



HAL
open science

Phylogénie, diversité et dynamique temporelle chez les ciliés tintinnidés marins

Charles Bachy

► **To cite this version:**

Charles Bachy. Phylogénie, diversité et dynamique temporelle chez les ciliés tintinnidés marins. Sciences agricoles. Université Paris Sud - Paris XI, 2012. Français. NNT : 2012PA112105 . tel-00769949

HAL Id: tel-00769949

<https://theses.hal.science/tel-00769949>

Submitted on 4 Jan 2013

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

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

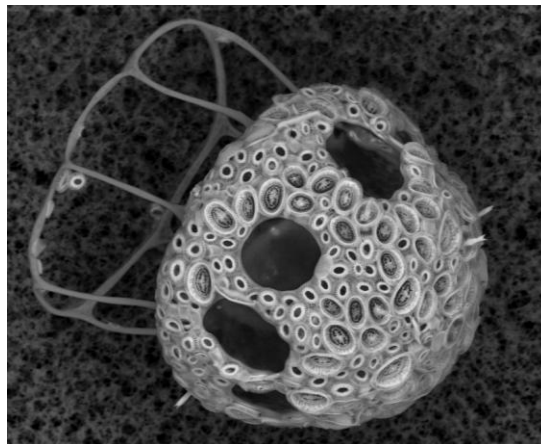
Thèse de doctorat de l'Université Paris-Sud (Paris 11)

École doctorale Gènes, Génomes, Cellules ED 426
Discipline : Microbiologie

Présentée par :
Charles BACHY

Pour obtenir le grade de :
Docteur ès sciences de l'Université Paris-Sud

Phylogénie, diversité et dynamique temporelle chez les ciliés tintinnidés marins



Thèse soutenue en juillet 2012, devant le jury composé de :

Laure GUILLOU
Jan PAWLOWSKI
Gwenaël PIGANEAU
Didier CASANE
John R. DOLAN
David MOREIRA
Purificación LOPEZ-GARCIA

Rapporteur
Rapporteur
Examinateur
Examinateur
Examinateur
Co-directeur de thèse
Co-directeur de thèse

Résumé

La diversité des protistes marins planctoniques, après avoir été historiquement étudiée sur des critères morphologiques, est depuis récemment sujette à une intense recherche à l'aide d'approches moléculaires. Notamment, les études basées sur l'amplification directe de marqueurs moléculaires à partir d'ADN environnemental ont révélées une exceptionnelle diversité, parfois nouvelle et non-caractérisée. Cependant, l'étendue de cette diversité reste encore peu comprise et ouvre de nombreux débats sur son interprétation. L'objectif central de ce travail est d'améliorer notre compréhension sur le lien existant entre la connaissance classique des eucaryotes unicellulaires et leur diversité estimée à partir des données moléculaires, en particulier pour une meilleure interprétation des processus évolutifs et écologiques. Pour cela, nous avons utilisé comme système modèle l'ordre des ciliés tintinnidés (Tintinnida) qui constituent un groupe riche en espèces, aisément identifiables au microscope grâce à leur coquille externe (lorica) et communément rencontrés dans l'ensemble des eaux marines et lacustres du monde. Un suivi approfondi sur deux ans des tintinnidés de la Baie de Villefranche-sur-Mer (Mer Méditerranée, France), couplant des analyses moléculaires de la diversité à partir de cellules individuelles et à partir d'ADN environnemental, a permis de caractériser la composition de ces communautés et leur dynamique temporelle aux échelles macro- et micro-évolutives. La première partie de ce travail a été destinée à la réalisation d'une phylogénie moléculaire de référence pour les tintinnidés en incorporant les séquences de 62 individus de morphologies diverses pour lesquelles des données moléculaires n'étaient pas disponibles. Nous avons amplifié et séquencé les gènes codant pour les ARN ribosomiques (ARNr 18S, 5.8S et 28S) et les espaces intergéniques correspondants (ITS1 et ITS2). La classification taxonomique a été réévaluée d'après les données moléculaires. Dans un deuxième temps, nous avons testé l'efficacité des approches moléculaires conventionnelles (amplification, clonage et séquençage Sanger du gène de l'ARNr 18S) et plus récentes (amplification et pyroséquençage de régions de l'ARNr 18S et de l'ITS), pour décrire la composition des communautés des tintinnidés dans des échantillons environnementaux en les comparant avec des estimations de la diversité par observation morphologique sur les mêmes échantillons. Si il existe de légères incongruences entre les approches et/ou les différents marqueurs employés, les approches culture-indépendantes s'avèrent efficaces pour décrire la diversité morphologique. En revanche, afin de ne pas surévaluer artificiellement le nombre d'espèces estimées à partir des données de pyroséquençage, il faut que des méthodes de débruitage et de regroupement en unités taxonomiques opérationnelles (UTOs) contraignantes soient appliquées. La troisième partie de ce travail a été dédiée au suivi temporel des communautés de tintinnidés à différentes profondeurs dans la baie de Villefranche, basé sur le clonage et le séquençage du gène de l'ARNr 18S et des régions ITS. Il apparaît des différences de distribution au cours de l'année à une même profondeur, en particulier en termes d'abondance de séquences pour une UTO donnée. Malgré un cadre phylogénétique solide et assez enrichi, l'approche moléculaire révèle des séquences éloignées des espèces déjà séquencées. La découverte de ces clades environnementaux souligne potentiellement l'importance écologique d'espèces encore mal connues. Enfin, le séquençage direct du gène de l'ARNr 18S et de l'ITS2 à partir des cellules individuelles de l'espèce *Undella claparedei* a offert l'opportunité d'une étude populationnelle sur une période de deux ans. La diversité intra-spécifique mesurée met à jour des phénomènes d'hybridation entre variantes génétiques. Une structuration génétique temporelle a également été observée pour le gène de l'ARNr 18S. Les implications de ces différentes recherches sont discutées dans le cadre de l'étude de la diversité et de l'écologie des tintinnidés, et plus largement, des protistes marins.

Abstract

The marine protistan diversity has been historically studied based on morphological characterization but has recently been the object of intense research using molecular approaches. Studies based on the amplification of molecular markers from environmental DNA revealed an outstanding diversity, partly new and uncharacterized. However, the actual extent of this diversity remains poorly known and highly debated. The main goal of this work was to improve our knowledge on protistan diversity to bridge the gap between molecular environmental surveys and classical protistology to better understand the ecology and evolution of unicellular eukaryotes. For this purpose, we used as a model the species-rich order of the tintinnid ciliates (Tintinnida, Ciliophora), which are easily distinguishable because of their secreted shell, the lorica, and commonly found in marine waters all around the globe. A two-year monitoring of the tintinnid populations in the Bay of Villefranche-sur-Mer (Mediterranean Sea, France), combining molecular analyses of the diversity based on single-cells and environmental DNA, gave us the opportunity to describe the tintinnid community composition and its temporal dynamics. In the first part of this work, we constructed a reference molecular phylogeny for the tintinnids including new sequences from 62 specimens of diverse morphologies, for which we amplified and sequenced the ribosomal coding genes (18S, 5.8S and 28S rRNA) and the corresponding intergenic spacers (ITS1 and ITS2). The taxonomic classification of the Tintinnida has been revised based on these molecular data. In the second part, in order to assess the accuracy of molecular-based approaches to describe the natural species assemblages of tintinnids, we compared the morphology-based diversity estimates with those derived from classical (amplification, cloning and Sanger sequencing of the 18S rRNA gene) and more recent (direct pyrosequencing of amplified 18S rRNA genes and ITS regions) molecular approaches. Even if there are still some disagreements between the different methods and/or molecular markers, the culture-independent approaches were efficient to describe the morphological diversity. However, a careful and rigorous analysis of pyrosequencing datasets, including sequence denoising and stringent sequence clustering in Operational Taxonomic Units (OTUs) with well-adjusted parameters, is necessary to avoid overestimating the species number. The third part of the thesis is dedicated to the study of the genetic diversity of tintinnids over a one-year survey in the Bay of Villefranche at five different depths by combining community fingerprinting analysis using denaturing gradient gel electrophoresis (DGGE) with direct PCR amplification and sequencing of 18S, 5.8S, and 28S rRNA genes and ITS regions. These analyses revealed marked seasonal changes, in particular in the sequence abundances of certain OTUs. In addition, despite an enriched phylogenetic reference sequence dataset for the tintinnids, we retrieved two abundant phylotypes without any closely related known species, highlighting the possible ecological relevance of unidentified species. Finally, we studied the intra-specific diversity of populations of the species *Undella claparedei* based on 18S rDNA and ITS direct sequencing of single-cells collected over a period of two years. We detected signals of hybridization and sexual recombination among different genetic variants. We also found genetic structuring of the 18S rRNA gene data differentiating populations collected at different times. The implications of all these results are discussed in the framework of the diversity and ecology of tintinnid ciliates and, more generally, of marine protists.

Remerciements

Je tiens à exprimer ici ma profonde reconnaissance à toutes les personnes qui, de près ou de loin, m'ont apporté leur soutien et leur confiance tout au long de cette aventure.

Aux hasards de la vie et des rencontres ...

A Céline, Thierry et Gwenaël, pour m'avoir mis sur les rails de la vie scientifique et poussé à continuer sur cette voie.

*A Puri, pour m'avoir accordé ta confiance du début jusqu'à la fin de ces 4 ans de thèse.
A David, pour m'avoir également accordé ta confiance du début jusqu'à la fin de cette thèse.*

Vous m'avez offert un cadre de travail rêvé, libre de mes choix, solide scientifiquement, et humainement génial.

Aux membres du laboratoire ESE, vous savez rendre l'environnement convivial et vos connaissances sont tellement vastes qu'on y apprend tous les jours quelque chose. A Hervé, Benjamin, Pierre, Gwendal, Amandine, Julien, Claire, Alexis, Alex, Yann, Elodie, Tatiana, Laetitia, Jacqui, Juan, Nathalie, JT, Lucie, Christian, Martha, Marina et Juergen.

A John, pour m'avoir offert l'opportunité de travailler à Villefranche-sur-Mer sur ces petites merveilles de la nature que l'on ne présente plus. A Jean-Yves, capitaine normand, pour m'avoir laissé la barre et accompagné pendant ces deux ans. A Corrine et Marie-Do pour m'avoir facilité la tâche. A Fernando, ses infos précieuses sur les protistes.

A Davide, grazie per la tua amicizia.

Aux membres de l'équipe microbio, passés et présents, Achim, William, Quique, Ludwig, Philippe, Marie, Céline, Estelle, Jonathan, Marianne, Pao. Je sais que l'on n'est pas beaucoup pour explorer le monde du vivant, mais on va y arriver. A François, pour s'être épris de mon sujet d'étude et avoir exploré ensemble les beautés de l'infiniment petit.

A la fine équipe du bureau 208, votre aide et votre motivation.

A la RATP, aux parisiens et parisiennes, pour partager avec moi 3 heures quotidiennes de votre joie de vivre.

A l'Agence Nationale de la Recherche, pour m'avoir financé pendant 4 ans.

A tous les membres du jury, pour avoir accepté d'évaluer mon travail de thèse.

A tous ceux que j'oublie, mais que je n'oublie pas.

A mes parents, mon frère et mes sœurs.

A Thierry et Isabelle, Vittoria et Nina.

A Alex, pour une dernière relecture salvatrice.

A Laura, pour ces dernières semaines éprouvantes, ces derniers mois pleins d'émotions et ces belles années passées ensemble...merci pour tes encouragements et ton amour. Merci surtout pour les belles années qui viennent, j'ai une folle envie de danser avec toi sur cette planète...

