of protists.

## Materials and methods

The construction of this database started >10 years ago, and our procedure have been optimized over time (for more details, recent history detailed at <http://ssu-rrna.org/method.html>). Here, we briefly describe the present general architecture of the database.

Entries containing at least one partial SSU rRNA gene sequence of eukaryotic origin are retrieved from three public databases using keywords. Our last update retrieved 484.657, 496.462 and 123 such entries from GenBank, EMBL and WGS-EMBL, respectively. An INSDC (<http://www.insdc.org/>) entry as defined by its accession number in public databases may contain several rRNA gene sequences, e.g. in long genomic fragments containing several partial or complete ribosomal operons. To allow such duplicated sequences within a single entry, each sequence was given a unique identifier, acc.p1.p2, where acc is the accession number of the entry containing the sequence, and p1 and p2 are the first and last positions of the sub-sequence within the complete sequence.

A majority of extracted sequences were shorter than 100 nucleotides or around 500 nucleotides (63% of retrieved sequences), likely resulting from the recent integration of short environmental sequences derived from clone libraries. Only sequences longer than 799 nucleotides were considered.

The first step was the identification of sequences originating from organelles. A reference database of SSU-rRNA gene sequences from chloroplasts and mitochondria was constructed using entire genomes or genomic fragments that contained a SSU-rRNA gene sequence and a protein-coding gene specific either of mitochondria or of chloroplasts. For derived-organelle sequences such as apicoplasts, hydrogenosomes and nucleomorphs, databases were manually built, using information found in scientific publications. These databases were used to determine by sequence similarity the origin of every sequence in the database. These sequences were assigned to a reduced taxonomic framework, including their location (such as: |Organelle|chloro-SSU| or |Organelle|mito-SSU|). These sequences are not more detailed in the database.

Introns were found to be a major problem in eukaryotic rRNA sequences compared with prokaryotic sequences (1,536 sequences with intron(s) described, 10,644 sequences with introns

found by computation). A dedicated C++ algorithm was developed to identify the presence of introns in the remaining sequences (Guillou *et al.* 2013). When detected, sequences with and without the intron(s) were generated (rRNA and rDNA sequences).

Sequences in the PR2 database are assigned an identifier in the form accession.p1.p2\_X, where accession is the accession number of an entry, p1 and p2 are the positions of this sequence in a larger genomic entry and X corresponding to introns treatment of the sequence [X = G: genomic sequence containing a described intron (rDNA); X=R: the previous genomic rRNA sequence, without the intron(s); X=U: no intron described, but intron(s) may be present; X=UC: introns were detected in silico and removed from the sequence (putative rRNA)].

### Taxonomy of nuclear-encoded sequences

As all SSU-rRNA genes are orthologs, a global phylogeny can be built, and essential past speciation events can be evidenced. This property is essential to build a ranked taxonomy. For example, at rank 1, there is a worldwide agreement to recognize three clades, Bacteria, Archaea and Eukaryota. We chose to additionally use “Organelle” as rank 1. Organelles have a eukaryote origin when they are nucleomorphs and a bacterial origin when they are mitochondrion and plastid. Because evolution of organelles and their hosts differ over time, their taxonomy is different too. In addition, scientists working on diversity are more interested in the identification of the cells that bear such organelles. Our choice was thus to allow their easy identification (and filtering out) during the first step of an analysis, targeting them as “Organelle” at rank 1.

Nomenclature and terms of the following ranks mainly follows the classification of eukaryotes proposed by Adl *et al.* (Adl *et al.* 2012). Thus, the second rank describes each eukaryotic “Super-Group” or Phylum (both terms are in use in different communities): Alveolata, Amoebozoa, Apusozoa, Archaeplastida, Excavata, Opisthokonta, Rhizaria or stramenopiles. The taxonomic descriptions are structured by the use of eight ranks, and following ranks mainly correspond to the division, class, order, family, genus and species.

The terms used for each rank are non-ambiguous (a term cannot be found in two different clades), contain no space (that may pose problems to computers) and whenever possible retained if monophyletic. When monophyly could not be insured, the term of rank above was used, appended with suffix \_X (suffix X if the above rank was already \_X). As the same species name frequently occurs in different genera, the species name is composed of the genus and species, using “+” as a separator (e.g. genus = *Diderma*, species = *Diderma+niveum*). Genus and species names from public databases are stored in separate fields for comparison.

For protists and unicellular fungi, a taxonomy was proposed by the group of experts, authoring this article. For multicellular fungi, plants and metazoans, the taxonomy was built mostly using the

taxonomy assigned in National Center for Biotechnology Information (NCBI)'s GenBank database entries. We first built a core reference database containing 23,116 manually analysed sequences representative of eukaryotic diversity. These analyses included reading published articles and phylogenetic analyses done by the authors of this article when necessary. This core reference database was subsequently used to automatically annotate the remaining sequences using different methods.

We are aware that for some clades such as metazoa, plants and fungi, our eight terms taxonomy is probably not as precise as it should be. Barcoding of metazoa and plants using SSU-rRNA sequences is not often used (normally only to complement Internal Transcribed Spacer (ITS) sequences). We will therefore try in a next release to propose an extended, still ranked and unified, taxonomy for fungi.

An outcrop of PR2 is the web-based tool KeyDNATools (<http://keydnatools.com/>). It uses 159,982 specific short (15nt) oligonucleotide sequences (named keys) generated from the core reference database. Each key is a signature present in sequences of a given clade, but not in those of other clades. Besides providing a very fast taxonomic identification, it also allows for detecting putative chimeric sequences, as when different identifications are obtained from the 5' and 3' ends of sequences.

Specific new computer programs mostly in C, C++ and Python have been developed. First, a new parallel distributed computing Needleman–Wunsch-based C program allowing to compute pairwise distances not taking into account terminal gaps (partially overlapping sequences) and long internal gaps (introns). This was coupled to a newly rewritten C average linkage-clustering program. Second, a new parallel distributed computing Needleman–Wunsch-based C++/Python program allowing to assign a consensus taxonomy to new sequences by comparison to a reference database (Crunch\_Assign).

When a conflict between taxonomies assigned using the different methods was found, it was manually solved. In the end, each nuclear encoded sequence is assigned an identifier in the form of this example:

```
>AY827845.1.1765_U|Eukaryota|Apusozoa|Hilomonadea|Planomonadida|Planomonadidae|Planomonadidae_Group1|Ancyromonas|Ancyromonas+sigmoides
```

## Results

In total, we found 136,866 nuclear encoded sequences, five pseudo-genes (FJ854546, FJ854545, D14632, AF310844, AJ404858, not included in PR2) and 34 sequences we could only assign as putative rRNA sequences (HM538255, GU385678, AB275106, AJ628837, AY180011, CP000499, CP000499, AY256215, EU402432, AB017015, GQ330639, GU820811, JF488788, AF239231, DQ423737, DQ104596, AY835700, DQ423728, EU545797, GU072272, GU072526, GQ247249, HM174255, DQ104594, EU174762, FN598473, EU726200, EF695080, GQ483783, GQ462590, EU173354, EF567390, EF695215, HQ871039, not included in PR2). Manual analyses of some of them allowed concluding for the presence of artefactual sequence internal or at the 5' or 3' end. Among nuclear-encoded sequences, we detected 1,756 putative chimeric sequences, either using the KeyDNATools and/or by manual inspection (listed on the website). For example, sequence EF023694.1.1975\_U is a chimera between parent sequences of Opisthokonta, Amoebozoa and Rhizaria in position 179-471, 623-1264 and 1536-1925, respectively. Other “18S” sequences are nucleomorphs (262 sequences). In all, 9,657 sequences have a chloroplastic origin, 33,051 are from mitochondria, six from hydrogenosomes (AJ237907, AJ237908, AJ871215, AJ871217, AJ871267, Y16670) and 26 from apicoplasts (U87145, AB471801, AB471802, AB471803, AB471804, AB471805, AB471806, AB471807, AB471808, AB471809, AB471810, AB471811, AB471812, AB649417, AB649418, AB649419, AB649420, AB649421, AB649422, AB649423, AB649424, HQ110105, JQ437257, JQ437258, JQ437259, U28056).

Within nuclear-encoded sequences, 54 data entries remained unassigned at the Super-Group level (Table 1), meaning that they could not be assigned to any specific taxon group within the domain Eukaryota (Eukaryota\_X). The Super-Group “Eukaryota\_Mikro” was created for sequences HM563060, AF477623 and HM563061, for which no consensus has been reached for their affiliation, although Haplosporidiidae has been suggested (Carnegie *et al.* 2003). BLAST analyses conducted at NCBI against non-redundant or at DNA Data Bank of Japan (DDBJ) against all showed extremely weak sequence similarity with sequences of fungi. Using our global similarity tool (Crunch\_Assign) showed no other sequence similar at  $\geq 80\%$  along the entire sequence. These results conducted to the creation of this new Super-Group (rank 2). For unassigned nuclear-encoded sequences (Eukaryota\_X), either no other similar sequence was found or similar sequences were detected but also annotated by us as Eukaryota\_X. A BLAST on NCBI non-redundant (excluding environmental sequences) and at DDBJ (all) revealed that a large number of them probably contained undescribed introns. Therefore, these sequences probably require a manual curation, but again highlight the importance of intron identification in eukaryotic sequences.

For lower taxonomic ranks, there were primarily two types of cases resulting in a failure to assign a taxonomic identity:

- (1) No agreement between experts to resolve at a given rank. For example, the genus (rank 7) is assigned, the order (rank 5) is assigned, but a family (rank 6) has not yet been described, or this rank is in fact polyphyletic, with no proper descriptions of the different families.

| Super-group                     | n1      | n2      |
|---------------------------------|---------|---------|
| Alveolata                       | 20,760  | 20,255  |
| Amoebozoa                       | 1,902   | 1,880   |
| Apusozoa                        | 254     | 242     |
| Archaeplastida                  | 16,309  | 16,092  |
| Eukaryota_Mikro                 | 3       | 3       |
| Eukaryota_X                     | 54      | 54      |
| Excavata                        | 2,871   | 2,869   |
| Hacrobia                        | 2,192   | 2,132   |
| Opisthokonta                    | 75,056  | 74,484  |
| Rhizaria                        | 7,581   | 7,459   |
| Stramenopiles                   | 9,884   | 9,640   |
| Total nuclear-encoded Eukaryota | 136,866 | 135,110 |
| Apicoplast                      | 26      | 26      |
| Chloroplast SSU                 | 9,657   | 9,657   |
| Hydrogenosome SSU               | 6       | 6       |
| Mitochondrion SSU               | 36,051  | 36,051  |
| Nucleomorph SSU (18S)           | 264     | 262     |

**Table 1:** Number of nuclear-encoded sequences in PR2 as annotated at the Super-Group taxonomic level n1, total number; n2, excluding putative chimera; Super-Group, rank 2 taxonomy.

- (2) A given sequence is similar at the family level with several sequences from different families; however, they agree at the order level.

In such cases, this sequence was assigned as: ...|Order|Order\_X|Genus|Genus+species. If a genus was not described (i.e. uncultured), the taxonomy becomes: ...Order|Order\_X|Order\_XX|Order\_XX+sp.