SOMMAIRE

<u>I. INTRODUCTION</u>	1
<u>A. La systématique des eucaryotes unicellulaires basée sur la morphologie</u>	1
1. Premières observations morphologiques	1
2. Microscope électronique et classification taxonomique des eucaryotes.....	4
3. L'apport de la phylogénie moléculaire à la classification des protistes.....	5
4. Le concept de morpho-espèce chez les protistes	8
<u>B. Les approches moléculaires pour étudier la diversité des eucaryotes</u>	10
1. L'amplification de marqueurs moléculaires à partir d'ADN environnemental	11
a. Marqueurs moléculaires employés	11
b. Comment décrire la diversité des eucaryotes non cultivables ?.....	11
c. L'apport des nouvelles technologies de séquençage.....	14
2. L'amplification de marqueurs moléculaires à partir de cellules uniques.....	16
a. Couplage des études de diversité moléculaires et morphologiques	17
b. Études de diversité intra-spécifique	18
<u>C. Les ciliés tintinnidés (Ciliophora, Tintinnida)</u>	19
1. Une longue histoire taxonomique	20
2. Description morphologique.....	22
3. Des dynamiques temporelles étudiées depuis longtemps	26
4. Les approches moléculaires	26
<u>II. OBJECTIFS</u>	29
<u>III. MATERIEL ET METHODES</u>	31
<u>A. La rade de Villefranche-sur-Mer : un site d'étude marin côtier</u>	31
1. Paramètres environnementaux	31
2. Stratégie d'échantillonnage.....	32
a. Trait de filet	33
b. Prélèvement d'eau à la bouteille Niskin	34
<u>B. Méthodes d'analyse moléculaire</u>	36
1. Choix des marqueurs.....	36

2. À partir de cellules isolées, amplification par PCR directe des gènes d'ARNr 18S, 5.8S, 28S et des ITS	37
3. Extraction de l'ADN environnemental	39
4. Clonage et séquençage des gènes d'ARNr et régions ITS	40
5. Pyroséquençage des amplicons d'ADNr 18S	40
6. DGGE (Denaturing Gel Gradient Electrophoresis)	41
<u>C. Analyses <i>in silico</i> des séquences</u>	42
1. Contrôle de qualité des séquences Sanger et regroupement en UTOs.....	42
2. Reconstruction d'arbres phylogénétiques.....	43
3. Filtrage et débruitage des pyroséquences.....	44
4. Analyses statistiques des données génétiques intra-spécifiques.....	45
<u>IV RESULTATS</u>	47
<u>A. Phylogénie moléculaire des ciliés tintinnidés marins</u>	47
1. Contexte et objectifs	47
2. Manuscrit de l'article 1 : Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora).....	48
3. Synthèse et éléments de discussion.....	81
<u>B. Analyse comparative des études moléculaires environnementales de diversité et des études morphologiques classiques chez les ciliés tintinnidés</u>	85
1. Contexte et objectifs	85
2. Manuscrit de l'article 2 : Accuracy of protist diversity assessments: morphology compared to cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study.....	86
3. Synthèse et éléments de discussion.....	120
<u>C. Diversité spatio-temporelle des ciliés tintinnidés dans la rade de Villefranche-sur-Mer</u>	123
1. Contexte et Objectifs.....	123
2. Manuscrit de l'article 3 : Seasonal dynamics of free-living tintinnid ciliate communities revealed by environmental sequences from the North-West Mediterranean Sea	124
3. Synthèse et éléments de discussion.....	159
<u>D. Diversité temporelle intra-spécifique chez l'espèce <i>Undella claparedei</i></u>	161
1. Contexte et Objectifs.....	161

2. Manuscrit de l'article 4 : Temporal patterns of polymorphism at 18S rRNA genes and internal transcribed spacers in populations of <i>Undella claparedei</i> (Tintinnida) from Villefranche Bay, France	162
3. Synthèse et éléments de discussion.....	193
4. Essais expérimentaux additionnels et Perspectives	194
<u>V. DISCUSSION GENERALE ET PERSPECTIVES.....</u>	197
<u>A. Intégration de l'information morphologique et les données moléculaires chez les ciliés tintinnidés</u>	197
<u>B. Dynamique spatio-temporelle de la diversité génétique à l'échelle de la communauté et à l'échelle d'une morpho-espèce</u>	201
<u>VI. CONCLUSION</u>	205
<u>BIBLIOGRAPHIE</u>	209
<u>TABLE DES ANNEXES</u>	219

LISTE DES ABREVIATIONS

ADN, DNA: Acide désoxyribonucléique

ADNr 18S: gène codant pour l'ARN de la petite sous-unité du ribosome

ADNr 28S: gène codant pour l'ARN de la grande sous-unité du ribosome

ARN, RNA: Acide ribonucléique

ARNr 28S: ARN de la grande sous-unité du ribosome

AMOVA: Analyse de Variance Moléculaire

AU: Approximately Unbiased

DGGE: Denaturing Gradient Gel Electrophoresis

GTR: General Time Reversible

ITS: Internal Transcribed Spacer

LSU rRNA: Large Sub Unit ribosomal RNA,

LSU rDNA: Large Sub Unit ribosomal RNA gene

OTU, UTO: Unité Taxonomique Opérationnelle

PC: Principal Component

PCR: Polymerase Chain Reaction

PCA: Principal Component Analysis

PCoA: Principal Correspondence Analysis

RDA: Redundancy Analysis

SOMLIT: Service d'Observation en Milieu Littoral

SSU rRNA: Small SubUnit ribosomal RNA

SSU rDNA: Small SubUnit ribosomal RNA gene

WGA: Whole Genome Amplification

I. Introduction

Les eucaryotes sont des organismes caractérisés par la présence d'un noyau entouré d'une membrane, qui contient l'essentiel du génome eucaryote, et par la présence d'organites (notamment les mitochondries et, dans le cas des eucaryotes photosynthétiques, les chloroplastes), qui dérivent des anciennes bactéries endosymbiontes et qui possèdent des génomes très réduits. Les eucaryotes unicellulaires, par opposition aux organismes multicellulaires que sont les animaux, une bonne partie des champignons et les plantes terrestres, sont généralement appelés protistes au sens large. La grande diversité des tailles, formes et fonctions des protistes sont un objet de fascination pour les chercheurs qui travaillent sur le monde microbien. Cependant, les scientifiques ont toujours eu des difficultés pour classer et définir les limites des espèces des eucaryotes unicellulaires, comme en témoigne l'histoire mouvementée de leurs découvertes, de leurs classifications et de la perception de leur diversité au cours des derniers siècles.

A. La systématique des eucaryotes unicellulaires basée sur la morphologie

1. Premières observations morphologiques

Les améliorations du microscope par Antoni van Leeuwenhoek permettent pour la première fois d'observer et de décrire les micro-organismes individuellement (1676). L'intérêt porté à la découverte de nouvelles formes de micro-organismes, appelés animalcules ou infusoires, sera perpétué au cours du 18^e siècle par quelques naturalistes, notamment le danois Otto Friedrich Müller (1776), qui accompagnait ses descriptions précises par des planches dessinées. Les études taxonomiques des organismes unicellulaires se développent ensuite de manière exponentielle au 19^e siècle, avec l'utilisation ordinaire du microscope comme instrument de laboratoire. Par exemple, dès 1845, pratiquement chaque nouvelle forme de micro-organisme observée dans les échantillons collectés en Mer du Nord par Johannes Müller est nouvelle pour la science. C'est en 1866 que le zoologiste allemand Ernst Haeckel

I. Introduction

propose l'établissement du règne des Protista pour regrouper l'ensemble des organismes dits «inférieurs», qui diffèrent morphologiquement des plantes et des animaux. Ce règne regroupe la plupart des organismes unicellulaires, y compris la majorité des bactéries qui sont rassemblées dans un groupe appelé Monera (Figure 1A). C'est à cette période que les eucaryotes unicellulaires vont commencer à être organisés en groupes sur la base de leur morphologie générale et leur mode de nutrition. En raison de la division de l'étude des micro-organismes entre les zoologistes et les botanistes, des cadres taxonomiques différents vont être créés indépendamment pour les taxons photosynthétiques (appelés algues ou phytoplancton) et les taxons hétérotrophes (appelés protozoaires). Par exemple, la première classification des formes hétérotrophes (Protozoa) dans les années 1880 n'inclut aucune forme photosynthétique (Bütschli 1880–1889). Depuis Darwin, la systématique a un objectif qui dépasse le simple classement, elle fait en sorte que la classification témoigne des affinités évolutives entre les taxons. Ainsi, en se fondant sur le degré de ressemblance morphologique, plus il y a de caractères partagés entre deux taxons, plus le lien de parenté qui les unit est étroit, et cette proximité devrait se traduire dans la classification. Elle deviendrait donc une classification naturelle, c'est à dire phylogénétique, fondée sur des liens de parenté. Cependant, la morphologie comparée s'avère difficile pour les eucaryotes unicellulaires jusqu'au milieu du 20^e siècle à cause des limitations des méthodes d'observation microscopique. L'intérêt se porte alors principalement sur les groupes d'eucaryotes qui possèdent une enveloppe externe facilement identifiable au microscope. Ainsi, les nombreuses campagnes océanographiques de cette période permettent de décrire un grand nombre d'espèces chez les diatomées (entourées d'une coque siliceuse, la frustule), les radiolaires (squelette siliceux), les foraminifères (présence d'un test), les silicoflagellés, les amibes à thèque ou encore les ciliés tintinnidés (présence d'une coquille externe, la lorica). La distinction fondamentale entre les organismes à cellules avec noyau, les eucaryotes, et les organismes à cellules sans noyau, les procaryotes, est vaguement évoquée pour la première fois par Edouard Chatton en 1937, qui introduit les termes procaryote et eucaryote, mais ce n'est que deux décennies plus tard qu'ils atteignent leur définition moderne d'après Roger Stanier et Cornelis B. van Niel (1962). Cependant, dès la première partie du 20^e siècle, Herbert F. Copeland (1938) propose que le groupe des Monera soit élevé au rang de royaume, distinct du royaume des protistes regroupant les organismes unicellulaires avec un noyau, du royaume des plantes et de celui des animaux.