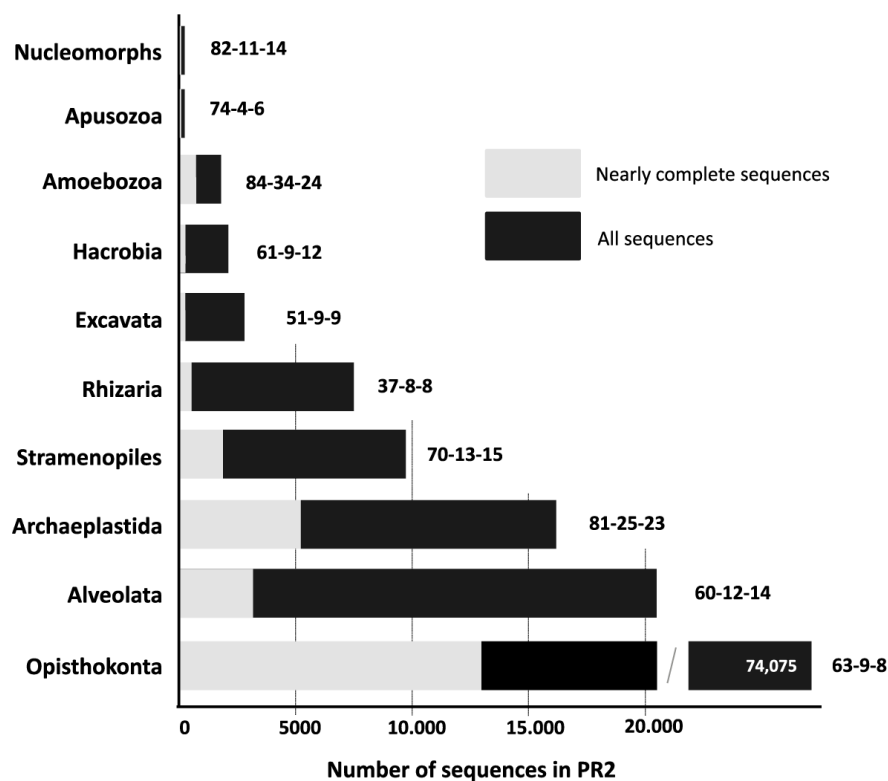
More than 74,000 sequences (54% of total number of sequences in the PR2 database) belong to Opisthokonta (Figure1). Alveolata and Archaeplastida are second in abundances (15 and 12%, respectively). Stramenopiles and Rhizaria represent 7.2 and 5.6%, respectively. Others Super Groups represent less than 2.2%. Only 29.4% are complete or nearly complete. In total, 63.7% of sequences include the V4 region and only 12.1% and 11.7% include the V9 region as recognized by primers Biomarks and Wamps (see the legend of Figure 1), respectively. Apusozoa, Hacrobia, Excavata and Opisthokonta have <10% of their sequences that include the V9 region. V9 region of Amoebozoa and Archaeplastida are better represented (34% and 25%, respectively, using the Biomarks primers).

## Downloads

We provide several different ways of downloading the database or part of it.

(See more explanations at [http://ssu-rrna.org/downloads\\_eukaryotic\\_main\\_page.html](http://ssu-rrna.org/downloads_eukaryotic_main_page.html))

- (1) The entire database or sequences of a specific clade can be downloaded using a taxonomy browser under fasta format, with sequence identifiers as described above. Putative chimeras have been removed.
- (2) The entire database or sequences of major groups can be downloaded under fasta format, with only the short unique identifier. The corresponding taxonomy is then downloaded as a tabulated file. This fasta format is appropriate to use in tools that do not allow for long sequence identifiers. They are also easier to use in large computations, as they spare the memory required. Finally, they are easier to use in pipelines or web sites (see below).
- (3) The entire database, taxonomies and sequences under tabulated format, for easy import in relational databases.
- (4) The entire database or sequences of a specific clade under fasta format, with sequence identifiers as described above, but after a clustering by sequence similarity (98, 96, 92%) and choosing only the longest sequence as representative of the cluster.
- (5) Phylogenetic trees are available for the main groups. They were built using pairwise distance computations (not taking introns as differences as explained above) and FastMe (Gascuel & Desper 2002).
- (6) Finally, we provide an “arb” filter that allows to easily import a fasta file (with taxonomy in the identifier) into an arb database, separating sequences and taxonomy as required.
- (7) In silico extracted domains corresponding to regions widely used in published articles and corresponding to several couples of primers.



**Figure 1:** Total number of SSU rDNA gene sequences in the PR2 database for each main eukaryotic lineage (all sequences = grey+black, complete or nearly complete sequences in light-grey). Note that nucleomorphs were extracted from Archaeplastida. Numbers indicated after bars indicate percentages of sequences that include the following: (i) the V4 region as defined by primers forward CCAGCASCYGGGTAATTCC and reverse ACTTTCGTTCTTGATYRA used during the European Biomarks project; (ii) the V9 region as defined by primers forward GTACACACCGCCGTC and reverse TGATCCTTCTGCAGGTTACCTAC used during the European Biomarks project; and (iii) the V9 region defined by primers forward TTGTACACACCGCCC and reverse CCTTCYGCAGGTTACCTAC used by the WAMPS project. For Opisthokonta, number in white=total number of sequences.

## Searching the database

We provide the following additional kinds of tools:

- (1) A search by keywords, allowing to search according to taxonomy, accession number and PMID (PubMed ID: retrieval of sequences described in a given publication). Retrieve sequences can be filtered according to length, quality and when containing the variable V4 of V9 domains (often used in conjunction with deep sequencing).
- (2) A search by “sequence signature”, with a link to the KeyDNAtools website (<http://keydnatools.com/>). This tool provides very fast results even for files containing many sequences. It also allows for detection of putative chimera as explained above.
- (3) A BLAST search against the database, as usually found on most sites.
- (4) A search (Crunch\_Assign) using our modified global (Needleman–Wunsch based) algorithm that returns the most similar hits based on the entire alignment of the sequences, and not based on a



good local alignment (high scoring pair, in BLAST). As a result, the percentage of similarity computed is more in agreement with what would be found using a Multiple Sequence Alignment [Clustal (Larkin *et al.* 2007), Muscle (Edgar 2004), MAFFT (Katoh & Toh 2010), ...] before computing distances. It allows or does not allow accounting for introns as described above.

- (5) A search of one or two primer motifs in sequences, returning every sequence that contains the primer(s) with International Union of Pure and Applied Chemistry (IUPAC) encoding allowed and also the possibility of mismatches between primer and sequence (a C program).
- (6) In silico extracted domains corresponding to regions widely used in published articles and corresponding to several couples of primers.

Both BLAST and Crunch\_Assign similarity searches are coupled to BLAST2Tree or Crunch\_Assign2Tree that use our Scriptree software (Christen *et al.* 2010b). Similarity search results can simply be copied and then pasted in the “2Tree section”; a phylogenetic tree is built and displayed on the fly, with taxonomic assignments (as chosen by the user) displayed in regard of each leaf. This section also allows downloading the sequences that have been pasted and the taxonomy as a tabulated file (Croce *et al.* 2010).

## Conclusion and perspectives

There are presently three databases, SILVA (Pruesse *et al.* 2007), RDP (Cole *et al.* 2005) and Green genes (DeSantis *et al.* 2006), offering a curated taxonomy for prokaryotic SSU rRNA sequences. Only SILVA additionally provides reference sequences for SSU-rRNA sequences of eukaryotic origin, curated for sequence quality but using the NCBI taxonomy (although recently a “SILVA” taxonomy is now proposed). Because our sequence identifier, i.e. accession.p1.p2, is similar to that used by SILVA, both databases can be easily compared.

Based on the last release 111, 1,518 of the 71,787 eukaryotic SILVA reference sequences are not present in the PR2 database. Manual checks showed that these sequences correspond to sequences extracted from entries in which no annotation allowed to identify the presence of a SSU-rRNA sequence, annotated as mRNA or annotated as prokaryotes. In all, 670 sequences identified as mitochondria were not in PR2; none of the SILVA chloroplast sequences was absent from PR2. Missing sequences will be soon analysed and incorporated in PR2. On the other hand, 53,735/7,774 nuclear, 31,492/29,763 mitochondrial, 462/18 chloroplastic and 133/80 other organelle sequences present in PR2 were not in SILVA reference sequences and SILVA entire database, respectively. This can be largely explained by the use of drastic filtering steps used by SILVA both in minimal length and sequence quality. However, because we are also users of such databases to analyze NGS data sets, we detected

two major reasons not to use too drastic quality filtering. First, representatives of novel environmental clades are often found within clone libraries with length of <1,000 nucleotides. Also, use of extreme quality filters may remove important sequences representatives of environmental groups, too short and/ or having poor quality at one of the end of a sequence (one-step Sanger sequencing without enough noise treatment for example). In PR2, sequence quality was indirectly inferred by the quality of the taxonomic assignation because bad-quality sequences became poorly assigned. Again, as sequence identifiers are similar between both databases, sequences can be easily compared between both databases.

The PR2 database possesses several valuable complementary tools or databases lacking in other databases.

### A ranked taxonomy

As for the PR2 database, SILVA taxonomy for eukaryotes now offers a taxonomy based on the structure proposed by Adl et al. (Adl *et al.* 2012). However, contrarily to SILVA, we proposed a normalized eight terms ranked taxonomy for every sequence in the database. We proceeded to this “normalization” from our experience in dealing with very large data sets using automated pipelines, and a depth of sequencing that revealed organisms spanning the entire spectrum of known living organisms. When considering the NCBI taxonomy for example, two sequences of Perciformes were found described using 22 ranks (AY263842 and EF470892 for Perciformes), whereas another Perciforme (AF112595) was described using only 15 ranks, and 10,360 sequences of Perciformes had between 16 and 21 ranks. Numerous examples exist for protists. A very good example is for the genus *Carpediomonas*. NCBI classify this genus within Eukaryota (rank 1), Fornicata (rank 2), *Carpediomonas* (rank 3). However, sequence AY117416 (*Carpediomonas membranifera*, (Simpson *et al.* 2002)) has no rank 2 taxonomy in its entry. As a result, it becomes extremely difficult using a computer and the lists of terms provided by a non-ranked taxonomy to identify for two different sequences, which members of the two lists indeed correspond to the same rank. This is the problem solved by our ranked taxonomy, thanks to a worldwide list of taxonomic experts. As an example, taxonomy of sequence AY117416 becomes Eukaryota|Excavata|Metamonada|Fornicata|Fornicata\_Group\_2|Carpediomonas\_like|Carpediomonas|Carpediomonas+membranifera in PR2. In SILVA, this sequence is linked to a 7 terms taxonomy, but taxonomy is seemingly not ranked and unified.

When occurring, missing ranks are automatically replaced in PR2 (labeled as clade-i\_X, where clade-i is the term for the next higher rank). This strategy allows rapidly inferring the taxonomy at the most probable higher rank and provides a rapid method for screening putative novel lineages at each taxonomic level.

### Introns

Most SSU rRNA databases and biodiversity analyses of prokaryotes understandably neglect introns. Although found even in *Escherichia coli* (Ferat *et al.* 1994; Sunde 2005), introns are rare in Bacteria and not very abundant in Archaea. Even when present, they have not yet been, to our knowledge, described in rRNA gene sequences. However, in Eukaryota, introns can be relatively abundant in rRNA gene sequences at least in some groups (Guillou *et al.* 2013). This led us to incorporate in our database both the rRNA and the rDNA sequences. As most NGS (or clone library) analyses of the biodiversity are dealing with PCR amplification of extracted gDNA, introns may represent a large part of the variability observed. Having genomic sequences, in addition to the rRNA transcript, in the database is important, not only for searching by similarity but also for the *in silico* estimation of expected amplicon lengths.

### Organelles

Organelles are often poorly treated in reference databases. For hydrogenosomes (AJ237907, AJ871215, AJ871217, AJ871267, Y16670), only sequence AJ871217 can be found in SILVA labeled as “Unclassified”. For GreenGenes, sequences were not found when searching by accession number. At RDP, the classifier resulted in every case into “unclassified\_Bacteria”. For the 26 apicoplast sequences, none was found in SILVA reference sequences or in the “ssu-accession-parc.acs”, release 111 (3,186,762 accession numbers). Even for better-known organelles, taxonomic assignment is not really better. For example, sequence AB000109 mitochondrion of *Dictyostelium discoideum* is labeled as “Unclassified” in SILVA. Chloroplasts are generally well identified in SILVA. However, among the chloroplastic sequences detected in this study, 263 were found in SILVA reference sequences as chloroplasts. Our approach to build independent databases for these organelles allowed us to probably reach a more precise taxonomic affiliation of organelles. Having such prokaryotic organelles in our database is essential with NGS data sets of both prokaryotes and eukaryotes because the use of “Bacteria” or “Eukaryota” specific primers resulted in some cases in a significant proportion of amplicons that are in fact of Organelle origin (Christen *et al.* 2010a; 2011; Edgcomb *et al.* 2011; Lecroq *et al.* 2011; Hartmann *et al.* 2012). Even if Organelle sequences are simply discarded from the final analysis, this database avoids identifying these sequences as some new deep lineages.

### Chimeric sequences

Chimeric sequences are PCR-generated hybrid products between multiple parent sequences that can be falsely interpreted as novel organisms, thus inflating apparent diversity (Hugenholtz & Huber 2003; Berney *et al.* 2004). The two algorithms most widely used for 16S chimera detection are

Pintail (Ashelford *et al.* 2005), included in RDP and SILVA databases, and Bellerophon (Huber *et al.* 2004) included in GreenGenes. In all cases, chimeras are detected by comparing independent regions of a sequence alignment. The KeyDNATools does not require the prior alignment of sequences, and it is particularly efficient to detect complex chimera having more than two parent sequences, or between two closely related parents. This tool can be used in concert with other detection methods. Our database, which has been screened for putative chimera, offers two possibilities of download: either including or excluding putative chimeric sequences.

### Similarity searches

BLAST is a widely used tool that finds regions of local similarity between sequences. However, such search based on a good local high scoring pair could lead to very bad results. We thus developed two independent methods of assignation. The first one, the Crunch\_Assign software is using a Needleman–Wunsch algorithm. It is also faster than BLAST and returns a score computed on the entire alignment. Because we are working on Eukaryotes, we also included the possibility of ignoring putative introns (to our knowledge, this possibility is not included in any other software). The second one, the KeyDNATools is also very fast and offers additionally chimera detection as discussed above. In >95% of cases, both assignations provide similar results. Sequences not annotated by the KeyDNATools likely result from the absence of the corresponding clade in the core reference database, low quality sequences or novel variants of the gene present in newly available sequences, not yet included in the core data set. Conversely, sequences not assigned by the Crunch\_Assign software are often chimera or low-quality sequences. After a search by similarity, we offer the possibility to build a phylogenetic tree on the fly, using most similar sequences found by BLAST or Crunch\_Assign.

### Updates

We have developed a pipeline that allows to analyze a GenBank new release within a week. Most of the time spent is indeed in manual checking of conflicts after average linkage clusterings, as explained previously. As a result, updates of the PR2 database will be done shortly after each GenBank new release. As a result, numbers provided in this article will probably differ from that available from PR2 at publication time of this manuscript.

---

### Acknowledgments:

The authors warmly thank Marion Viprey for her pioneering helps with the construction of the PR2 database. Computations have been done on the “Mésocentre SIGAMM” machine, hosted by Observatoire de la Côte d’Azur, Nice, France.

### Funding

The European Union’s Seventh Framework Programmes (FP7) BIOMARKS (2008-6530, ERA-net Biodiversa) and MicroB3 [287589] and the following ANR (France) projects: AQUAPARADOX, PARALEX and GIME. Funding for open access charge: ANR Paralex (French) and BIOMARKS (FP7).

## References

- Adl SM, Hoppenrath M, Simpson AGB *et al.* (2012) The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, **59**, 429–493.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, **71**, 7724–7736.
- Berney C, Fahrni J, Pawlowski J (2004) How many novel eukaryotic “kingdoms?” Pitfalls and limitations of environmental DNA surveys. *BMC Biology*, **2**, 1–13.
- Carnegie RB, Meyer GR, Blackburn J *et al.* (2003) Molecular detection of the oyster parasite *Mikrocytos mackini*, and a preliminary phylogenetic analysis (vol 54, pg 219, 2003). *Diseases of Aquatic Organisms*, **54**, 219–227.
- Christen R, Behnke A, Engel M *et al.* (2010a) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.
- Christen R, Chevenet F, Croce O, Hebrard M, Berry V (2010b) ScripTree: scripting phylogenetic graphics. *Bioinformatics*, **26**, 1125–1126.
- Christen R, Pawlowski J, Bachar D *et al.* (2011) Eukaryotic Richness in the Abyss: Insights from Pyrotag Sequencing. *PLoS ONE*, **6**, e18169.
- Cole JR, Chai B, Farris RJ *et al.* (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic acids research*, **33**, D294–D296.
- Croce O, Christen R, Chevenet F (2010) A new web server for the rapid identification of microorganisms. *J Microbial Biochem Technol*, **2**, 84–88.
- DeSantis TZ, Huber T, Hugenholtz P *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–1797.
- Edgcomb V, Orsi W, Bunge J *et al.* (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME Journal*, **5**, 1344–1356.
- Ferat J-L, Gouar M, Michel F (1994) Multiple group II self-splicing introns in mobile DNA from *Escherichia coli*. *Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie*, **317**, 141–148.
- Gascuel O, Desper R (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of computational biology*, **9**, 687–705.
- Guillou L, Bachar D, Christen R (2013) Detection of Introns in Eukaryotic Small Subunit Ribosomal RNA Gene Sequences. *Dataset Papers in Biology*, **2013**, 1–12.
- Guillou L, Massana R, Viprey M *et al.* (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, **10**, 3349–3365.
- Hartmann M, Howes CG, VanInsberghe D *et al.* (2012) Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. *The ISME Journal*, **6**, 2320–2320.

- Huber T, Hugenholtz P, Faulkner G (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, **20**, 2317–2319.
- Hugenholtz P, Huber T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 289–293.
- Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, **26**, 1899–1900.
- Larkin A-M, Blackshields G, Brown NP, Chenna R (2007) Clustal W and Clustal X version 2.0. *bioinformatics applications note*, **23**, 2947–2948.
- Lecroq B, Lejzerowicz F, Bachar D *et al.* (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *PNAS*, **108**, 13177–13182.
- Lopez-Garcia P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**, 603–607.
- Moon-van der Staay SY, De Wachter R, Vaultot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, **409**, 607–610.
- Pruesse E, Quast C, Knittel K *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research*, **35**, 7188–7196.
- Santini S, Logares R, Audic S *et al.* (2012) Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *The ISME Journal*, **6**, 1823–1833.
- Simpson AGB, Roger AJ, Silberman JD *et al.* (2002) Evolutionary history of “early-diverging” eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Molecular biology and evolution*, **19**, 1782–1791.
- Sunde M (2005) Class I integron with a group II intron detected in an *Escherichia coli* strain from a free-range reindeer. *Antimicrobial agents and chemotherapy*, **49**, 2512–2514.

## Liste des 1 191 séquences de références V9 constituant la base DinR2

>EF016919|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelBlastodinales\_01|OodiniaceaelPiscinoodinium|sp.

>AF080096|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelBlastodinales\_02|OodiniaceaelAmyloodinium|ocellatum

>DQ490256|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelBlastodinales\_02|OodiniaceaelAmyloodinium|ocellatum

>DQ490257|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelBlastodinales\_02|OodiniaceaelAmyloodinium|ocellatum

>DQ490258|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelBlastodinales\_02|OodiniaceaelAmyloodinium|ocellatum

>AB211357|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophyceae-XI|Dinophyceae-XXI|Pileidinium|ciceropse

>HM483398|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophyceae\_XI|Dinophyceae-XXI|Duboscquodinium\_01|collinii

>HM483399|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophyceae\_XI|Dinophyceae-XXI|Duboscquodinium\_02|collinii

>HM853811|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_01|DinophysiaceaelDinophysis\_01|lodia

>HM853812|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_01|DinophysiaceaelDinophysis\_01|lodia

>GU647166|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_02|DinophysiaceaelDinophysis\_02|lacuminata

>AB073117|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_03|lacuminata

>AJ506972|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_03|lacuminata

>EU130569|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_03|lacuminata

>GU647163|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_03|lacuminata

>AB073118|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|fortii

>AB073119|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|norvegica

>AB366002|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|infundibulus

>AF239261|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|norvegica

>AJ506973|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|lacuta

>AJ506974|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|norvegica

>EU780644|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|caudata

>HM853806|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|phalacromoides

>HM853815|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|caudata

>HM853816|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_04|tripos

>HM853813|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_05|hastata

>HM853814|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_05|hastata

>HM853810|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_06|monacantha

>HM853807|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_07|sp.

>HM853808|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_07|sp.

>HM853809|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_07|pusilla

>HM853805|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelDinophysis\_08|hastata

>HM853802|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelHistioneis\_01|gubernans

>HM853803|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelHistioneis\_02|longicollis

>HM853801|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelHistioneis\_03|cymbalaria

>HM853804|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelHistioneis\_04|longicollis

>EU780646|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelHistioneis\_05|sp.

>EU780649|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|magnificus

>HM853793|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|heteroporus

>HM853794|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|heteroporus

>HM853795|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|heteroporus

>HM853796|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|heteroporus

>HM853797|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|magnificus

>HM853798|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|quadratus

>HM853799|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|quadratus  
 >HM853800|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelOrnithocercus\_01|quadratus  
 >HM853782|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelOxyphysisloxytoxoides  
 >HM853786|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_01|favus  
 >HM853787|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_01|favus  
 >AJ506975|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|rotundata  
 >EU780657|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|isp.  
 >FJ477084|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|isp.  
 >HM853772|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|parvulum  
 >HM853773|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|parvulum  
 >HM853779|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|doryphorum  
 >HM853780|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|doryphorum  
 >HM853781|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|doryphorum  
 >HM853784|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|rotundata  
 >HM853788|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|porodictyum  
 >HM853789|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|porodictyum  
 >HM853790|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|porodictyum  
 >HM853791|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|porodictyum  
 >HM853792|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_02|porodictyum  
 >AB551248|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|mitra  
 >EU780655|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|irapa  
 >FJ477082|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|irapa  
 >HM853774|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|irapa  
 >HM853775|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|mitra  
 >HM853776|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|mitra  
 >HM853777|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|mitra  
 >HM853785|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_03|DinophysiaceaelPhalacroma\_03|isp.  
 >HM853764|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_04|AmphisoleniaceaelAmphisolenia\_01|isp.  
 >HM853767|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_05|AmphisoleniaceaelTriposolenialbicornis  
 >HM853768|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_05|AmphisoleniaceaelTriposolenialbicornis  
 >HM853769|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_05|AmphisoleniaceaelTriposolenialbicornis  
 >HM853770|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_05|AmphisoleniaceaelTriposolenialbicornis  
 >HM853763|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_06|AmphisoleniaceaelAmphisolenia\_02|lidentata  
 >HM853765|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_06|AmphisoleniaceaelAmphisolenia\_02|globifera  
 >HM853766|EukaryotalAlveolatalPyrrophytopythaldinophyceaelDinophysiales\_06|AmphisoleniaceaelAmphisolenia\_02|schauinslandii  
 >DQ322643|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_01|CrypthecodiniaceaelCrypthecodinium|isp.  
 >FJ821501|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_01|CrypthecodiniaceaelCrypthecodinium|cohnii  
 >M64245|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_01|CrypthecodiniaceaelCrypthecodinium|cohnii  
 >AY443013|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_02|GonyaulacaceaelGonyaulax\_01|verior  
 >AB375868|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_03|GonyaulacaceaelAmylax|lboxus  
 >AB375869|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_03|GonyaulacaceaelAmylax|tricantha  
 >AF274269|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_04|GonyaulacaceaelLingulodinium|polyedrum  
 >AF377944|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_04|GonyaulacaceaelLingulodinium|polyedrum  
 >AJ415511|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_04|GonyaulacaceaelLingulodinium|polyedrum  
 >AY421788|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_04|GonyaulacaceaelLingulodinium|polyedrum  
 >EF492507|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_04|GonyaulacaceaelLingulodinium|polyedrum  
 >GU647183|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_05|CeratiaceaelNeoceratium\_01|isp.  
 >AY443014|EukaryotalAlveolatalPyrrophytopythaldinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|hirundinella



>AY460574|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >DQ487192|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >GU647172|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >GU647173|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >GU647181|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >GU647182|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >GU647188|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_06|CeratiaceaelNeoceratium\_02|Isp.  
 >AF274275|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_07|GoniodomataceaelPyrodiniumbahamense  
 >AY456115|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_07|GoniodomataceaelPyrodiniumbahamense  
 >AB088298|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|linsuetum  
 >AB088302|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|pseudogonyaulax  
 >AB538439|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >AJ535376|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|tamutum  
 >AJ535377|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|tamutum  
 >AJ535378|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|tamutum  
 >AJ535379|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|tamutum  
 >AJ535380|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >AJ535381|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >AJ535382|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >AJ535383|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >AJ535384|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >AJ535385|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|taylori  
 >AJ535388|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >AJ535389|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|taylori  
 >AJ535390|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|taylori  
 >AMU27499|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >AOU27500|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >AY641564|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|hiranoi  
 >AY641565|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|leei  
 >AY641566|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|satoanum  
 >AY831408|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >AY883005|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|monilatium  
 >AY883006|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >EU418967|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|Isp.  
 >JF521619|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|landersoni  
 >JF521620|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|landersoni  
 >JF521621|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >JF521630|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|linsuetum  
 >JF521631|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >JF521632|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >JF521633|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >JF521634|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >JF521635|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|minutum  
 >JF521636|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >JF521637|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lostenfeldii  
 >JF521638|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|pseudogonyaulax  
 >JF906999|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|lusitanicum  
 >U27498|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_08|GonyaulacaceaelAlexandrium\_01|margalefii

>U27500|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_08|Gonyaulacaceae|Alexandrium\_01|lostenfeldii

>AF244939|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_09|Ostreopsidaceae|Ostreopsis|sp.