I. Introduction

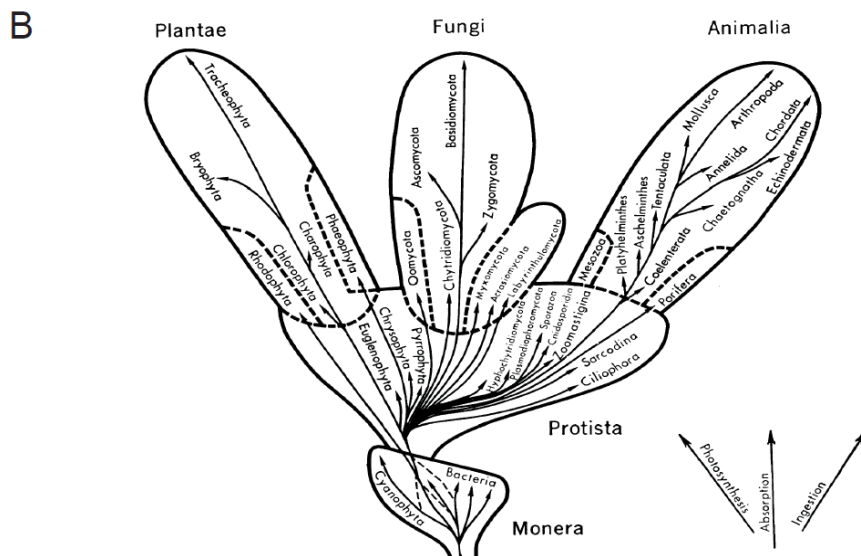
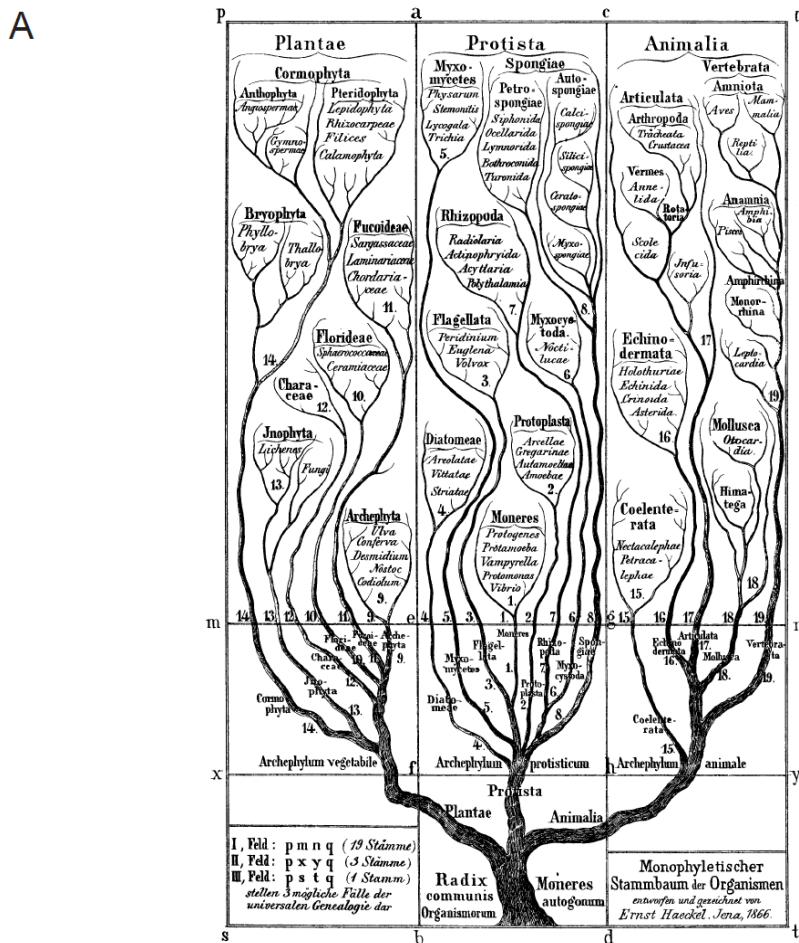


Figure 1. (A) Arbre de la vie selon Haeckel (1866). Premier arbre représenté avec un tronc commun à l'ensemble des organismes vivants. La colonne du centre comprend tous les membres du groupe des Protista *sensu* Haeckel (procaryotes + protistes). **(B)** Arbre de la vie selon Whittaker (1969) divisé en cinq royaumes. Les protistes *sensu* Whittaker sont une étape évolutive de transition entre les Monera (procaryotes) et les plantes, champignons et animaux.

2. Microscope électronique et classification taxonomique des eucaryotes

Jusque dans les années 1950/1960, les descriptions morphologiques réalisées au microscope optique ne vont pas permettre de reconstructions phylogénétiques pour les eucaryotes unicellulaires sur suffisamment de caractères. Les descriptions morphologiques vont être largement améliorées avec l'apparition du microscope électronique permettant, avec son haut pouvoir de résolution, l'identification de caractères morphologiques à l'échelle subcellulaire. Ainsi, au sein des protistes, on va tenter de reconstruire la phylogénie des espèces et des groupes d'eucaryotes sur la base de caractères ultrastructuraux partagés. John O. Corliss (1984) reconnaît ainsi 45 de ces groupes dont la monophylie *a priori* ne fait pas de doute : les ciliés, les dinoflagellés, les apicomplexes, les kinetoplastidés, les euglènes, les parabasalidés, les diplomonades, les diatomées, les chrysophytes, les chlorophytes, les ascomycètes... Bien qu'il existait un consensus sur l'existence de ces groupes sur la base d'un ensemble de caractères morphologiques partagés, les relations de parenté entre eux restaient énigmatiques. Dans les années 1970, les eucaryotes unicellulaires étaient encore considérés par les systèmes de classification dominants comme un groupe évolutif de transition entre les Monera (procaryotes) d'une part, et les plantes, animaux et champignons d'autre part (Whittaker 1969) (Figure 1B). Dans les années 1980, Thomas Cavalier-Smith met en avant un autre système de classification, basé sur l'interprétation de nombreux critères d'ultrastructure cellulaire, qui divise les eucaryotes en six royaumes (Cavalier-Smith 1987, 1993) : les Archezoa, les Protozoa, les Chromista, les Plantae (algues vertes, rouges et glaucophytes), les Fungi et les Animalia. Le royaume des Archezoa regroupe alors toutes les espèces anaérobies et amitochondriales (Metamonadida, Parabasalia, Microsporidia et Archamoebae) supposées avoir divergé avant l'origine commune de la mitochondrie à tous les autres eucaryotes par un événement endosymbiotique. Ce groupe sera finalement réintroduit dans le royaume des Protozoa après la découverte que l'absence de mitochondries chez ces espèces est due à une perte secondaire. Les eucaryotes sont alors divisés en cinq royaumes, le sixième royaume étant, sous son système de classification, celui des bactéries (Cavalier-Smith 1998, 2004).

3. L'apport de la phylogénie moléculaire à la classification des protistes

Dans les années 1960, le séquençage et la comparaison des premières séquences protéiques et d'ADN ont conduit Zuckerkandl et Pauling à proposer ces séquences comme source d'information pouvant être utilisée pour reconstruire leur histoire évolutive et donc, celle des organismes qui les portent (Zuckerkandl et Pauling 1965). C'est le début de l'ère moléculaire pour les classifications phylogénétiques. Carl R. Woese et George E Fox ont été les premiers en 1977 à utiliser une molécule choisie en raison de son abondance dans les cellules, et surtout de sa présence dans l'ensemble des organismes cellulaires : l'ARN de la petite sous-unité du ribosome (ARNr SSU), ARNr 18S chez les eucaryotes. Ceci a eu des conséquences notables pour l'étude de la diversité et la phylogénie du monde vivant, avec la découverte de la division des procaryotes en deux grandes lignées évolutives : les Bacteria et les Archaea, aussi éloignées entre elles qu'avec les Eucarya (eucaryotes) (Woese et al. 1990). Notre compréhension des relations entre groupes d'eucaryotes a été également bouleversée par l'utilisation des données moléculaires pour reconstruire les phylogénies de ces organismes (Sogin 1989; Sogin et al. 1986). Les premières phylogénies basées sur le gène codant pour l'ARNr 18S, puis, dans certains cas, sur un ensemble plus large de gènes conservés dont les séquences ont été déterminées à partir d'un nombre de protistes restreint mais bien distribué dans les taxons connus à l'époque, ont permis d'établir un cadre phylogénétique plus fiable pour la classification et ont mené à une révision de la structure générale de l'arbre des eucaryotes.

Notre vision actuelle de la phylogénie des eucaryotes comprend une diversité relativement importante de phylums bien soutenus par les analyses de phylogénie moléculaire (groupes monophylétiques). Chacun comprend une grande diversité de structures, de modes de nutrition et de comportements, qui peuvent se regrouper dans un petit nombre de grands « super-groupes » ou royaumes dont les relations phylogénétiques sont bien moins solides (Adl et al. 2005; Keeling et al. 2005; Simpson et Roger 2002) (Figure 2). Les trois grands groupes d'organismes multicellulaires (plantes, animaux et champignons) se retrouvent dispersés dans l'arbre des eucaryotes avec le reste des eucaryotes unicellulaires (Figure 2). Dans ce cadre phylogénétique, de nombreux groupes d'eucaryotes considérés comme monophylétiques sur des critères morphologiques ultrastructuraux (*i.e.* modes d'organisation cellulaire et subcellulaire révélés par microscopie électronique, tels que la forme des crêtes

I. Introduction

mitochondriales, la longueur et l'orientation des corps basaux et leur association avec d'autres organites, le nombre de membranes délimitant le chloroplaste s'il y en a un, etc.) (Patterson 1999) ont été confirmés. Ainsi, la monophylie des ciliés (présence de cils et de deux types de noyaux) a été confirmée par les données moléculaires. Cependant, d'autres groupements morphologiques se sont avérés rassembler des taxons considérés maintenant comme éloignés (e.g., les héliozoaires et les radiolaires étaient autrefois groupés ensemble), ou à l'inverse, des taxons considérés comme éloignés sont maintenant considérés comme ayant partagé une histoire évolutive récente commune (e.g., les ciliés, les dinoflagellés et les apicomplexes qui sont maintenant groupés au sein des alvéolés) (Adl et al. 2005; Baldauf 2008; Walker et al. 2011). Ce qui paraît certain aujourd'hui, c'est que l'on est loin d'avoir une phylogénie résolue pour de nombreux eucaryotes. Les données phylogénomiques, basées sur l'analyse phylogénétique d'un très grand nombre de gènes, discutent l'existence de certains groupes proposés sur la base de caractères morphologiques, comme les chromalvéolés (Burki et al. 2012; Burki et al. 2008). Toutefois, les séquences de génomes complets disponibles sont encore peu nombreuses et inégalement réparties au sein des eucaryotes. Il faudra réaliser des analyses plus complètes qui incorporent des représentants d'un plus grand nombre de taxons pour essayer de résoudre les relations évolutives parmi les lignées eucaryotes.