>AJ415509|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_10|Ostreopsidaceae|Coolia|monotis

>EF492487|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_10|Ostreopsidaceae|Coolia|monotis

>EF492488|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_10|Ostreopsidaceae|Coolia|sp.

>HQ897281|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_10|Ostreopsidaceae|Coolia|sp.

>HQ897282|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_10|Ostreopsidaceae|Coolia|canariensis

>AY443024|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_11|Pyrophacaceae|Pyrophacus|steinii

>AF033869|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_12|Pyrophacaceae|Fragilidium|subglobosum

>EF492489|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_12|Pyrophacaceae|Fragilidium|mexicanum

>FJ405355|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_12|Pyrophacaceae|Fragilidium|mexicanum

>FJ405356|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_12|Pyrophacaceae|Fragilidium|mexicanum

>AB499537|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|sp.

>AB548851|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|sp.

>DQ388463|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202846|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202847|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202848|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202849|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202850|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202851|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202852|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|lysumotoi

>EF202853|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202854|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202855|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202856|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202857|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202858|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202859|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202860|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|ruetzleri

>EF202862|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|pacificus

>EF202866|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|belizeanus

>EF202867|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|belizeanus

>EF202868|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|belizeanus

>EF202869|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|belizeanus

>EF202870|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|belizeanus

>EF202874|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|pacificus

>EF202875|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|pacificus

>EF202878|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202879|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202881|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202882|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202883|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202884|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202885|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202886|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202887|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus

>EF202888|Eukaryota|Alveolata|Pyrrophycophyta|Dinophyceae|Gonyaulacales\_13|Goniodomataceae|Gambierdiscus|toxicus



>AF052190|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_17|GonyaulacaceaelGonyaulax\_05|spinifera  
 >DQ867107|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_17|GonyaulacaceaelGonyaulax\_05|spinifera  
 >AF274258|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_18|GonyaulacaceaelGonyaulax\_06|cochlea  
 >DQ388465|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_18|GonyaulacaceaelGonyaulax\_06|cochlea  
 >AY421790|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_19|CeratiaceaelProtoceratium\_01|reticulatum  
 >AF022154|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_20|CeratacoryaceaelCeratacorys|horrida  
 >DQ388456|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_20|CeratacoryaceaelCeratacorys|horrida  
 >AF274273|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_21|CeratiaceaelProtoceratium\_02|reticulatum  
 >DQ217789|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_21|CeratiaceaelProtoceratium\_02|reticulatum  
 >AF022153|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|ifusus  
 >AF022192|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|itenua  
 >AJ276699|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|ifurca  
 >DQ388462|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|longipes  
 >EF527109|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|isp.  
 >FJ402942|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|massiliense  
 >FJ402943|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|hexacanthum  
 >FJ402944|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|concilians  
 >FJ402945|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|candelabrum  
 >FJ402946|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|leuarcuatum  
 >FJ402947|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|symmetricum  
 >FJ402948|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|pentagonum  
 >FJ402949|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|declinatum  
 >FJ402950|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|concilians  
 >FJ402951|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|petersii  
 >FJ402952|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|limulus  
 >FJ402953|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|petersii  
 >FJ402954|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|lazoricum  
 >FJ402955|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|candelabrum  
 >FJ402956|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|arietinum  
 >FJ402957|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|extensum  
 >FJ402958|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|ifusus  
 >FJ402959|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|contrarium  
 >FJ402960|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|horridum  
 >FJ402961|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|gravidum  
 >FJ402962|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|limulus  
 >FJ402963|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|kofoidii  
 >FJ402964|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|minutum  
 >FJ402965|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|paradoxides  
 >FJ824910|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|digitatum  
 >FJ824911|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_22|CeratiaceaelNeoceratium\_03|platycorne  
 >AB088280|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella  
 >AB088281|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella  
 >AB088282|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|affine  
 >AB088283|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella  
 >AB088284|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella  
 >AB088285|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella  
 >AB088286|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella  
 >AB088287|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|catenella











>JN098316|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098317|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098318|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098319|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098320|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098321|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098322|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098323|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098324|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098325|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098326|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098327|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098328|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098329|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098330|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098331|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098332|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098333|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098334|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098335|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >JN098336|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Isp.  
 >U09048|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Ifundyense  
 >X54946|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_23|GonyaulacaceaelAlexandrium\_02|Itamarensis  
 >GU647179|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_24|CeratiaceaelNeoceratium\_04|Isp.  
 >GU647187|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGonyaulacales\_24|CeratiaceaelNeoceratium\_04|Isp.  
 >EF058240|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_01|TovelliaceaelJadwigialaplanata  
 >EF058239|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_02|GymnodiniaceaelGymnodinium\_01|Iimpatiens  
 >AY443016|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_03|HemidiniaceaelHemidinium|Inasutum  
 >JQ439938|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_04|TovelliaceaelEsoptrrodinium|Isp.  
 >JQ439940|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_04|TovelliaceaelEsoptrrodinium|Isp.  
 >JQ439941|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_04|TovelliaceaelEsoptrrodinium|Isp.  
 >JQ439942|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_04|TovelliaceaelEsoptrrodinium|Isp.  
 >JQ439943|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_04|TovelliaceaelEsoptrrodinium|Isp.  
 >JQ439944|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_04|TovelliaceaelEsoptrrodinium|Isp.  
 >AB103389|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Isp.  
 >AB103390|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Isp.  
 >AB107845|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Isp.  
 >AF274251|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Icarterae  
 >AF274255|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Imassartii  
 >AJ415512|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Icarterae  
 >AY443011|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Isp.  
 >AY443012|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Isp.  
 >EF057406|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Ioperculatum  
 >EF057407|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Icarterae  
 >EF492485|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Icarterae  
 >EU046334|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Icarterae  
 >EU046335|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Iklesii  
 >EU046336|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_05|GymnodiniaceaelAmphidinium\_01|Isp.

>EU046337|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_05|Gymnodiniaceae|Amphidinium\_01|sp.  
 >EU046338|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_05|Gymnodiniaceae|Amphidinium\_01|sp.  
 >EU046340|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_05|Gymnodiniaceae|Amphidinium\_01|sp.  
 >FR865623|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_05|Gymnodiniaceae|Amphidinium\_01|carterae  
 >FR865624|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_05|Gymnodiniaceae|Amphidinium\_01|carterae  
 >L13719|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_05|Gymnodiniaceae|Amphidinium\_01|belauense  
 >FJ947046|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_06|Warnowiaceae|Warnowia\_01|sp.  
 >FJ467491|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_07|Warnowiaceae|Erythrospidinium|lagile  
 >FJ467492|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_08|Warnowiaceae|Warnowia\_02|sp.  
 >EU293236|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_09|Gymnodiniaceae|Apicoporus\_01|parvidiaboli  
 >EU293237|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_09|Gymnodiniaceae|Apicoporus\_01|parvidiaboli  
 >EU293238|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_09|Gymnodiniaceae|Apicoporus\_01|parvidiaboli  
 >AF274252|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_10|Gymnodiniaceae|Togulajjolla  
 >AY443010|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_10|Gymnodiniaceae|Togulabritannica  
 >FJ473380|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_11|Gymnodiniales-XI|Chytridium|laffine  
 >JF790997|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_12|Kareniaceae\_01|Karlodinium\_01|micrum  
 >JF790995|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_12|Kareniaceae\_01|Takayama\_01|sp.  
 >AB288380|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_13|Gymnodiniaceae|Cochlodinium\_01|fulvescens  
 >AB288381|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_13|Gymnodiniaceae|Cochlodinium\_01|fulvescens  
 >HQ896315|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_13|Gymnodiniaceae|Cochlodinium\_01|fulvescens  
 >JF791049|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_14|Kareniaceae\_02|Karlodinium\_02|micrum  
 >GU295202|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_15|Gymnodiniaceae|Amphidinium\_02|mootonorum  
 >AF274253|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_16|Gymnodiniaceae|Amphidinium\_03|herdmanii  
 >FJ947036|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_17|Warnowiaceae|Proterothropsis|sp.  
 >FJ947037|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_17|Warnowiaceae|Proterothropsis|sp.  
 >AF022196|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_18|Gymnodiniaceae|Gymnodinium\_02|sp.  
 >AY999082|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_18|Gymnodiniaceae|Gymnodinium\_02|laureolum  
 >AB120001|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_19|Gymnodiniaceae|Gyrodinium\_01|spirale  
 >AB120002|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_19|Gymnodiniaceae|Gyrodinium\_01|fusiforme  
 >FJ947038|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_20|Warnowiaceae|Nematodinium|sp.  
 >FJ947039|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_20|Warnowiaceae|Nematodinium|sp.  
 >FJ947040|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_21|Warnowiaceae|Warnowia\_03|sp.  
 >JN098208|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_21|Warnowiaceae|Warnowia\_03|sp.  
 >FN392226|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_22|Gymnodiniaceae|Gymnodinium\_03|laureolum  
 >HQ270472|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_23|Gymnodiniaceae|Gymnodinium\_04|sp.  
 >HQ270473|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_23|Gymnodiniaceae|Gymnodinium\_04|sp.  
 >FJ473378|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_24|Gymnodiniaceae|Dissodinium|pseudolunula  
 >DQ837534|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_25|Gymnodiniaceae|Gymnodinium\_05|dorsalisulcum  
 >AF022199|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_26|Gymnodiniaceae|Lepidodinium|viride  
 >AM184122|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_26|Gymnodiniaceae|Lepidodinium|chlorophorum  
 >AY331681|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_26|Gymnodiniaceae|Lepidodinium|chlorophorum  
 >DQ499645|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_26|Gymnodiniaceae|Lepidodinium|viride  
 >AB626150|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_27|Gymnodiniaceae|Spiniferodinium|sp.  
 >GU295203|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_27|Gymnodiniaceae|Spiniferodinium|lgaleiforme  
 >AB265962|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_28|Gymnodiniaceae|Gymnodinium\_06|catenatum  
 >AB265963|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_28|Gymnodiniaceae|Gymnodinium\_06|sp.  
 >AB265964|Eukaryota|Alveolata|Pyrrophyceophyta|Dinophyceae|Gymnodiniales\_28|Gymnodiniaceae|Gymnodinium\_06|microreticulatu

m

>AB265965|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|microreticulatum

>AF022193|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>AY421783|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>AY421784|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>AY421785|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>DQ779989|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>DQ779990|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>DQ785883|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>EU418954|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|isp.

>EU418972|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|isp.

>EU418973|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|catenatum

>GU647167|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_28|GymnodiniaceaelGymnodinium\_06|isp.

>AF022194|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|GymnodiniaceaelGymnodinium\_07|fuscum

>AY421789|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPheopolykrikos\_01|hartmannii

>EU418966|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPolykrikos\_01|isp.

>DQ371293|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPolykrikos\_02|lebourae

>DQ822481|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPolykrikos\_02|herdmanae

>DQ975470|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPolykrikos\_02|herdmanae

>DQ975471|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPolykrikos\_02|lebourae

>DQ975472|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_29|PolykrikaceaelPolykrikos\_02|lebourae

>EU293235|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_30|GymnodiniaceaelApicoporus\_02|glaber

>AB120003|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_31|GymnodiniaceaelGyrodinium\_02|rubrum

>DQ371291|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_32|PolykrikaceaelPolykrikos\_03|kofoidii

>DQ371292|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_32|PolykrikaceaelPolykrikos\_03|kofoidii

>DQ371294|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_33|PolykrikaceaelPheopolykrikos\_02|beauchampii

>DQ371295|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_33|PolykrikaceaelPheopolykrikos\_02|beauchampii

>AF009131|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_34|Kareniaceae\_03|Karenia\_01|mikimotoi

>AF272046|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_04|Karlodinium\_03|micrum

>JF791096|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_05|Karlodinium\_04|micrum

>AF009216|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_06|Karenia\_02|isp.

>AF352821|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_07|Karenia\_03|brevis

>AF352822|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_08|Karenia\_04|brevis

>DQ779991|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_09|Karenia\_05|brevis

>AF352820|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_10|Karenia\_06|brevis

>HM067005|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_35|Kareniaceae\_11|Karenia\_07|papilionacea

>AF272045|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_36|Kareniaceae\_12|Karlodinium\_05|micrum

>HM067002|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_36|Kareniaceae\_13|Karenia\_08|bidigitata

>EF492504|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_36|Kareniaceae\_14|Karenia\_09|brevis

>EF492505|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_36|Kareniaceae\_14|Karenia\_09|mikimotoi

>JF791035|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_37|Kareniaceae\_15|Karenia\_10|mikimotoi

>JF791040|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_37|Kareniaceae\_15|Karlodinium\_06|micrum

>JF791044|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_37|Kareniaceae\_15|Karlodinium\_06|micrum

>JF791045|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_37|Kareniaceae\_15|Karlodinium\_06|micrum

>JF791046|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_37|Kareniaceae\_15|Karlodinium\_06|micrum

>JF791048|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_37|Kareniaceae\_15|Karlodinium\_06|micrum

>AF172712|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_38|GymnodiniaceaelGymnodinium\_10|galatheanum

>AM494500|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodinales\_38|Kareniaceae\_16|Karlodinium\_07|micrum

>AJ415516|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_17|Karlodinium\_08|micrum  
 >JN986577|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_18|Karlodinium\_09|veneficum  
 >AF352819|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_19|Karenia\_11|brevis  
 >AF172713|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_20|Karenia\_12|mikimotoi  
 >FJ587220|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_21|Karenia\_13|mikimotoi  
 >AF172714|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_22|Karenia\_14|brevis  
 >AJ415518|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_22|Karenia\_14|brevis  
 >AF022195|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_23|Karenia\_15|mikimotoi  
 >HM067007|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_24|Karenia\_16|selliformis  
 >FR865627|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_38|Kareniaceae\_25|Karenia\_17|mikimotoi  
 >AF274254|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_39|GymnodiniaceaelAmphidinium\_04|longum  
 >AF272049|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_39|Kareniaceae\_26|Karlodinium\_10|micrum  
 >EF492503|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_39|Kareniaceae\_27|Karenia\_18|brevis  
 >EF492506|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_40|Kareniaceae\_28|Karlodinium\_11|micrum  
 >AF352818|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_41|Kareniaceae\_29|Karenia\_19|brevis  
 >EF492502|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_41|Kareniaceae\_30|Karenia\_20|brevis  
 >EF036540|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_42|Kareniaceae\_31|Karlodinium\_12|micrum  
 >FN357291|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_43|Kareniaceae\_32|Karlodinium\_13|sp.  
 >EF492501|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_43|Kareniaceae\_33|Karenia\_21|brevis  
 >AY121855|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_44|GymnodiniaceaelGymnodinium\_08|sp.  
 >AF274262|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_45|Kareniaceae\_34|Karlodinium\_14|micrum  
 >AY245692|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_45|Kareniaceae\_35|Karlodinium\_15|micrum  
 >AF274259|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_45|Kareniaceae\_36|Karenia\_22|brevis  
 >AB001438|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_46|GymnodiniaceaelGyrodinium\_03|sp.  
 >AY800130|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_46|Kareniaceae\_37|Takayama\_02|sp.  
 >HM067010|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_46|Kareniaceae\_37|Takayama\_02|acrocha  
 >AF274260|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_47|GymnodiniaceaelGymnodinium\_09|sp.  
 >AJ415517|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_47|Kareniaceae\_38|Karenia\_23|mikimotoi  
 >AF272050|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_48|Kareniaceae\_39|Karlodinium\_16|micrum  
 >AB564298|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564299|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564300|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564301|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564302|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564303|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564304|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AB564305|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_49|PtychodiscaceaelHerdmaniallitoralis  
 >AF274261|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|dorsum  
 >AF274263|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|luncatenatum  
 >AY421786|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|linstriatum  
 >AY443015|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|linstriatum  
 >AY721981|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|linstriatum  
 >DQ084522|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|linstriatum  
 >DQ388457|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|luncatenatum  
 >DQ847433|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|linstriatum  
 >EF492495|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|luncatenatum  
 >EF492497|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|dorsum  
 >EF492498|EukaryotalAlveolatalPyrrophytodyphytalDinophyceaelGymnodiniales\_50|GymnodiniaceaelGyrodinium\_04|luncatenatum

>AB183672|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_51|GymnodiniaceaelAkashiwo\_01|sanguinea  
 >AJ415513|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_51|GymnodiniaceaelAkashiwo\_01|sanguinea  
 >AY421770|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_51|GymnodiniaceaelAkashiwo\_01|sanguinea  
 >AY421771|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_51|GymnodiniaceaelAkashiwo\_01|sanguinea  
 >AY831410|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_51|GymnodiniaceaelAkashiwo\_01|sanguinea  
 >AY831411|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_51|GymnodiniaceaelAkashiwo\_01|sanguinea  
 >AF276818|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >AY456106|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >AY456108|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >AY831412|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >DQ779987|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >DQ779988|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >EF492486|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >U41085|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_52|GymnodiniaceaelAkashiwo\_02|sanguinea  
 >AM503930|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_53|GymnodiniaceaelGyrodinium\_05|sp.  
 >FN669510|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_53|GymnodiniaceaelGyrodinium\_05|dominans  
 >FN669511|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_53|GymnodiniaceaelGyrodinium\_05|sp.  
 >AF022197|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_54|GymnodiniaceaelGymnodinium\_08|impudicum  
 >DQ779992|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_54|GymnodiniaceaelGymnodinium\_08|impudicum  
 >DQ779993|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_54|GymnodiniaceaelGymnodinium\_08|impudicum  
 >DQ785884|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_54|GymnodiniaceaelGymnodinium\_08|impudicum  
 >EU418974|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_54|GymnodiniaceaelGymnodinium\_08|impudicum  
 >GU362426|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_54|GymnodiniaceaelGymnodinium\_08|impudicum  
 >AF274256|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_55|GymnodiniaceaelAmphidinium\_05|semilunatum  
 >AB120004|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_56|GymnodiniaceaelGyrodinium\_06|helveticum  
 >AF274250|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_57|GymnodiniaceaelAmphidinium\_06|lasymmetricum  
 >HM066998|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_58|Kareniaceae\_40|Brachidinium|capitatum  
 >AY421779|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >AY421780|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >DQ779984|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >DQ779985|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >DQ779986|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418940|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418941|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418942|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418943|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418944|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418946|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418950|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418951|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418952|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418953|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418955|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418956|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418957|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418958|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418959|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides  
 >EU418960|EukaryotalAlveolatalPyrrophycohytalDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides

- >EU418961|EukaryotalAlveolatalPyrrophytalinDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides
- >EU418962|EukaryotalAlveolatalPyrrophytalinDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides
- >EU418963|EukaryotalAlveolatalPyrrophytalinDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides
- >EU418964|EukaryotalAlveolatalPyrrophytalinDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides
- >EU418965|EukaryotalAlveolatalPyrrophytalinDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides
- >EU418971|EukaryotalAlveolatalPyrrophytalinDinophyceaelGymnodiniales\_59|GymnodiniaceaelCochlodinium\_02|polykrikoides
- >AY443025|EukaryotalAlveolatalPyrrophytalinDinophyceaelLophodiales\_01|LophodiaceaelWoloszynskia\_01|leopoliensis
- >EF058253|EukaryotalAlveolatalPyrrophytalinDinophyceaelLophodiales\_02|LophodiaceaelWoloszynskia\_02|pascheri
- >GU355678|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_01|NoctiluaceaelAbedinium|dasypus
- >GU355679|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_01|NoctiluaceaelAbedinium|dasypus
- >GU355680|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_02|KofoidiniaceaelKofoidinium\_01|sp.
- >GU355681|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_03|KofoidiniaceaelKofoidinium\_02|sp.
- >GU355682|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_04|KofoidiniaceaelSpatulodinium\_01|sp.
- >GU355683|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_04|KofoidiniaceaelSpatulodinium\_02|pseudonoctiluca
- >GU355684|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_04|KofoidiniaceaelSpatulodinium\_02|pseudonoctiluca
- >GU355685|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_04|KofoidiniaceaelSpatulodinium\_02|pseudonoctiluca
- >AF022200|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_04|NoctiluaceaelNoctilucal|scintillans
- >DQ388461|EukaryotalAlveolatalPyrrophytalinDinophyceaelNoctilucales\_04|NoctiluaceaelNoctilucal|scintillans
- >AB033717|EukaryotalAlveolatalPyrrophytalinDinophyceaelOxyrrhinales|OxyrrhinaceaelOxyrrhis|marina
- >AF280077|EukaryotalAlveolatalPyrrophytalinDinophyceaelOxyrrhinales|OxyrrhinaceaelOxyrrhis|marina
- >AF330215|EukaryotalAlveolatalPyrrophytalinDinophyceaelOxyrrhinales|OxyrrhinaceaelOxyrrhis|marina
- >AF482425|EukaryotalAlveolatalPyrrophytalinDinophyceaelOxyrrhinales|OxyrrhinaceaelOxyrrhis|marina
- >AY566416|EukaryotalAlveolatalPyrrophytalinDinophyceaelOxyrrhinales|OxyrrhinaceaelOxyrrhis|marina
- >AY566417|EukaryotalAlveolatalPyrrophytalinDinophyceaelOxyrrhinales|OxyrrhinaceaelOxyrrhis|marina
- >AB181881|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_01|ProtoperidiniaceaelProtoperidinium\_01|labei
- >AB181882|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_01|ProtoperidiniaceaelProtoperidinium\_01|labei
- >AB181890|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_01|ProtoperidiniaceaelProtoperidinium\_01|denticulatum
- >AB181891|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_01|ProtoperidiniaceaelProtoperidinium\_01|denticulatum
- >AB181907|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_01|ProtoperidiniaceaelProtoperidinium\_01|thorianum
- >AB181908|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_01|ProtoperidiniaceaelProtoperidinium\_01|thorianum
- >AB181883|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lconicum
- >AB181884|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lconicum
- >AB181885|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lconicum
- >AB181886|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lconicum
- >AB181887|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lconicum
- >AB181888|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lcrassipes
- >AB181892|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|ldivergens
- >AB181893|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|ldivergens
- >AB181894|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lleonis
- >AB181895|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lleonis
- >AB181896|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lleonis
- >AB181897|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lleonis
- >AB181898|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lleonis
- >AB181899|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lpallidum
- >AB181900|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lpallidum
- >AB181901|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lpallidum
- >AB181902|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lpellucidum
- >AB181903|EukaryotalAlveolatalPyrrophytalinDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lpellucidum

>AB181904|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|punctulatum  
>AB181905|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|punctulatum  
>AB181906|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|punctulatum  
>AB255835|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|elegans  
>AB255836|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|elegans  
>AB255837|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|pentagonum  
>AB255838|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|pentagonum  
>AB261515|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|crassipes  
>AB261516|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|pallidum  
>AB261517|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|punctulatum  
>AB261518|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|thulesense  
>AB261519|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|thulesense  
>AB261520|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|thulesense  
>AB261521|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|thulesense  
>AB261522|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|thulesense  
>AB275355|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lexcentricum  
>AB284159|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|bipes  
>AY443020|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|conicum  
>AY443021|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|lexcentricum  
>AY443022|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|pellucidum  
>EF526884|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_02|ProtoperidiniaceaelProtoperidinium\_02|sp.  
>AB212091|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_03|OxytoxaceaelAmphidiniellalsedentaria  
>AF274257|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_04|GlenodiniopsidaceaelGlenodiniopsis|steinii  
>DQ317536|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_05|BlastodiniaceaelBlastodinium\_01|contortum  
>DQ317537|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_05|BlastodiniaceaelBlastodinium\_01|contortum  
>JN257680|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_05|BlastodiniaceaelBlastodinium\_01|contortum  
>ZZ00193632|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_05|BlastodiniaceaelBlastodinium\_01|contortum  
>AB246746|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_06|PeridiniaceaelPeridinium\_01|quinquecorne  
>Y10566|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_06|PeridiniaceaelPeridinium\_01|balticum  
>Y10567|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_06|PeridiniaceaelPeridinium\_01|foliaceum  
>JF790993|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_07|PeridiniaceaelPentapharsodinium\_01|tyrrhenicum  
>JF790989|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_08|PeridiniaceaelPentapharsodinium\_02|tyrrhenicum  
>AB255833|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_09|ProtoperidiniaceaelProtoperidinium\_03|claudicans  
>AB255834|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_09|ProtoperidiniaceaelProtoperidinium\_03|depressum  
>AB261512|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelDiplopsalis|lebourae  
>AB261513|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelDiplopsalis|bomba  
>AB261514|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelGotoius|lexcentricus  
>AB273721|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelOblealacanthocysta  
>AB273722|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelOblealacanthocysta  
>AB273723|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelOblealtorta  
>AB273724|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelOblealtorta  
>AB273725|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_10|ProtoperidiniaceaelOblealtorta  
>FJ888593|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_11|PodolampadaceaelBlepharocystalsp.  
>FJ888594|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_11|PodolampadaceaelPodolampaspalmipes  
>FJ888595|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_11|PodolampadaceaelPodolampaspibipes  
>FJ888596|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_11|PodolampadaceaelPodolampaslelegans  
>FJ888597|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_11|PodolampadaceaelPodolampaspinifera  
>AF521101|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_11|PodolampadaceaelRoscoffialcapitata