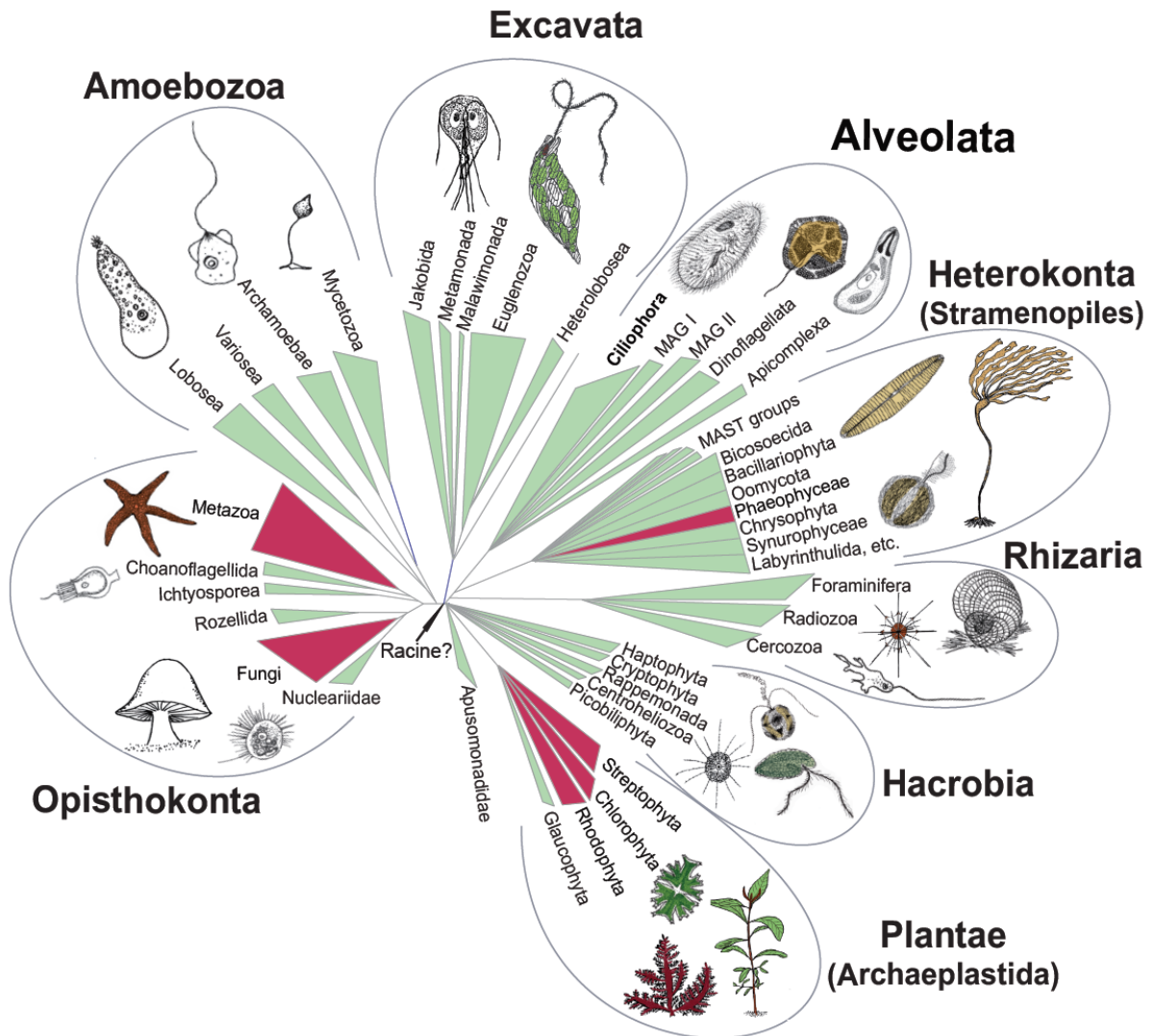


Figure 2. Arbre phylogénétique schématisé présentant la diversité du domaine des eucaryotes. La reconstruction phylogénétique est basée sur différents marqueurs conservés, en particulier l'ADNr 18S. Les groupes en vert ne comptent que des représentants unicellulaires tandis que les groupes où l'on trouve des organismes pluricellulaires avec tissus ou pseudo-tissus (les colonies sont exclues) sont colorés en rouge. Dans cet arbre simplifié, qui ne représente pas la totalité des eucaryotes connus, une partie des différentes grandes lignées sont associées avec une représentation d'une espèce représentative, illustrant la diversité morphologique au sein des eucaryotes, notamment chez les protistes. Figure modifiée d'après (López-García et Moreira 2008) et (Walker et al 2011).

I. Introduction

Dans son usage actuel, le terme protiste se réfère au sens large aux organismes eucaryotes généralement unicellulaires, mais pas toujours (par exemple les algues laminaires sont des protistes). C'est un terme qui décrit une certaine réalité phylogénétique par exclusion, puisqu'on considère comme protistes l'ensemble des lignées eucaryotes à l'exception des animaux et des plantes et, pour certains auteurs, aussi des champignons (Adl et al. 2005). Donc le terme vernaculaire de protiste recouvre un groupe polyphylétique trop hétérogène pour être taxonomiquement sensé, mais reste néanmoins très utile pour désigner les eucaryotes unicellulaires. Bien que la classification des protistes s'est révélée, et se révèle, compliquée depuis leur découverte, les scientifiques n'ont jamais cessé de les étudier et de les observer pour décrire leur diversité.

4. Le concept de morpho-espèce chez les protistes

Comprendre la notion d'espèce est central quand on parle de diversité ; cependant cette question est très souvent discutée et difficile à appréhender. La définition la plus communément admise, du moins pour les eucaryotes, est celle d'Ernst Mayr (Mayr 1948). Basée sur le concept dit « biologique », elle énonce qu'une espèce est une population dont les individus peuvent se reproduire entre eux et engendrer une descendance viable et féconde en conditions naturelles. Le concept de morpho-espèce se base sur la description et l'identification d'une espèce sur la base de critères morphologiques ; ainsi, les espèces sont des groupes d'individus morphologiquement similaires et clairement distinguables des individus d'autres groupes. Traditionnellement, c'est ce concept qui a été favorisé par les taxonomistes chez les protistes pour des raisons pratiques (e.g. Alverson 2008), puisque la morphologie reste relativement informative, surtout au niveau ultrastructural, comparée à celle des procaryotes. Cette approche morphologique des espèces a été et est encore utilisée pour approcher le concept d'espèce biologique chez les protistes, quand bien même le cycle de vie des espèces étudiées n'est pas connu. Ce n'est que plus récemment que l'on a introduit le concept phylogénétique de l'espèce : les espèces sont définies comme un groupe d'individus partageant l'histoire évolutive commune la plus récente (Baum et Donoghue 1995). Avec l'étude des phylogénies moléculaires, ce concept se traduit chez les protistes en termes de pourcentage de similarité de séquences des marqueurs moléculaires.

I. Introduction

La délimitation morphologique des espèces entre elles est en partie arbitraire et repose largement sur l'expertise et l'appréciation générale du taxonomiste. Les taxonomistes sont d'ailleurs souvent qualifiés par leurs collègues de « lumpers » (rassembleurs) ou de « splitters » (diviseurs) selon leur tendance à maintenir un nombre réduit d'espèces en intégrant différentes variabilités morphologiques au sein d'une espèce, ou au contraire, à créer de nouvelles espèces qui correspondent à des variantes morphologiques qu'ils jugent discriminantes. Cependant, dans la pratique quotidienne du chercheur, l'identification précise à l'aide du microscope est problématique. Les diatomées ou d'autres groupes présentant des spécimens de grande taille et qui possèdent des exosquelettes observables peuvent être aisément identifiés au rang de l'espèce, mais comme au temps des premières observations, de nombreux protistes ne peuvent pas l'être. Chez les ciliés par exemple, des méthodes qui permettent d'accéder à des détails morphologiques ultrastructuraux (par exemple l'imprégnation au nitrate d'argent ou l'observation par microscopie électronique) sont nécessaires pour une identification précise au niveau de l'espèce (Montagnes et Lynn 1987). Mais ces méthodes ne peuvent être appliquées de façon routinière dans les études écologiques car elles demandent des moyens trop conséquents. Enfin, un autre problème pour une identification précise des protistes est que les espèces sont souvent décrites de façon inadéquate dans la littérature ancienne. Pour une espèce donnée, la seule description est souvent une gravure accompagnée d'un petit paragraphe descriptif, ceci étant parfois la seule information disponible pour les naturalistes contemporains pour identifier un spécimen collecté.

Les différents concepts d'espèces sont en partie à l'origine du débat qui existe sur la diversité et la distribution géographique des microorganismes. D'après une certaine école de pensée (Finlay 2002), les protistes en dessous d'une certaine taille (~1 mm) devraient se comporter comme des bactéries et se soumettre au modèle de Baas-Becking ("tout est partout, mais l'environnement sélectionne") (1934), impliquant une dispersion cosmopolite sur laquelle l'effet de la sélection environnementale locale fera le tri des espèces qui se développent dans un écosystème donné. En effet, du fait de leur petite taille et de la grande taille de leurs populations, les protistes auraient une grande capacité de dispersion qui ne serait pas limitée par les barrières géographiques. En conséquence de quoi, les phénomènes de spéciation par isolement géographique ne devraient pas avoir lieu, résultant dans un petit nombre d'espèces

I. Introduction

(comparés aux animaux et aux plantes) ayant une répartition mondiale (Finlay 2002). Ce discours s'est largement appuyé sur l'étude de la morphologie des protistes, en particulier les ciliés, pour lesquels une même morpho-espèce peut être retrouvée de façon ubiquitaire (Finlay 2002; Finlay et Fenchel 2004). Cette école de pensée s'est confrontée à celle de Foissner qui a promu un « modèle d'endémicité modéré » (1999, 2006), qui propose que si certaines morpho-espèces ont bien une distribution ubiquitaire, d'autres ont une distribution géographique restreinte. Les avancées des études moléculaires ont ensuite stimulé la recherche sur ces questions de la diversité microbienne (Bass et al. 2007; Fenchel 2005). En effet, les études moléculaires ont montré que certains organismes, identiques d'un point de vue morphologique, présentent des divergences phylogénétiques et/ou écologiques importantes (ce que l'on nomme désormais la diversité cryptique) parfois associées à des régions géographiques différentes (e.g. Boenigk et al. 2005; de Vargas et al. 1999; Katz et al. 2005; Lowe et al. 2005; Nanney et al. 1998; Rodriguez et al. 2005; Slapeta et al. 2006). C'est pourquoi les morpho-espèces traditionnelles doivent être traitées avec beaucoup de précaution au sein des protistes, spécialement celles qui présentent un nombre limité de spécificités morphologiques.

B. Les approches moléculaires pour étudier la diversité des eucaryotes

Les travaux de phylogénie moléculaire des protistes ont permis de réorganiser les relations de parenté entre et au sein des groupes taxonomiques déjà connus. Ces travaux ont également confirmé l'origine commune de tous les eucaryotes en étudiant le gène codant pour l'ARN de la petite sous-unité du ribosome (ADNr 18S pour les eucaryotes) (Woese et Fox 1977). Toutefois, la résolution des relations de parenté entre eucaryotes n'est pas la seule réussite de la biologie moléculaire. L'analyse de la diversité dans le milieu naturel par amplification de marqueurs moléculaires à partir d'ADN environnemental a eu un profond impact sur la perception de la diversité des protistes, en particulier en milieu marin (Massana 2011; Moreira et López-García 2002).

1. L'amplification de marqueurs moléculaires à partir d'ADN environnemental

a. Marqueurs moléculaires employés

Tous les protistes partagent avec les autres organismes cellulaires le système de l'information héréditaire (l'ADN) et une partie de la machinerie complexe d'expression de cette information, notamment la machinerie de traduction : le ribosome. Celui-ci est un arrangement complexe de protéines et d'ARN, constitué d'une grande et d'une petite sous-unité, qui permet de traduire les ARN messagers en protéines par le biais du code génétique universel. La grande sous-unité (LSU, *Large SubUnit*) contient entre autres l'ARNr 28S et 5.8S. La petite sous-unité (SSU, *Small SubUnit*) contient l'ARNr 18S. Ces différents ARNr sont codés en une seule unité de transcription appelée le cistron ribosomique (Hillis et Dixon 1991). Historiquement, les gènes codant pour l'ARN des sous-unités du ribosome se sont donc imposés comme des marqueurs phylogénétiques de premier choix (Long et Dawid 1980; Woese et Fox 1977). L'ARNr 18S présente le meilleur compromis en termes de taille, d'environ 1900 bp (même si certains protistes possèdent des insertions qui peuvent augmenter considérablement la longueur de ce marqueur), assurant un grand nombre de positions informatives comparables entre les différentes lignées, et de variabilité de séquence, permettant de discriminer les organismes à différents niveaux taxonomiques. L'un des avantages majeurs de l'ADNr 18S est qu'il est possible de comparer un très grand nombre de séquences entre elles grâce à la disponibilité d'un nombre croissant et très élevé des séquences de ce gène, entières ou partielles, dans les bases de données publiques. Cependant, pour distinguer et observer de la variabilité entre organismes phylogénétiquement très proches ou faisant partie de la même espèce, des marqueurs plus variables que l'ARNr 18S sont utilisés. Par exemple, les deux régions intergéniques du cistron ribosomique (ITS, *Internal Transcribed Spacer*) sont des régions non codantes et donc a priori soumises à une pression de sélection plus faible que les régions bordantes, présentant en général un haut taux de polymorphisme.

b. Comment décrire la diversité des eucaryotes non cultivables ?