>FJ541189|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|pruvoti  
 >HQ226069|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|inornatum  
 >HQ226070|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|spinulosum  
 >HQ226071|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|spinulosum  
 >HQ226072|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|spinulosum  
 >ZZ00193633|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|spinifera  
 >ZZ00193634|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|spinifera  
 >ZZ00193635|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|inornatum  
 >ZZ00193636|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_12|BlastodiniaceaelBlastodinium\_02|sp.  
 >AF274249|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_13|Peridinales-XI|Adenoidesleludens  
 >EF492484|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_13|Peridinales-XI|Adenoidesleludens  
 >AF274270|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_14|PeridiniaceaelPentapharsodinium\_03|sp.  
 >AF022201|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_15|PeridiniaceaelPentapharsodinium\_04|tyrrhenicum  
 >AY443017|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelPeridinium\_02|polonicum  
 >AB183674|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_01|sp.  
 >AB183675|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_01|sp.  
 >AB183677|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_01|sp.  
 >AF274276|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_01|sweeneyae  
 >AF274277|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_01|trochoidea  
 >AJ415515|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_01|trochoidea  
 >AB183671|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_02|sp.  
 >AY743960|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_02|sp.  
 >DQ847435|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_02|precaria  
 >HM483396|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_03|trochoidea  
 >AY421792|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_04|trochoidea  
 >EF492513|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_04|trochoidea  
 >FR865630|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_04|sp.  
 >HQ845330|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_04|trochoidea  
 >HQ845331|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_04|sweeneyae  
 >JQ246506|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_16|PeridiniaceaelScrippsiella\_04|trochoidea  
 >AM494499|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_17|PeridiniaceaelScrippsiella\_06|sp.  
 >AY829528|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_18|PeridiniaceaelGlenodinium|sp.  
 >GU647178|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_18|PeridiniaceaelGlenodinium|sp.  
 >JN257672|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_19|BlastodiniaceaelBlastodinium\_03|sp.  
 >ZZ00193637|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_19|BlastodiniaceaelBlastodinium\_03|sp.  
 >AB353770|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_20|PeridiniaceaelPeridiniopsis\_01|sp.  
 >AF274271|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_21|PeridiniaceaelPeridinium\_03|lumbonatum  
 >FR865631|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_21|PeridiniaceaelPeridinium\_03|inconspicuum  
 >AB195668|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_21|Peridinales-XI|Galeidinium|rugatum  
 >AB183670|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|triquetra  
 >AF022198|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|triquetra  
 >AF274267|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|rotundata  
 >AJ415514|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|triquetra  
 >AY421787|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|triquetra  
 >DQ388464|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|rotundata  
 >GU594638|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|triquetra  
 >JN098193|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_22|HeterocapsaceaelHeterocapsa\_01|rotundata  
 >FJ541187|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_23|BlastodiniaceaelBlastodinium\_04|galatheanum



>FJ541188IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_23IBlastodiniaceaelBlastodinium\_04Igalatheanum  
>AB185114IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Icinctum  
>AB232669IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Iwillei  
>AF022202IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Isp.  
>AF231805IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Ibipes  
>AF274272IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Iwillei  
>AF274280IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Iwillei  
>DQ166208IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Igatunense  
>DQ166209IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Icinctum  
>DQ166210IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Isp.  
>DQ166211IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Isp.  
>DQ487197IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Isp.  
>DQ980482IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Ilimbatum  
>EF058242IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Ibipes  
>EF058243IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Icinctum  
>EF375879IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_24IperidiniaceaelPeridinium\_04Iwillei  
>AB353771IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_25IperidiniaceaelPeridiniopsis\_02Ipenardii  
>ZZ00193631IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_26IBlastodiniaceaelBlastodinium\_05Ioviforme  
>AF521100IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_27ILessardiaceaelLessardialelongata  
>AF274265IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_28IHeterocapsaceaelHeterocapsa\_02Iniei  
>AF274266IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_28IHeterocapsaceaelHeterocapsa\_02Ipygmaea  
>EF492499IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_28IHeterocapsaceaelHeterocapsa\_02Iniei  
>EF492500IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_29IHeterocapsaceaelHeterocapsa\_03Ipygmaea  
>EF492509IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_30IperidiniaceaelPeridinium\_05Isociale  
>FR877580IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_30IperidiniphycidaelAzadinium\_01Isp.  
>U52352IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Isp.  
>U52353IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Isp.  
>U52354IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Isp.  
>U52355IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Isp.  
>U52356IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Isp.  
>U52357IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Inutricula  
>U52911IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_31IperidiniaceaelScrippsiella\_07Isp.  
>AY238479IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_32IOxytoxaceaelThecadinium\_01Idragescoi  
>AB564308IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_32IProtooperidiniaceaelProtooperidinium\_04Iminutum  
>AB564309IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_32IProtooperidiniaceaelProtooperidinium\_04Iminutum  
>DQ975473IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_33IOxytoxaceaelSabulodiniumlundulatum  
>DQ975474IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_33IOxytoxaceaelSabulodiniumlundulatum  
>AY238477IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_34IOxytoxaceaelThecadinium\_02Imucosum  
>DQ388458IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_34IOxytoxaceaelThecadinium\_02Iinclinatum  
>EF492515IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_34IOxytoxaceaelThecadinium\_02Iinclinatum  
>JN680857IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_35IperidiniphycidaelAzadinium\_02Ispinosum  
>AY590478IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_36IPfiesteriaceaelLuciella\_01Isp.  
>AY443018IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_37IperidiniaceaelPeridinium\_06Iwierzejskii  
>AF231803IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_38IperidiniaceaelDurinskialbaltica  
>GU999528IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_38IperidiniaceaelDurinskialbaltica  
>AB183639IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_39IHeterocapsaceaelHeterocapsa\_04Isp.  
>AF033865IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_39IHeterocapsaceaelHeterocapsa\_04Isp.  
>EF492492IEukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_39IHeterocapsaceaelHeterocapsa\_04Isp.

>EF492494|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_39|Heterocapsaceae|Heterocapsa\_04|sp.  
 >FJ549370|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_39|Heterocapsaceae|Heterocapsa\_04|sp.  
 >HQ324897|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_40|Peridiniphytidae|Azadinium\_03|poporum  
 >HQ324898|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_40|Peridiniphytidae|Azadinium\_03|poporum  
 >HQ324899|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_40|Peridiniphytidae|Azadinium\_03|poporum  
 >JF791029|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_41|Peridiniaceae|Pentapharsodinium\_05|tyrrhenicum  
 >JF791053|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_41|Peridiniaceae|Pentapharsodinium\_05|tyrrhenicum  
 >AB246744|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_42|Peridiniaceae|Peridinium\_07|lunquecorne  
 >AF231804|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_43|Peridiniaceae|Kryptoperidinium|foliaceum  
 >AF274268|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_43|Peridiniaceae|Kryptoperidinium|foliaceum  
 >DQ847436|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_43|Peridiniaceae|Kryptoperidinium|foliaceum  
 >EF492508|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_43|Peridiniaceae|Kryptoperidinium|foliaceum  
 >AJ841809|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_44|Pfiesteriaceae|Stoeckeria\_01|lalgida  
 >AY238478|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_45|Oxytoxaceae|Thecadinium\_03|kofoidii  
 >GU295204|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_45|Oxytoxaceae|Thecadinium\_03|kofoidii  
 >AY664984|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >AY664985|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >AY664986|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >DQ317538|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|navicula  
 >JN257674|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|mangini  
 >JN257676|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|mangini  
 >JN257677|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >JN257678|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >JN257679|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >ZZ00193620|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >ZZ00193623|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|mangini  
 >ZZ00193625|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >ZZ00193628|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >ZZ00193629|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_46|Blastodiniaceae|Blastodinium\_06|sp.  
 >AY970653|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_47|Peridiniaceae|Peridinium\_08|laciculiferum  
 >AY970662|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_47|Peridiniaceae|Scrippsiella\_08|hangoei  
 >AJ968729|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Peridinales-XIPaulsenella|vonstoschii  
 >DQ991380|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_01|brodyi  
 >DQ991372|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >DQ991373|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >DQ991374|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >DQ991375|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >DQ991376|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >DQ991378|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >DQ991379|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_02|brodyi  
 >AF080097|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_03|brodyi  
 >DQ991377|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_04|brodyi  
 >AY590476|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Cryptoperidiniopsis\_05|brodyi  
 >AM050344|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Luciella\_02|sp.  
 >AM050345|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Luciella\_02|sp.  
 >AY033487|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Luciella\_02|sp.  
 >AY245689|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Luciella\_02|sp.  
 >AY251289|EukaryotalAlveolatalPyrrophytodyta|Dinophyceae|Peridinales\_48|Pfiesteriaceae|Luciella\_02|sp.

>AY590482|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelLuciella\_02|sp.  
 >AY590483|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelLuciella\_02|sp.  
 >AY590477|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelLuciella\_03|sp.  
 >AF077055|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF080098|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialshumwayae  
 >AF149793|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF218805|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialshumwayae  
 >AF330600|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330601|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330602|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330603|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330604|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330605|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330606|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330607|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330608|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330609|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330610|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330611|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330612|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330613|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330614|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330615|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330616|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330617|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330618|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330619|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AF330620|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AM231028|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AM231029|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AM231030|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AM231031|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AM231032|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AM231033|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AY033488|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AY112746|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >AY121846|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >DQ991381|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >DQ991382|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelPfiesterialpiscicida  
 >FN557541|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPeridinales\_48|PfiesteriaceaelStoeceria\_02|sp.  
 >L13716|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPhytodinales\_01|GloeodiniaceaelGloeodinium\_01|lviscum  
 >EF058235|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPhytodinales\_02|PhytodiniaceaelCystodinium|phaseolus  
 >EF058251|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPhytodinales\_02|PhytodiniaceaelPhytodinium|sp.  
 >EF058238|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPhytodinales\_03|GloeodiniaceaelGloeodinium\_02|montanum  
 >DQ388460|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_01|ProrocentraceaelProrocentrum\_01|cassubicum  
 >GU327677|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_02|ProrocentraceaelProrocentrum\_02|pseudopaname  
 nse  
 >Y16233|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_02|ProrocentraceaelProrocentrum\_02|panamensis