Il est admis qu'une majorité de la diversité microbienne n'est pas ou est difficilement cultivable ; dans le cas des procaryotes, on estime que moins de 1% des espèces seraient

I. Introduction

cultivables en fonction des environnements (Rappe et al. 2003). Afin d'accéder à cette diversité dans le milieu naturel, les recherches se sont basées sur l'amplification directe de marqueurs moléculaires à partir d'ADN environnemental. De façon expérimentale, l'ADNr 18S est amplifié par réaction de polymérisation en chaîne (PCR, *Polymerase Chain Reaction*) à partir de l'ADN de toute la communauté microbienne, à l'aide d'amorces qui ciblent les gènes eucaryotes. Les produits de PCR sont ensuite séparés et séquencés par plusieurs méthodologies. La méthode classiquement utilisée consiste à cloner et séquencer par une stratégie Sanger le gène d'ARNr 18S de chaque clone. Les banques ainsi construites, riches de plusieurs dizaines ou centaines de clones, sont supposées être représentatives de la diversité spécifique du gène amplifié à partir de l'ADN environnemental, et donc par extension, de la communauté présente dans l'échantillon environnemental.

Pour estimer la richesse spécifique, les séquences de clones, c'est à dire les ADNr 18S, sont alignées à l'aide des logiciels d'alignement multiple, et regroupées par similarité (pourcentage d'identité) de séquences. Les groupes de séquences partageant plus de 97-99% d'identité (le niveau d'identité est souvent défini par le chercheur en fonction du taxon étudié; >97% est en général utilisé chez les procaryotes; >98 ou 99% chez les eucaryotes) sont assimilés à des Unités Taxonomiques Opérationnelles (UTO) et sont considérés comme une approximation phylogénétique des espèces. Les études environnementales ont montré l'existence de plusieurs milliers d'UTOs cohabitant dans quelques litres d'eau de mer (Edgcomb et al. 2011). Chez les eucaryotes, il existe peu d'études qui ont justifié le pourcentage de similarité à employer pour construire les UTOs afin de décrire au mieux la richesse spécifique. Par exemple, à partir des séquences d'ADNr 18S disponibles dans GenBank, Caron et collaborateurs défendent l'emploi de 98% de similarité pour les eucaryotes en général (Caron et al. 2009), ainsi que Nebel et collaborateurs (2010) qui défendent également l'emploi de 98% de similarité pour définir une UTO chez les ciliés. Cependant, ces études concèdent qu'il n'existe pas de pourcentage unique qui pourrait objectivement décrire la diversité des protistes, puisque la variabilité de l'ADNr 18S est très hétérogène en fonction des groupes taxonomiques considérés.

Cette approche a d'abord été appliquée avec succès chez les procaryotes en révélant de nombreuses lignées pour lesquelles il n'existait pas d'informations génétiques et pour

I. Introduction

lesquelles les informations morphologiques n'étaient pas discriminantes (Giovannoni et al. 1990). Ce n'est qu'environ une dizaine d'années plus tard que cette stratégie de caractérisation de la diversité microbienne dans les écosystèmes naturels a été appliquée à l'étude de la diversité de microorganismes eucaryotes. Comme chez les procaryotes, cette approche a permis de mettre en évidence une exceptionnelle diversité chez les eucaryotes. Les études effectuées à partir d'échantillons marins ont montré la présence de séquences affiliées à des taxons déjà connus, parfois en abondance comme pour la classe des prasinophytes (Diez et al. 2001; Romari et Vaultot 2004). De façon plus inattendue, les reconstructions phylogénétiques ont mis en évidence la présence, parfois abondante, de séquences très éloignées des groupes taxonomiques connus. Ainsi, de nouvelles lignées ont été révélées parmi les alvéolés avec la découverte de deux groupes de séquences (les Marine Alveolates Group I et II, MAGI et MAGII) proches des dinoflagellés, mais sans séquences de référence disponibles pour les identifier sans ambiguïté (López-García et al. 2001; Moon-van der Staay et al. 2001). De la même manière, des groupes de séquences environnementales appartenant à des lignées phylogénétiques nouvelles ont été mis en évidence parmi les stramenopiles (MARine STRamenopiles, MAST, Massana et al. 2004; Massana et al. 2002). Depuis ces premières découvertes, de nouvelles lignées non caractérisées par des méthodes traditionnelles ont été mises en évidence avec par exemple les rappemonadés (Kim et al. 2011b) ou les picobiliphytes (Not et al. 2007) qui branchent parmi le phylum controversé des Hacrobia. Ces découvertes ont profondément modifié notre vision et notre approche expérimentale de la diversité, l'écologie et l'évolution des protistes de la biosphère.

Toutefois, l'approche moléculaire n'est pas exempte de biais. Il a été montré en particulier que certains alvéolés, qui possèdent un grand nombre de copies du cistron ribosomique, ont tendance à dominer les banques de clones. Ces analyses moléculaires de la diversité fondées sur l'amplification, le clonage et le séquençage des gènes d'ARNr 18S permettent facilement de décrire les organismes majoritaires dans un échantillon. Cependant, en partie à cause du nombre limité de séquences produites (de l'ordre de la dizaine ou de la centaine) et à la surdominance de certaines lignées (réelle ou due au nombre élevé de copies des gènes qui codent l'ARNr 18S), elles ne permettent pas de décrire la totalité de la diversité présente dans un écosystème, et notamment, celle des espèces rares ou minoritaires. Pour pallier à ce problème, il est possible de choisir des amorces qui ciblent plus spécifiquement

I. Introduction

certaines lignées. De cette manière, il est possible d'amplifier des gènes de certaines espèces ou groupes taxonomiques spécifiques. L'approche groupe-spécifique a montré une diversité phylogénétique insoupçonnée au sein de groupes connus sur des critères morphologiques, comme les cercozoaires (Bass et Cavalier-Smith 2004), les haptophytes (Liu et al. 2009), les diploméidés (Lara et al. 2009) ou les télonémides (Brate et al. 2010). Cependant, ces études étant rarement couplées avec des approches morphologiques, il n'est pas possible de savoir si ces nouvelles séquences ne reflètent pas une diversité morphologique déjà connue pour laquelle des données génétiques de référence ne sont pas encore disponibles.

c. L'apport des nouvelles technologies de séquençage

La stratégie d'amplification, clonage et séquençage des gènes d'ARNr 18S s'est révélée très utile pour réaliser des analyses de diversité dans des écosystèmes variés. Cependant, il existe des biais associés à cette stratégie entraînant parfois une sur- ou sous-représentation de certains taxons dans les banques de clones. Parmi ces biais, on peut noter l'extraction différentielle d'ADN des différents eucaryotes présents dans l'échantillon à cause de lyse différentielle, ou le nombre de copies des gènes qui peut être variable entre différentes lignées mais aussi au sein d'une même espèce (Potvin et Lovejoy 2009).

Depuis quelques années, une recherche méthodologique très active a abouti à la mise au point de nouvelles technologies de séquençage, qui sont à la fois plus performantes en terme de quantité de séquences produites par échantillon et en terme de coût. Parmi ces méthodes de séquençage de nouvelle génération à haut-débit se trouvent la technologie Illumina (HiSeq 2000 qui délivre 600 Gbp en lectures de 100 bp par *run*) ou la technologie 454 de Roche, le pyroséquençage (un *run* FLX+ produit actuellement 1 million de séquences d'environ 700 bp). Un avantage additionnel est que ces méthodes permettent de séquencer les fragments de gènes amplifiés sans passer par l'étape de clonage. Aussi, pour un même échantillon, il est possible d'obtenir plusieurs milliers à centaines de milliers de séquences.

Ces méthodes présentent malheureusement des désavantages : la petite taille des séquences générées par Illumina, qui ne pose pas problème dans le cas du séquençage de génome du fait de la profondeur de lecture (plusieurs lectures pour une même région), ne permet pas d'attribuer correctement à un taxon les séquences du gène amplifié, du moins rarement au niveau spécifique. La technique d'Illumina a donc été moins employée pour des

I. Introduction

études de diversité, même si elle commence à être utilisée pour décrire la diversité de certaines lignées eucaryotes, mettant en évidence une diversité insoupçonnée (foraminifères, Lecroq et al. 2011). Une limitation majeure de la technologie 454 est liée aux homopolymères, ou répétitions consécutives d'une même base (GGG ou AAA par exemple), qui peuvent entraîner des erreurs de séquençage par délétion ou insertion (Shendure et Ji 2008), entraînant un taux d'erreur mesuré à 0,5% dans les analyses de diversité (Huse et al. 2007).

Les analyses de diversité eucaryote en utilisant le séquençage d'amplicons (fragments de gènes amplifiés par PCR) par pyroséquençage 454 se sont multipliées pendant ces dernières années. Une observation commune dans les études de diversité microbienne utilisant ces nouvelles technologies confirme une observation déjà faite à partir des études de clonage et de séquençage, mais à plus petite échelle : de nombreuses communautés sont composées d'un petit nombre de taxons qui dominent par leur abondance et d'une majorité de lignées détectées en faible abondance. Les espèces représentées dans ces études par une ou deux séquences sont souvent considérées comme faisant partie d'une « biosphère microbienne rare » (Pedrós-Alió 2007; Sogin et al. 2006). Cette biodiversité rare, d'abord observée chez les procaryotes, est maintenant bien documentée chez les protistes (Caron et Countway 2009; Stoeck et al. 2009; Stoeck et Epstein 2009). Cette idée de rareté dans l'observation de la biodiversité n'est pas nouvelle et a été longuement étudiée en biologie de la conservation chez les macro-organismes (Gaston 1994). Toutefois, si une biodiversité "rare" formée par des espèces peu abondantes qui sont soit inactives (par exemple, en train de disperser ou en dormance dans l'attente des conditions favorables), soit actives (des espèces non abondantes peuvent réaliser des fonctions importantes du point de vue écologique) existe bel et bien, on peut se demander si une partie de cette "biosphère rare" n'est pas le produit des erreurs de séquençage liés à la méthode 454 ou à d'autres biais techniques, comme cela a été soulevé (Kunin et al. 2010; Quince et al. 2009; Reeder et Knight 2009).

L'amplification de marqueurs moléculaires à partir d'ADN environnemental pour caractériser la diversité a pour objectif de répondre à la question « combien d'espèces différentes et quelles espèces se trouvent dans un échantillon naturel ? » et ce, sans avoir d'information au préalable sur les individus présents dans l'échantillon. Cette approche a

généralisé une quantité considérable de données génétiques et a conduit à un consensus sur le fait que la diversité en espèces et phylogénétique des protistes est bien plus grande qu'on ne le pensait. Pour autant, l'efficacité de ces méthodes moléculaires pour l'identification des protistes dépend intégralement du nombre de séquences dans la base de données de référence liées sans équivoque à des espèces décrites sur des critères taxonomiques classiques.

2. L'amplification de marqueurs moléculaires à partir de cellules uniques

La difficulté de maintenir en culture de nombreux protistes implique que de nombreuses descriptions d'espèces ont été basées sur la seule observation de spécimens collectés sur le terrain. Alors qu'il est possible de faire des observations morphologiques détaillées sur ces populations naturelles, il est difficile de les mettre en relation avec l'information moléculaire obtenue sur ces mêmes populations à partir d'études de diversité moléculaire environnementale. En premier lieu, il apparaît nécessaire d'attribuer les séquences environnementales aux morpho-espèces connues et décrites qui n'ont pas été caractérisées génétiquement. Ensuite, en relation avec le problème posé par la diversité cryptique, il est devenu indispensable de savoir si derrière une morpho-espèce donnée se cache une diversité génétique plus grande que celle attendue au sein d'une espèce phylogénétique, et quelle est son ampleur. Les études de génétique des populations peuvent fournir des informations importantes sur la diversité génétique au sein et entre des populations. Cependant, si l'on veut interpréter une telle information, il est essentiel de distinguer la diversité intra-spécifique qui existe au sein d'une population et la diversité cryptique au sein des espèces (diverses espèces phylogénétiques pour une seule morpho-espèce). Il est donc important d'élucider les limites des espèces avant d'évaluer les variations au niveau populationnel. Pour répondre à ces deux points, les recherches s'intensifient pour obtenir de l'information génétique par amplification et séquençage du gène de l'ARNr 18S et d'autres marqueurs moins conservés à partir de cellules individuelles.

Il existe plusieurs méthodes pour isoler et séparer les cellules de leur milieu naturel pour l'étude de la diversité des communautés d'eucaryotes. Pour les espèces de taille inférieure à 20 μm , la cytométrie en flux est une technique qui permet de faire défiler les cellules présentes dans un milieu à grande vitesse dans un capillaire devant un faisceau laser

I. Introduction

et de les trier en fonction des différentes caractéristiques (taille, fluorescence naturelle ou émise par des fluorochromes divers, ...) (Heywood et al. 2010). D'autre part, de nombreux protistes de taille supérieure à 20 μm sont différenciables sous microscope optique et dissociables de la communauté à l'aide de micropipettes ou capillaires. Après cette étape, une photo ou une vidéo peuvent être prises pour caractériser morphologiquement la cellule et fournir une archive pour l'identification. Une fois la cellule isolée de son environnement, les marqueurs moléculaires d'intérêt peuvent être amplifiés par PCR directement sur le matériel récolté. Toutefois, une pré-amplification de l'ADN par amplification complète du génome (WGA) est parfois nécessaire pour fournir suffisamment de matériel génétique (Binga et al. 2008). Cette méthodologie est théoriquement applicable à n'importe quelle cellule eucaryote.

a. Couplage des études de diversité moléculaires et morphologiques

L'amplification de l'ADNr 18S par PCR directe sur des cellules isolées manuellement et la construction de phylogénies moléculaires ont permis d'avancer dans la reconstruction de l'arbre de certaines lignées d'eucaryotes. Chez les protistes, la première référence de l'usage de techniques moléculaires sur des cellules individuelles de protistes semble être Orias et collaborateurs (1991). Ils ont étudié des cellules individuelles de l'espèce *Tetrahymena thermophila* pour détecter du polymorphisme dans les gènes d'ARNr. La majorité des études qui ont suivi ont visé le séquençage des régions codantes et non codantes de la région de l'ADNr (18S, 5.8S, 28S et ITS) pour des analyses de biodiversité et de phylogénie. En conséquence, des représentants d'une partie des principaux groupes d'eucaryotes identifiés jusqu'à aujourd'hui ont été sujets à ce type d'approche : chez les Apicomplexa (Pecher et al. 2004), les Ciliophora (Andreoli et al. 2009; Auinger et al. 2008; Van Hoek et al. 2006), les Dinoflagellata (Gómez et al. 2010b; Gómez et al. 2011; Gómez et al. 2012; Guillou et al. 2002), les Phaeodarea (Yuasa et al. 2006), les Foraminifera (Pawlowski et al. 2002), les Haptophyta (Takano et al. 2006), les Oxymonadida (Heiss et Keeling 2006), les Chrysophyceae et Synurophyceae (Auinger et al. 2008). Les phylogénies moléculaires de certaines lignées de protistes ont été depuis peu enrichies de façon significative en une seule étude avec des cellules isolées représentatives d'un grand nombre d'espèces les plus diverses possibles, en particulier chez les protistes présentant des parties morphologiques dures et décorées (Decelle et al. 2012; Krabberød et al. 2012, chez les Radiolaria; Goussard et al. 2012; Lara et al. 2008,

I. Introduction

chez les amibes à thèques Arcellinida). Les travaux d'isolement manuel ont également l'avantage de permettre l'accès à l'information génétique d'espèces très rarement observées, à l'image des dinoflagellés porteurs d'ocelloïde *Erythrospidinium* et *Warnowia* pour lesquelles l'ADNr 18S a pu être amplifié et séquencé par PCR directe pour la première fois (Gómez et al. 2009).

Rares sont les études qui ont essayé de coupler les approches d'amplification d'ADNr 18S à partir d'ADN environnemental et à partir de cellules individuelles pour lier sans ambiguïté l'information morphologique et génétique lors d'un même échantillonnage (Medinger et al. 2010). La nécessité de convergence des informations de la protistologie classique et des études environnementales est une tâche urgente. Le déficit actuel est de démêler la correspondance entre les groupes de séquences d'ADN et les espèces biologiques. Cela permettrait entre autres de coupler un corpus très vaste de connaissances classiques sur l'écologie, le cycle de vie et la biologie cellulaire des organismes avec les bases de données de séquences obtenues à partir de suivis environnementaux. Un bon exemple de ce mariage est l'utilisation récente de l'isolement de cellules individuelles de protistes pour avoir des représentants décrits de certains groupes de séquences environnementales. Il est apparu que le genre parasitaire *Euduboscquella*, dont la morphologie et le cycle de vie sont connus depuis le début du 20^e siècle (Chatton 1920), est affilié au clade environnemental des Marine Alveolate Group I, abondamment détecté dans les banques de clones d'échantillons d'eaux de mer (Bachvaroff et al. 2012; Harada et al. 2007; López-García et al. 2001). De façon analogue, l'espèce marine coloniale *Solenicola setigera*, considérée comme un épiphyte de la diatomée *Leptocylindrus mediterraneus*, est affiliée aux séquences d'un groupe des Marine STRamenopiles (MAST-3) (Gómez et al. 2010a).

b. Études de diversité intra-spécifique

La possibilité de collecter depuis l'environnement plusieurs cellules appartenant à la même morpho-espèce ouvre de nouvelles possibilités sur l'étude de la diversité intra-spécifique. Les études cherchant à révéler la variabilité génétique intra-spécifique de l'ADNr 18S ou des séquences des ITS chez les protistes ont été nombreuses, mais jusqu'à maintenant basées sur un faible nombre de représentants, souvent issus de cultures clonales (e.g. Gentekaki et Lynn 2009). Chez les ciliés, par exemple, des études ont révélé des taux de

I. Introduction

diversité génétique variables en fonction de l'espèce considérée. Certaines espèces sont apparemment sujettes à une diversité génétique faible (e.g. *Tetrahymena thermophila*, Katz et al. 2006 ; *Laboea strobila*, Katz et al. 2005), d'autres présentent à l'inverse une diversité génétique élevée (e.g. *Paramecium caudatum*, Barth et al. 2006 ; *Cyclidium glaucoma*, Finlay et al. 2006). Jusqu'à maintenant, la plupart des études de génétique des populations, basées sur l'amplification des gènes nucléaires et/ou de marqueurs microsatellites, se sont concentrées sur un nombre relativement petit d'espèces d'intérêt écologique et/ou économique comme les algues toxiques, les coccolithophoridés ou les diatomées (Casteleyn et al. 2009b; Evans et al. 2005; Iglesias-Rodríguez et al. 2006; Rynearson et Armbrust 2005). Cependant, ces études sont toutes limitées par la nécessité d'établir des lignées cellulaires par culture clonale. D'autre part, très peu d'études de diversité à l'échelle populationnelle se sont basées sur l'amplification de marqueurs ribosomiques par PCR directe à partir de spécimens isolés (e.g. Pawlowski et al. 2007). Cette dernière approche pourrait permettre de décrire la diversité intra-spécifique de nombreuses espèces sans passer par une phase difficile de culture.

C. Les ciliés tintinnidés (Ciliophora, Tintinnida)

Si nous choisissons de travailler sur un groupe taxonomique riche en espèces décrites, sera-t-il possible de concilier les phylogénies moléculaires et celles dérivées des traits morphologiques ? Est-ce que la diversité génétique révélée par les approches environnementales sera cohérente avec la diversité morphologique observée ? Que nous apprendra la diversité génétique que la diversité morphologique ne peut nous révéler ?

Le développement des approches moléculaires basées sur l'amplification et séquençage des gènes d'ARNr 18S dans l'étude des eucaryotes a révolutionné notre perception de leur diversité et de leur histoire évolutive. Cependant, de nouvelles questions sont apparues, et l'un des défis de l'étude contemporaine des protistes est de réussir à accorder les nouvelles approches moléculaires avec les masses de données taxonomiques, comportementales et écologiques accumulées pendant l'ère pré-moléculaire. Travailler à

I. Introduction

l'échelle des protistes en général a plusieurs désavantages, en particulier du fait de leur grande diversité morphologique, d'habitats, de cycle de vie et des données très parcellaires sur certains groupes. J'ai choisi de m'intéresser à un groupe présentant une cohérence phylogénétique et fonctionnelle, les ciliés tintinnidés, afin de répondre à certaines problématiques de la protistologie :

- La diversité génétique révélée par les approches environnementales est-elle corrélée avec la diversité morphologique ?
- Que nous apprend le couplage des approches moléculaires environnementales (amplification à partir d'ADN environnemental et isolement de cellules individuelles) sur la diversité et la phylogénie de groupes de protistes bien connus morphologiquement ?
- Les méthodes d'amplification de marqueurs moléculaires à partir d'ADN environnemental sont-elles efficaces pour décrire et comprendre les dynamiques temporelles des communautés de protistes ?
- La méthode d'amplification de marqueurs moléculaires sur cellule isolée est-elle efficace pour décrire la diversité au niveau intra-spécifique ? Existe-t-il une dynamique temporelle de cette diversité ?

1. Une longue histoire taxonomique

Les ciliés tintinnidés sont des organismes unicellulaires planctoniques caractérisés par la présence d'une coquille externe, appelée lorica. Il existe d'autres organismes planctoniques présentant des structures observables en guise de squelette interne ou externe tels que les foraminifères ou les radiolaires, mais peu présentent la grande diversité de formes rencontrées chez les tintinnidés. Cette morphologie particulière a attiré la curiosité de quelques uns des premiers protozoologistes. Ainsi, on retrouve pour la première fois la description d'un tintinnidé par O. F. Müller en 1776, qui mentionne la présence d'une loge dans laquelle vit un cilié. Il nomme cet organisme *Trichoda inquilinus* ; tous les ciliés étaient placés dans le genre *Trichoda* par Müller et *inquilinus*, qui signifie en latin l'occupant, se réfère au cilié comme

I. Introduction

l'habitant d'une structure. Le terme *Tintinnus* est mentionné pour la première fois par Schrank (1803) pour reclasser l'espèce *Trichoda inquilinus* dans un nouveau genre. Le terme *Tintinnus* en latin signifie clochette. La première étude détaillée sur les tintinnidés décrit morphologiquement 17 espèces présentant des loricae de différents types ; étude dans laquelle les auteurs considèrent déjà ces différentes espèces comme faisant partie d'un même groupe naturel (Claparède et Lachmann 1868). Plusieurs observations, descriptions et monographies de nouvelles espèces de tintinnidés sont réalisées depuis les eaux côtières aux larges de Messine et Lanzarote (Haeckel 1873), de Villefranche-sur-Mer (Fol 1881) et de Naples (Daday 1886; Entz 1884; Entz 1885). La première revue des espèces de tintinnidés incluant une clé de classification est publiée par Eugene von Daday en 1887. À l'instar de l'étude d'autres groupes planctoniques comme les diatomées, les dinoflagellés ou les copépodes, l'étude des tintinnidés et la description de nouvelles espèces vont largement profiter des grandes expéditions océanographiques du début du 20^e siècle qui sillonnent les mers du globe (Brandt 1906, 1907; Jörgensen 1924; Laackmann 1907, 1910; Meunier 1910). Les deux monographies de Charles Kofoid et Arthur Campbell (1929, 1939) renomment et réarrangent de nombreux genres de tintinnidés et décrivent de nouvelles formes de tintinnidés sur la base du plancton collecté durant l'expédition Agassiz en 1904-1905. La taxonomie et la phylogénie des tintinnidés sont entièrement basées sur la morphologie de la lorica, et l'identification des espèces est encore aujourd'hui basée sur ce principe. Suivront plusieurs rapports taxonomiques remarquables jusqu'à la fin du 20^e siècle (Campbell 1942; Hada 1937; Paulmier 1995, 1997), mais de manière générale, l'intérêt porté aux tintinnidés pendant la seconde moitié du 20^e siècle se portera principalement sur leur rôle écologique au sein du réseau trophique marin (Gifford et al. 1988; Heinbokel 1978; Heinbokel et Beers 1979). Il est maintenant reconnu que les tintinnidés font partie du microzooplancton, qui constitue avec les autres ciliés et les dinoflagellés hétérotrophes, le consommateur majoritaire de la production primaire du réseau trophique pélagique (Calbet et Landry 2004).