>AB189773|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189774|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189775|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189776|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189777|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189778|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189779|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >AB189780|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >DQ238043|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lvis  
 >EU287486|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|isp.  
 >EU287488|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|isp.  
 >FJ842379|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lconsumum  
 >Y16234|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|larenarium  
 >Y16235|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lima  
 >Y16236|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lmaculosum  
 >Y16237|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_03|ProrocentraceaelProrocentrum\_03|lconconvum  
 >EF017804|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_04|ProrocentraceaelProrocentrum\_04|lminimum  
 >AY551273|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_05|ProrocentraceaelProrocentrum\_05|lidentatum  
 >EF657885|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_06|ProrocentraceaelProrocentrum\_06|ltsawwassenens  
 e  
 >AY551272|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_07|ProrocentraceaelProrocentrum\_07|ldonghaiense  
 >AJ841810|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_08|ProrocentraceaelProrocentrum\_08|ldonghaiense  
 >Y16238|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_08|ProrocentraceaelProrocentrum\_08|lminimum  
 >rs10857|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_08|ProrocentraceaelProrocentrum\_08|lidentatum  
 >DQ028763|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_09|ProrocentraceaelProrocentrum\_09|lminimum  
 >AY421791|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >AY803740|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >AY803741|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >AY803742|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lidentatum  
 >AY803743|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|ldonghaiense  
 >DQ336054|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|ldonghaiense  
 >DQ336057|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lidentatum  
 >DQ336060|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >DQ336063|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >DQ336066|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >DQ336069|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >DQ336072|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >EU780639|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >FJ587221|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_10|ProrocentraceaelProrocentrum\_10|lminimum  
 >AJ415520|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_11|ProrocentraceaelProrocentrum\_11|lminimum  
 >AB183673|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_12|ProrocentraceaelProrocentrum\_12|ltriestinum  
 >AY443019|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_12|ProrocentraceaelProrocentrum\_12|lgracile  
 >DQ004734|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_12|ProrocentraceaelProrocentrum\_12|ltriestinum  
 >EF492512|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_12|ProrocentraceaelProrocentrum\_12|ltriestinum  
 >GU327678|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_12|ProrocentraceaelProrocentrum\_12|lglenanicum  
 >GU327679|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_12|ProrocentraceaelProrocentrum\_12|lglenanicum  
 >AF238264|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_13|ProrocentraceaelProrocentrum\_13|lminimum  
 >EU196418|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProrocentrales\_14|ProrocentraceaelProrocentrum\_14|lemarginatum

>EU196420|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_14|ProcentraceaelProcentrum\_14|fukuyoi  
>Y16239|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_14|ProcentraceaelProcentrum\_14|lemarginatum  
>AJ415519|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>AY585526|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>AY803739|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>AY833514|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>DQ004735|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>EF492510|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imexicanum  
>EF492511|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>EU287485|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imexicanum  
>EU287487|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Irhathymum  
>EU780638|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>FJ842096|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Irhathymum  
>M14649|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imicans  
>Y16232|EukaryotalAlveolatalPyrrophycohytalDinophyceaelProcentrales\_15|ProcentraceaelProcentrum\_15|Imexicanum  
>AF022156|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPyrocystales\_01|PyrocystaceaelPyrocystis|Inocluca  
>AF274274|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPyrocystales\_01|PyrocystaceaelPyrocystis|Inocluca  
>FR865628|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPyrocystales\_01|PyrocystaceaelPyrocystis|Inocluca  
>FR865629|EukaryotalAlveolatalPyrrophycohytalDinophyceaelPyrocystales\_01|PyrocystaceaelPyrocystis|Inocluca  
>AB275020|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelBiecheleria|Inocluca  
>EF058252|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelBiecheleria|Inocluca  
>FR690459|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelBiecheleria|Inocluca  
>EF417317|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelPolarella|Inocluca  
>EF434275|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelPolarella|Inocluca  
>EF434276|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelPolarella|Inocluca  
>EF434277|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_01|SuessiaceaelPolarella|Inocluca  
>AF099183|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_02|SuessiaceaelPolarella|Inocluca  
>GQ375263|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_02|SuessiaceaelPolarella|Inocluca  
>EF492490|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelPelagodinium|Inocluca  
>U37365|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelPelagodinium|Inocluca  
>U37366|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelPelagodinium|Inocluca  
>U37367|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelPelagodinium|Inocluca  
>U37406|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelPelagodinium|Inocluca  
>U41087|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelPelagodinium|Inocluca  
>DQ388466|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelProtodinium|Inocluca  
>EF492491|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelProtodinium|Inocluca  
>EF492493|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelProtodinium|Inocluca  
>JF791031|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelProtodinium|Inocluca  
>U41086|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_03|SuessiaceaelProtodinium|Inocluca  
>AB016538|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016572|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016573|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016574|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016575|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016576|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016577|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016578|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca  
>AB016579|EukaryotalAlveolatalPyrrophycohytalDinophyceaelSuessiales\_04|SymbiodiniaceaelSymbiodinium|Inocluca

>AB016580|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016581|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016593|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016594|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016595|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016596|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016597|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016722|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016723|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB016724|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB085911|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB085912|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB085913|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB085914|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB085915|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AB183640|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF182822|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF225965|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|California  
 >AF238256|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF238257|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF238258|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF238261|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF274279|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AF379641|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AY160124|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AY165766|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AY443023|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >AY456111|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Microadriaticum  
 >AY456113|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >EF036539|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Goreau  
 >EF492496|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Microadriaticum  
 >EF492514|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Microadriaticum  
 >GU362427|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >HM067608|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >HM067611|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >HM067612|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >HM067613|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >JQ320136|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >L13717|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Coccolithum  
 >L13718|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Meandrinae  
 >M88509|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >M88521|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Microadriaticum  
 >U10893|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Sp.  
 >X62650|EukaryotalAlveolatalPyrrophytalinophytaDinophyceaeSuessiales\_04|SymbiodiniaceaeSymbiodinium|Pilosum  
 >AF274278|EukaryotalAlveolatalPyrrophytalinophytaDinophyceae|Thraosphaerales|Thraosphaeraceae|Thraosphaeraheimii  
 >AY665079|EukaryotalAlveolatalPyrrophytalinophytaDinophyceae|Uncultured\_01|Uncultured-XI|Uncultured-XX|Dinoflagellate  
 >AY665080|EukaryotalAlveolatalPyrrophytalinophytaDinophyceae|Uncultured\_01|Uncultured-XI|Uncultured-XX|Dinoflagellate  
 >AY665081|EukaryotalAlveolatalPyrrophytalinophytaDinophyceae|Uncultured\_01|Uncultured-XI|Uncultured-XX|Dinoflagellate



Submitted in *Environmental Microbiology*

## Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding

Noan Le Bescot<sup>1</sup>, Frédéric Mahé<sup>2</sup>, Stéphane Audic<sup>1</sup>, Dimier Céline<sup>1</sup>, Marie-José Garet<sup>1</sup>, Mikael Katinka<sup>3</sup>, Patrick Wincker<sup>3</sup>, Colombran de Vargas<sup>1</sup>, Raffaele Siano<sup>4</sup>

<sup>1</sup> CNRS UMR 7144 & Université Pierre & Marie Curie, EPEP - Evolution of Protists and Pelagic Ecosystems, Station Biologique de Roscoff, 29682 Roscoff, France

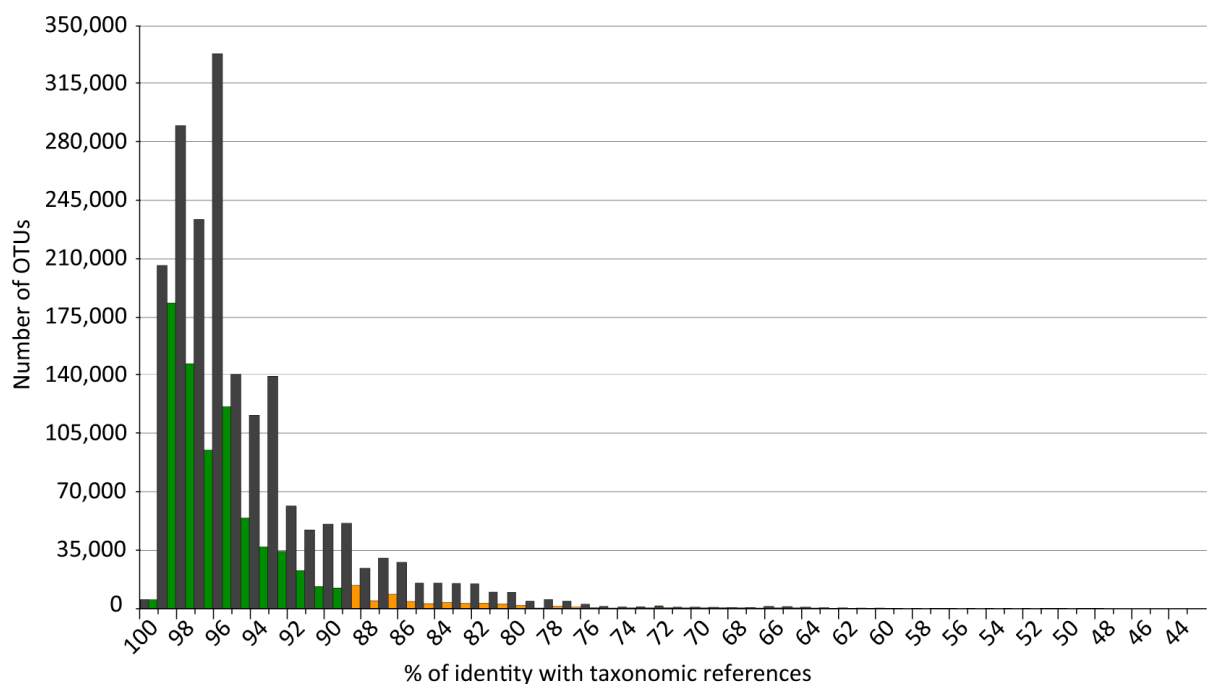
<sup>2</sup> Department of Ecology Technische Universität Kaiserslautern Erwin-Schrödinger Str. 14 D-67663 Kaiserslautern, Germany

<sup>3</sup> Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux CP5706 91057 Evry cedex, France

<sup>4</sup> IFREMER, Centre de Brest, DYNECO/Pelagos, BP 70, 29280 Plouzané, France

**Correspondence:** Raffaele Siano, raffaele.siano@ifremer.fr. Tel: 00 33 2 98 22 42 04  
Colombran de Vargas, vargas@sb-roscoff.fr, Tel: 00 33 2 98 29 25 28

### Supplementary (online)



**Figure S1:** Number of dinoflagellates OTU in relation to their percentage of identity (%ID) with taxonomic references. The gray bars represent the original dataset. The green + yellow bars represent the data base conserved after eliminating sequences present in less than 10 copies; the green bars only represent the final dataset obtained keeping only sequences with %ID  $\geq$  90% to any dinoflagellate reference sequence.