Les études réalisées à l'échelle de la planète ont montré une répartition globale des tintinnidés (Pierce et Turner 1993) (Figure 3). En revanche, certaines espèces présentent des patrons de distribution en fonction des grandes zones biogéographiques souvent rencontrés dans l'étude de la distribution du plancton : les espèces ubiquitaires que l'on retrouve partout (e.g. *Eutintinnus*, *Amphorides*), des espèces uniquement observées le long des côtes (e.g. *Metacylis*, *Tintinnopsis*), uniquement dans les eaux chaudes et tempérées (e.g. *Cyttarocylis*,

I. Introduction

Undella), ou uniquement dans les mers polaires (e.g. *Ptychocylis*, *Cymatocylis*). Des rapports basés sur les données d'occurrence de Pierce et Turner (1993) ont établi que les tintinnidés présentent un patron de diversité latitudinale communément observé chez les protistes, avec une diversité minimale aux pôles et des pics de diversité observés autour des tropiques (Dolan et al. 2006) (Figure 3).

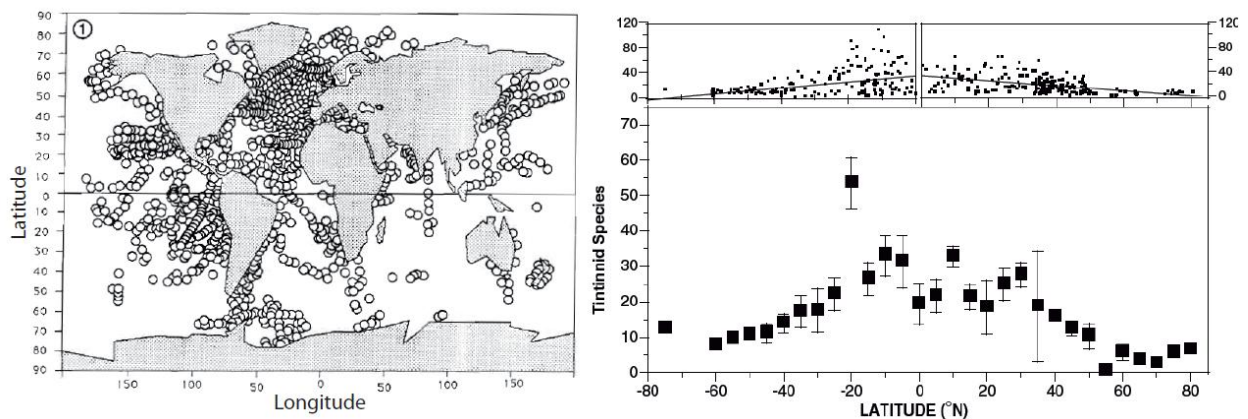


Figure 3. Répartition mondiale des 1800 sites d'échantillonnage où des espèces de tintinnidés ont été enregistrées (Pierce et Turner, 1993) (à gauche). Les tintinnidés montrent un gradient latitudinal typique de diversité spécifique. Les carrés représentent la richesse spécifique totale par bandes de 5° de latitude depuis la base de données de Pierce et Turner (1993), compilée par Dolan et collègues 2006 (à droite).

2. Description morphologique

On s'accorde à dire que le nombre total de genres décrits de tintinnidés est de 50 (Lynn 2008), avec un nombre d'espèces décrites se situant autour de 1 000 (Agatha et Struder-Kypke in press). Mais il n'existe pas actuellement de compedium listant de manière formelle toutes les formes de tintinnidés décrites jusqu'à aujourd'hui. Du point de vue de la systématique, les tintinnidés sont des ciliés (Ciliophora) de la sous-classe des Choreotrichida, constituant l'ordre des Tintinnida. Parmi les protistes, les ciliés se caractérisent par la possession de cils, de deux types de noyaux (un micronucleus diploïde impliqué dans la reproduction et un macronucleus polyplôïde impliqué dans l'expression de l'information

I. Introduction

génétique) et d'un type de recombinaison sexuelle appelée conjugaison. Au sein des ciliés, les Choreotrichida se caractérisent par une zone adorale avec des « polykinetids » servant pour la locomotion et la nutrition, organisés en rangées circulaires autour de la cavité orale (Lynn 2008). Enfin, le caractère partagé par tous les tintinnidés est la présence d'une coquille sécrétée par la cellule, la lorica. La partie postérieure de la cellule est contractile, et une ou plusieurs extensions cellulaires en forme de longues tiges connectent la cellule à sa lorica, appelées pédoncules (Figure 4). On distingue différents types de loricae chez les espèces de tintinnidés : hyaline, agglomérée avec des structures minérales ou biogéniques (frustules de diatomées ou coccolithophores), ou les deux à la fois (Figure 5). C'est sur ce critère d'agglomération que la première phylogénie des tintinnidés a été construite (Kofoid et Campbell 1939), considérant les loricae agglomérées comme un caractère ancestral. L'identification des spécimens de tintinnidés au niveau de l'espèce repose donc sur la forme, la taille et la capacité d'agglomérer ou non des particules de la lorica. Cependant, chez certaines espèces, les caractéristiques de la lorica utilisées pour établir une classification taxonomique des tintinnidés prêtent à confusion. En particulier, la lorica est depuis longtemps suspectée de présenter des variantes dans les populations naturelles de plusieurs espèces (e.g. Boltovskoy et al. 1990; Davis 1981; Santoferrara et Alder 2009; Wasik et Mikolajczyk 1994). L'exemple le plus marquant des limites de la taxonomie basée sur la lorica est l'observation faite sur des cultures qu'une seule espèce de *Favella* peut présenter des loricae que l'on pensait être spécifiques de genres différents (Laval-Peuto 1977, 1981). Malgré ces limitations, pour des raisons pratiques évidentes, l'identification des spécimens de tintinnidés au niveau du genre et de l'espèce se pratique encore aujourd'hui en fonction de la morphologie de la lorica.

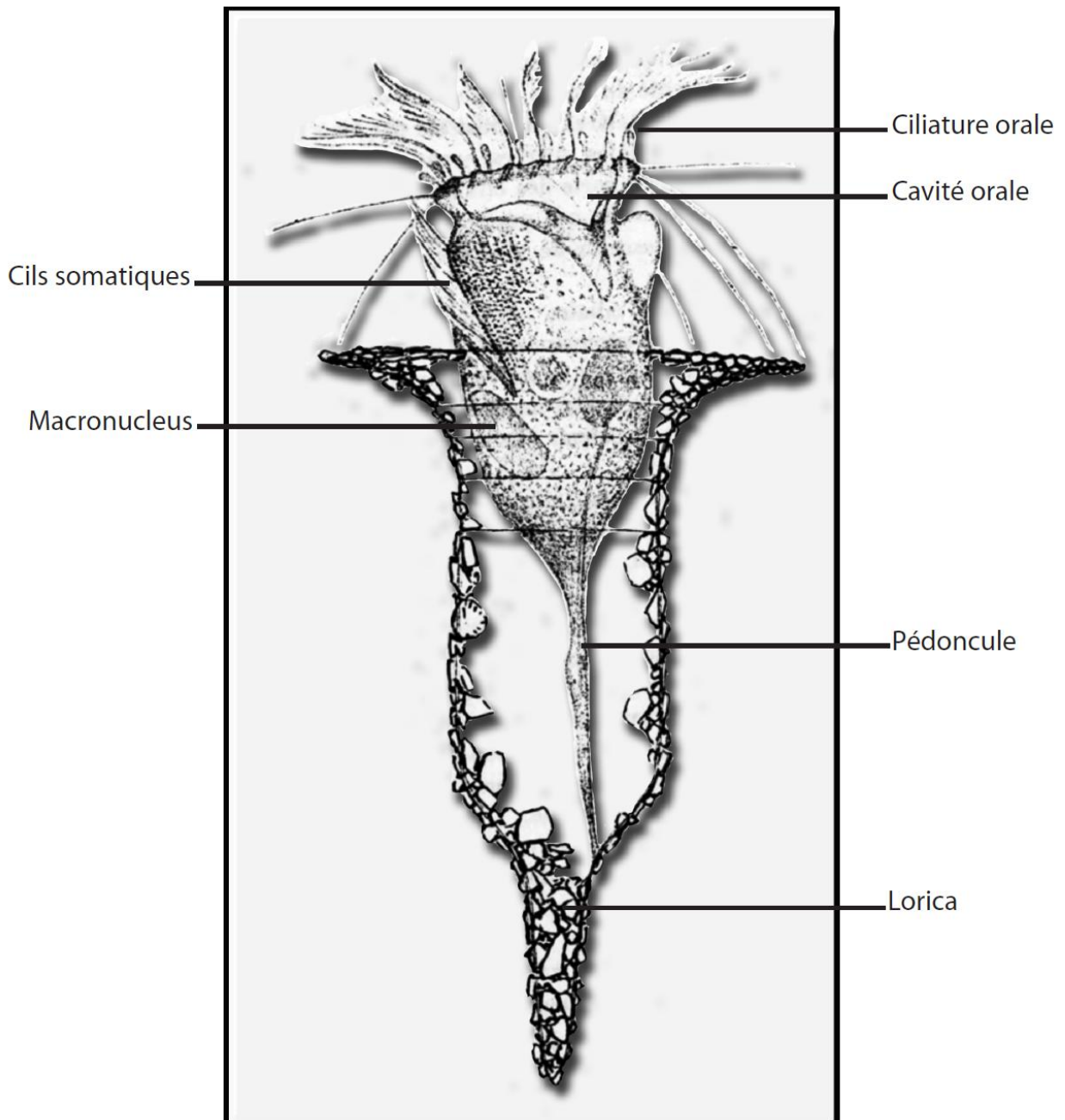


Figure 4. Morphologie de base d'un tintinnidé sur une représentation d'un *Tintinnopsis campanulla*, modifiée d'après Fauré-Fremiet (1924). La ciliature orale est composée de membranelles orales arrangées en cercle fermé autour de la cavité orale. La surface de la cellule est en partie recouverte de rangées de cils courts. Le macronucleus est souvent apparent sous le microscope optique (moins le micronucleus qui lui est associé). La partie postérieure de la cellule est attachée à la lorica par un pédoncule contractile.

I. Introduction

En revanche, il existe quelques caractéristiques morphométriques simples, comme le diamètre de l'ouverture orale, qui apparaissent être des critères taxonomiques conservatifs et donc suffisants pour distinguer des espèces entre elles (Laval-Peuto et Brownlee 1986). De plus, les dimensions de la lorica, typiquement le diamètre de l'ouverture orale, sont corrélées avec les caractéristiques de leur nourriture. Ainsi, le diamètre de l'ouverture orale de la lorica fait environ quatre fois le diamètre des proies ingérées le plus efficacement (Dolan et al. 2002). Ceci est un indicateur de leur rôle écologique puisque, en tant que brouteurs, ils peuvent ainsi contrôler les populations de certains organismes en fonction de leur taille.