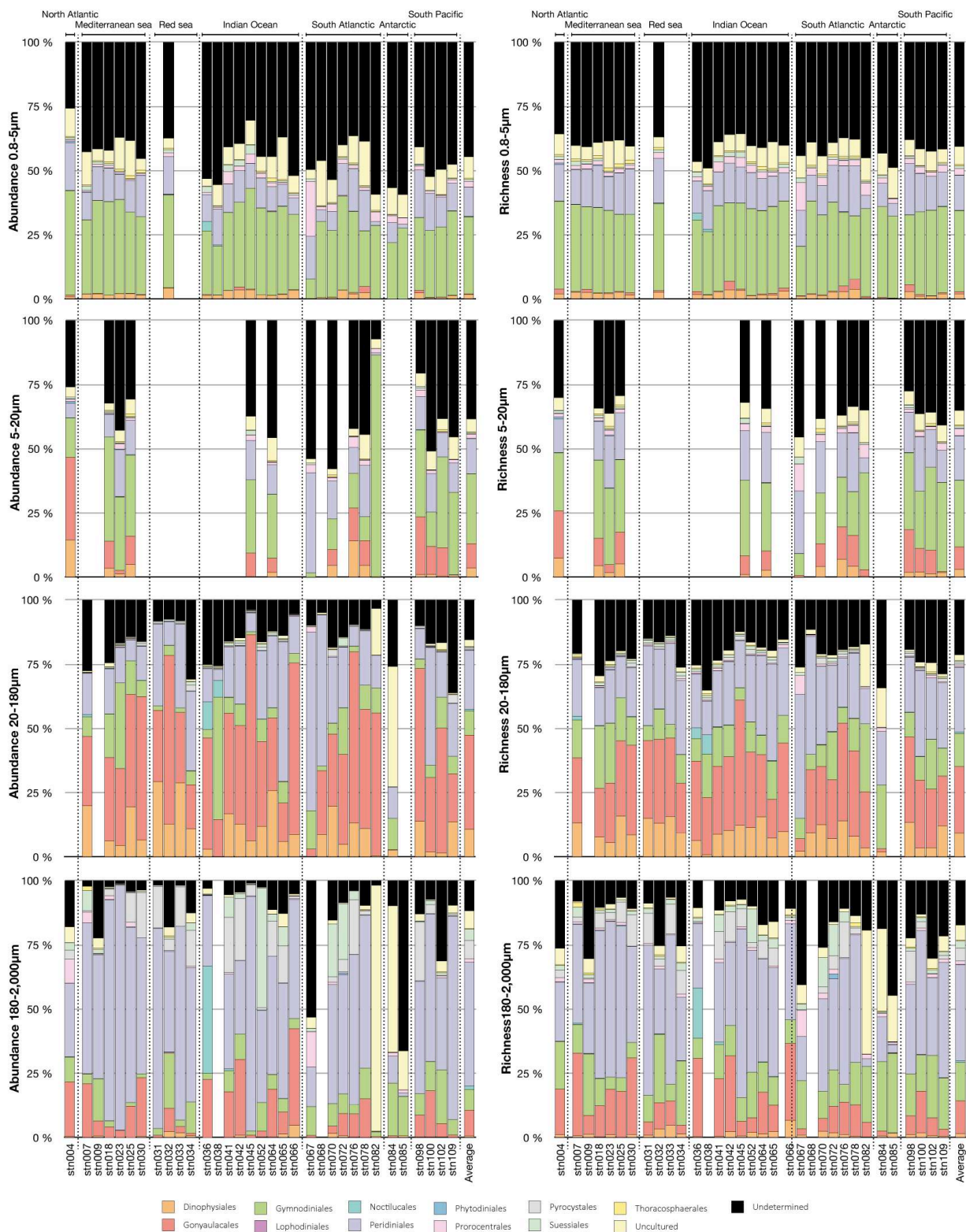
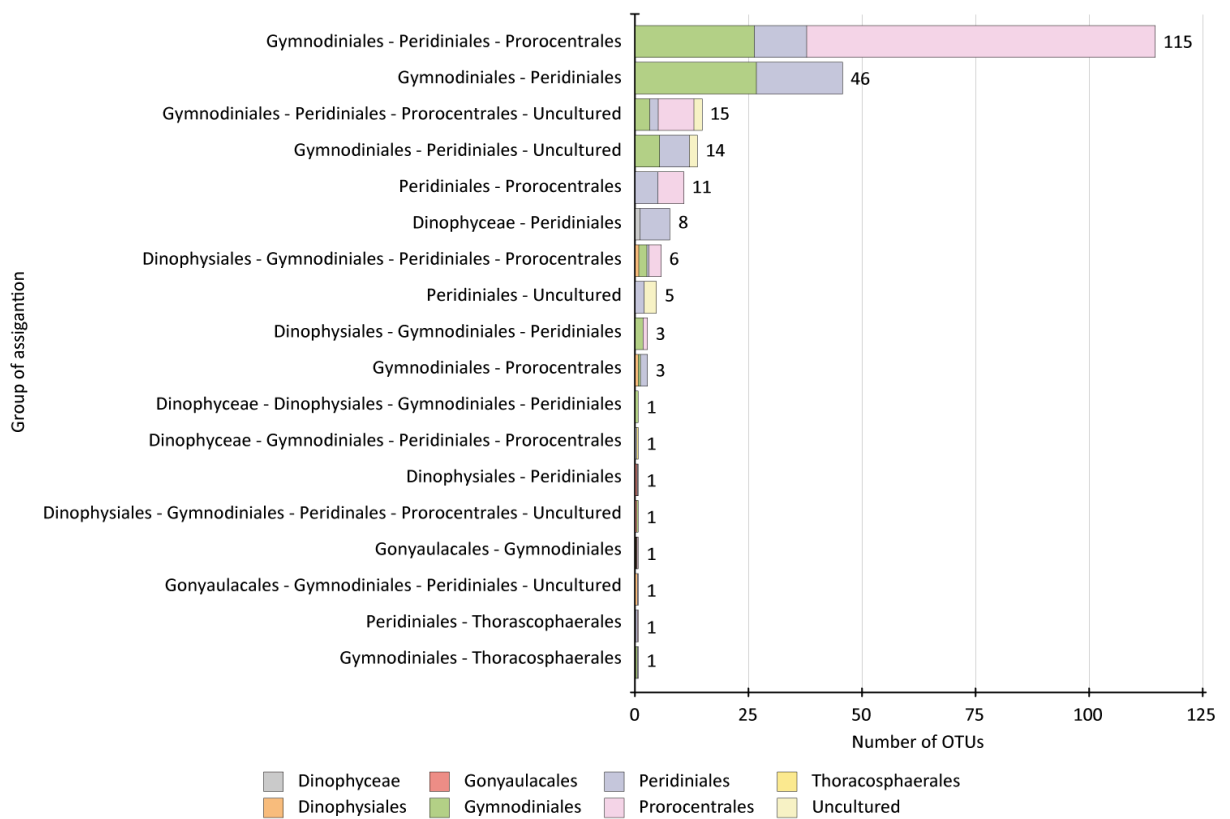


Figure S2: OTUs abundances and richnesses assigned to the different dinoflagellate orders across size fractions per sampling station.



**Figure S3:** Frequency of conflict assignment for undetermined dinoflagellates in size fraction 0.8-5µm. The relative importance of orders is represented in each bar.

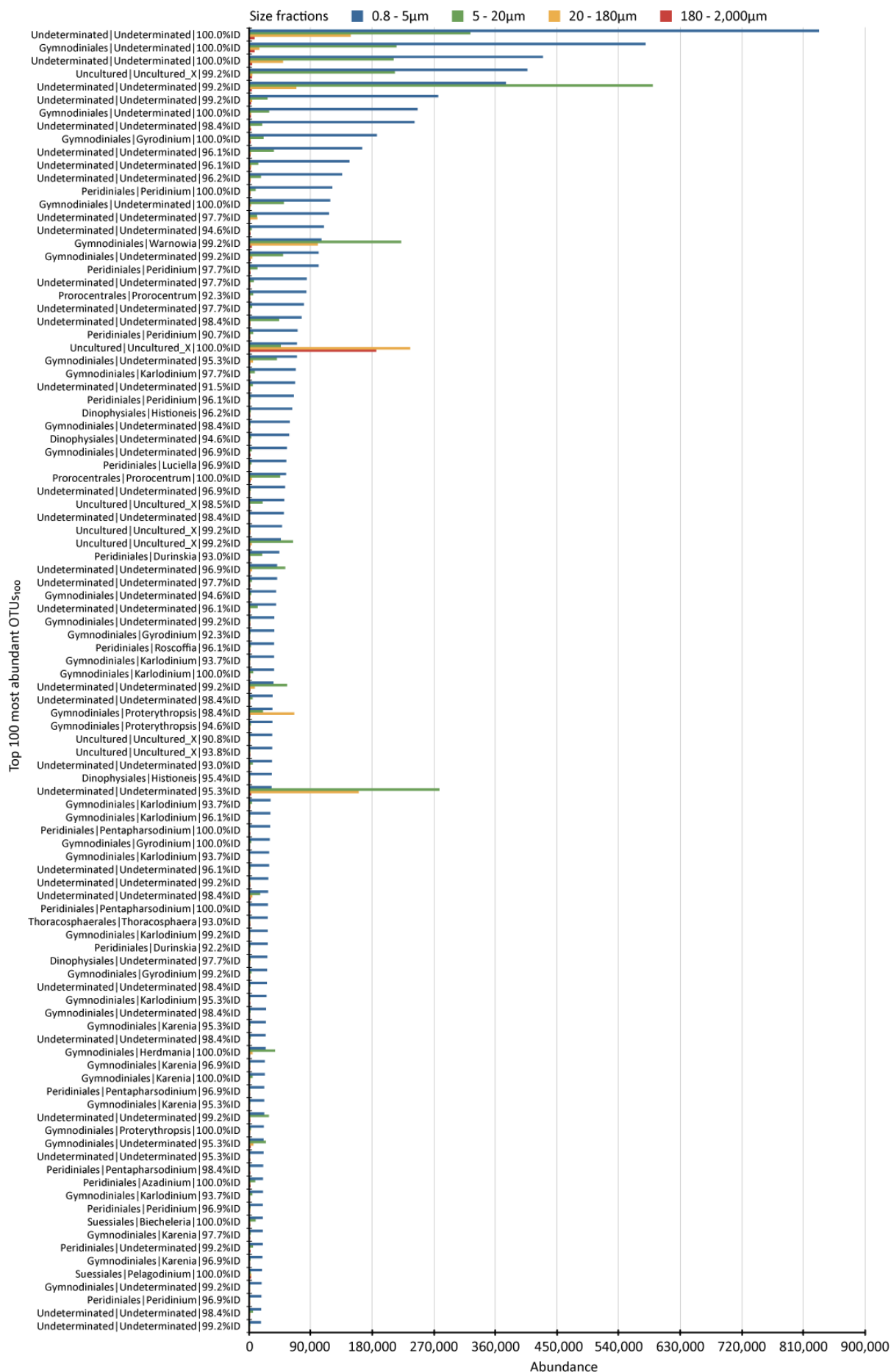


Figure S4: Abundances of top 100 OTUs (Orders/Genus/%ID) found in fraction 0.8-5µm across all size fractions.



## Abstract

Dinoflagellates (Alveolata, Dinophyceae) are a group of protists that first appeared over 400 million years ago, and contain  $\pm 2,400$  living species and  $\pm 2,500$  fossil species. About half of the current dinoflagellate species are photosynthetic and contribute significantly to the primary production of aquatic ecosystems, whereas other species acquire their carbon through a spectacular diversity of ecological relationships: symbiosis *sensu lato*, phagotrophy, mixotrophy). Dinoflagellates have a unique and phenomenal diversity of morphology, physiology, and complex behaviors that allow them to adapt to aquatic marine and freshwater environments. This bio-complexity has made it very difficult to perform integrative studies across the entire group and to establish general ecological or adaptive strategies. The main objective of this thesis was to search for global biodiversity patterns and structure of planktonic dinoflagellate communities across the world's oceans. We first participated in the development of a morphogenetic sampling protocol for all planktonic dinoflagellates, covering their entire spectrum of cell size from 3 to 500 $\mu\text{m}$ , and a significant portion of their spatio-temporal niches in the photic zone of global oceans (Not et al., submitted). This homogeneous and robust protocol was designed to operate on a sailboat, and was applied to each of the 154 sampling points of the circum-global Tara-Oceans planktonic ecosystems sampling project (2009-2012) (Karsenti et al., 2011). The first part of this work was to not only implement the new protocol directly on the boat, but also to explore various high-throughput analyses of the biodiversity of pelagic marine dinoflagellates within Tara-Oceans samples. A first approach, based on the automatic FlowCAM imaging technology carried aboard Tara, allowed assessment of diversity, relative abundance, and geographical distribution of dinoflagellate morpho-species within the genus *Neoceratium* from the Mediterranean Sea (Le Bescot et al., in prep). A second approach was the use of environmental ribosomal DNA fragment ("metabarcodes"), amplified from nucleic acids extracted from seawater, as markers of dinoflagellate biodiversity. The building of a reference database of dinoflagellate DNA barcodes (*DinR2* — Dinoflagellate Ribosomal Reference database), incorporated into a broader database for all marine protists (*PR2* — Protist Ribosomal Reference database) (Guillou et al, 2013), allowed us to assign a traditional taxonomy to meta-barcodes, thus linking classical knowledge of dinoflagellate biodiversity to our analysis based on high-throughput sequencing. We then explored whether general patterns in the structure of dinoflagellate biodiversity emerged across size classes and organismal and ecological niches (Le Bescot et al., submitted). The abundance of different orders of dinoflagellates varies primarily between size classes (pico-, nano-, micro- and meso-plankton), independently of the ecosystem studied and the sampling period. A remarkable and unexpected diversity of pico-dinoflagellates (< 5 $\mu\text{m}$ ) was identified in the order Gymnodiniales, with a large number of barcodes remaining to be identified. The meso-plankton fraction (180-2,000 $\mu\text{m}$ ), meanwhile, contained a high diversity and abundance of parasitic and photo-symbiotic genera of dinoflagellates, which play a key functional role in the maintenance of plankton communities (Le Bescot et al., submitted). Finally, a preliminary study of the impact of environmental conditions on the meta-community of dinoflagellates identified factors affecting the composition of dinoflagellate communities and their functional structure between the photic zone and the meso-pelagic zone (Le Bescot et al., in prep).

**Keywords:** plankton, dinoflagellate, meta-barcode, biodiversity, molecular ecology, global ocean, Tara-Oceans