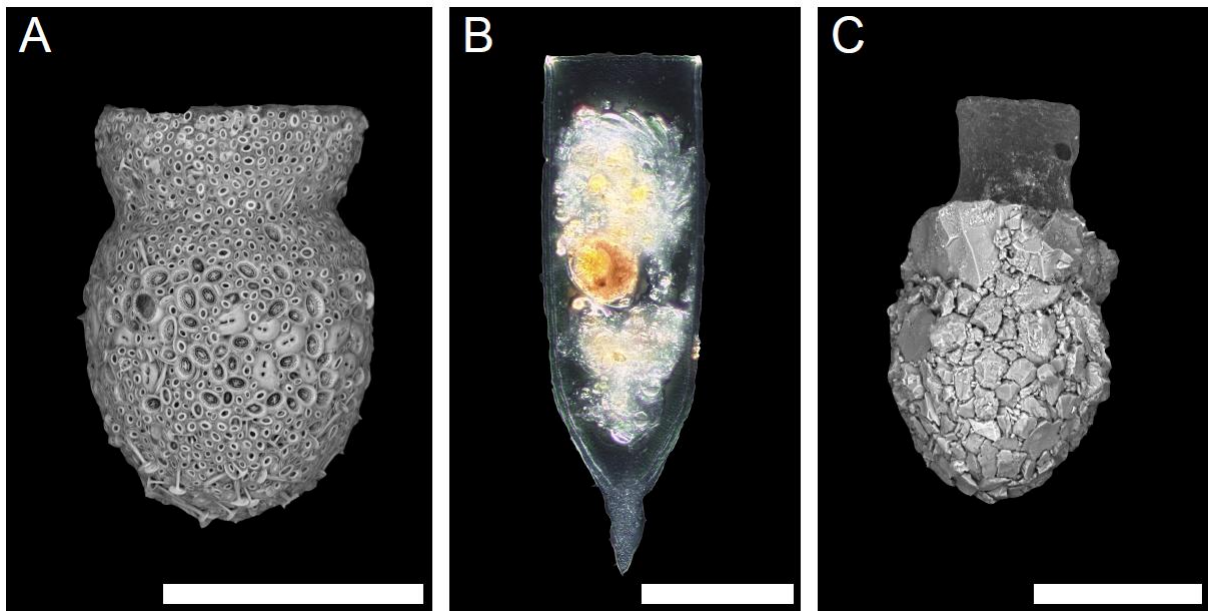


Figure 5. (A) Représentation d'un spécimen de *Codonella* sp. photographié au microscope électronique à balayage. La lorica est entièrement agglomérée avec des coccolithes provenant de différentes espèces de coccolithophoridés. (B) Représentation d'un spécimen de *Favella* sp. photographié au microscope optique. La lorica est entièrement hyaline. (C) Représentation d'un spécimen de *Codonellopsis* sp. photographié au microscope électronique à balayage. La lorica est en partie agglomérée avec des particules minérales et en partie hyaline au niveau du col.

3. Des dynamiques temporelles étudiées depuis longtemps

Au début du 20^e siècle, afin de comprendre les dynamiques des pêcheries, l'accent est mis sur l'aspect écologique de l'étude des micro-organismes, déjà considérés comme un maillon essentiel du réseau trophique marin. L'étude du plancton, après s'être focalisée sur la taxonomie, va se déplacer vers la biologie des communautés et des populations, avec une attention particulière portée sur les cycles annuels du plancton et les facteurs qui les contrôlent (croissance du phytoplancton, régénération des sels nutritifs, broutage par les hétérotrophes) (Davis 1955; Mills 1989). Les études quantitatives pour décrire les pics de croissance de certaines espèces planctoniques en milieu côtier ont été les précurseurs de l'océanographie biologique moderne, et avaient pour but d'expliquer les variations de composition des communautés planctoniques liées aux conditions environnementales. L'océanographie biologique a mis en évidence la variabilité temporelle des micro-organismes planctoniques, avec des successions d'assemblages d'espèces (Smayda 1980). Quelques-unes des premières études sur les changements dans les populations planctoniques ont concerné ou inclus les tintinnidés (e.g. Laackmann 1908; Lohmann 1908). De nombreux examens effectués dans des milieux côtiers et hauturiers ont mis en évidence des compositions de communautés changeantes en fonction des saisons. Les suivis temporels des tintinnidés ont été étudiés de façon approfondie en Mer Méditerranée (e.g. Bojanić et al. 2012; Margalef 1957; Sitran et al. 2007), et particulièrement dans la Baie de Villefranche-sur-Mer (Balech 1959; Dolan et al. 2006; Rassoulzadegan 1979), site propice à l'étude du plancton marin. Il semblerait que chez les tintinnidés, les espèces à lorica hyaline constituent la plupart de la communauté dans les mois d'été tandis que les espèces à lorica agglutinée dominent le reste de l'année (Dolan in press).

4. Les approches moléculaires

Les ciliés tintinnidés semblent être un groupe idéal pour explorer les liens entre la diversité morphologique et moléculaire. En effet, il est possible de les identifier par la morphologie de la lorica à l'aide d'un microscope optique, d'obtenir des images pour un archivage numérique et ensuite, d'amplifier un ou plusieurs gènes du même individu. Ceci permet de créer un lien direct entre les données moléculaires et morphologiques, et

I. Introduction

potentiellement conduire à la résolution de la controverse sur le choix de la lorica comme caractère pour correctement séparer les espèces de tintinnidés entre elles. Malgré la longue histoire de travaux taxonomiques et écologiques sur les tintinnidés, peu d'investigations ont été menées pour étudier la diversité génétique au sein de cet ordre. Les premières phylogénies moléculaires ont été basées sur l'amplification du gène d'ADNr 18S à partir de plusieurs cellules isolées à partir d'échantillons environnementaux et regroupées (Snoeyenbos-West et al. 2002). Comme les ciliés possèdent plusieurs copies de leur génome, l'amplification de marqueurs moléculaires est facilitée et ne nécessite pas d'étapes intermédiaires avant l'étape de PCR. Ainsi, plusieurs re-descriptions d'espèces de tintinnidés ou de travaux moléculaires s'intéressant à des espèces côtières de tintinnidés ont permis d'enrichir la base de données des séquences d'ADNr 18S par amplification par PCR directe sur cellule unique (Duff et al. 2008; Gao et al. 2009; Li et al. 2009; Snoeyenbos-West et al. 2002; Strüder-Kypke et Lynn 2003, 2008). Ces premiers résultats ont confirmé que l'agglomération des loricae semblait être un caractère ambivalent pour la reconstruction phylogénétique des tintinnidés. Cependant, la diversité génétique de peu de genres a été explorée et l'intérêt taxonomique s'est principalement porté sur le genre *Tintinnopsis*, délaissant la grande majorité de la diversité morphologique restante. Ensuite, des études de diversité moléculaire environnementale visant exclusivement les choreotrichidés (dont font partie les tintinnidés) et le groupe frère des oligotrichidés ont été menées afin d'étudier la répartition des communautés en fonction de gradients environnementaux (Doherty et al. 2007; Doherty et al. 2010; Tamura et al. 2011). Malheureusement, dû au faible nombre de séquences de référence et à la taille réduite des séquences générées (<700 nucléotides), les clones générés n'ont pas pu être identifiés au-delà du niveau taxonomique de l'ordre. Finalement, en dehors des études spécifiquement axées sur les ciliés mentionnées ci-dessus, une inspection rapide de la base de données de séquence GenBank révèle qu'il existe très peu de séquences environnementales de tintinnidés (approximativement une dizaine). Cela signifie que ces organismes sont peu détectés par les approches classiques de diversité utilisant des amorces générales eucaryotes. Ce qui peut s'expliquer par le fait que les tintinnidés sont rencontrés dans les eaux de surface du globe en relatives faibles concentrations (oscillant entre 1 et 10^4 individus par litre), des abondances qui caractérisent aussi d'autres groupes de protistes comme les radiolaires, foraminifères ou dinoflagellés parmi d'autres.

I. Introduction

II. Objectifs

L'objectif central de ce travail est d'améliorer notre compréhension sur le lien existant entre la connaissance classique des eucaryotes unicellulaires et leur diversité estimée à partir des données moléculaires, en particulier pour une meilleure interprétation des processus évolutifs et écologiques. Pour cela, j'ai choisi de travailler sur les ciliés tintinnidés comme groupe modèle car ils présentent une très grande diversité morphologique documentée depuis longtemps et sont communément rencontrés dans les écosystèmes marins. S'ajoute à ces avantages le fait qu'il est possible de les identifier aisément, de les isoler manuellement sous le microscope et d'amplifier des marqueurs ribosomiques (ADNr 18S et ITS) à partir de cellules individuelles.

Les objectifs plus spécifiques de ce travail sont :

- L'établissement d'une phylogénie solide des tintinnidés en incorporant les séquences d'individus représentatifs des différentes morphologies observées.

Quelle histoire évolutive nous raconte cette phylogénie moléculaire ? La diversité des formes observées est-elle directement liée à la diversité moléculaire ?

Pour répondre à ces questions, j'ai utilisé les gènes codants pour les ARN ribosomiques (ARNr 18S, 5.8S et 28S) et les espaces intergéniques correspondants (ITS1 et ITS2) amplifiés à partir de cellules individuelles de différents taxons.

- La comparaison des études environnementales basées sur la diversité morphologique et moléculaire.

La composition des communautés de tintinnidés observée dans un échantillon d'eau de mer avec un microscope est-elle efficacement décrite par les approches moléculaires d'amplification et de séquençage à partir d'ADN environnemental? Observe-t-on une grande richesse génétique que ne nous laissait pas soupçonner la diversité morphologique, c'est à dire, y-a-t-il des espèces cryptiques ?

Pour cela, je me suis concentré sur la comparaison détaillée de deux échantillons marins avec une observation exhaustive au microscope d'une part, et l'amplification et le séquençage des

II. Objectifs

régions de l'ADNr 18S et des ITS d'autre part en utilisant des méthodes classiques de clonage et séquençage Sanger ainsi que du pyroséquençage.

- La description de la dynamique temporelle des communautés de tintinnidés à différentes profondeurs dans la rade de Villefranche-sur-Mer (France).

Existe-t-il des différences de distribution ou retrouve-t-on toujours les mêmes espèces au cours d'une année ? Quelles sont les corrélations entre les assemblages d'espèces d'un échantillon et les conditions environnementales ? Jusqu'à quel niveau taxonomique ce type de suivi peut-il efficacement décrire la diversité ?

Nous avons utilisé une approche spécifique visant notre groupe modèle pour caractériser la diversité temporelle par des méthodes moléculaires d'analyse de banques de gènes de l'ARNr 18S et des régions des ITS.

- La description de la dynamique temporelle de l'espèce *Undella claparedei* à l'échelle intra-spécifique.

Quelle est l'ampleur de la diversité génétique intra-spécifique observée ? Cette diversité est-elle corrélée avec la période d'échantillonnage ? Avec les conditions environnementales ? Avec les variantes morphologiques de la lorica ? Que nous apprend cette variabilité sur l'écologie de l'espèce étudiée ?

Pour répondre à ces questions, nous avons amplifié l'ADNr 18S et l'ITS par PCR directe à partir d'un lot de cellules isolées sur une période de presque deux ans dans la rade de Villefranche-sur-Mer.

III. Matériel et Méthodes

A. La rade de Villefranche-sur-Mer : un site d'étude marin côtier

1. Paramètres environnementaux

La rade de Villefranche-sur-Mer est un lieu historiquement reconnu comme favorable à l'étude du plancton. Elle se situe au nord de la Mer Ligure (Figure 6). Elle est orientée dans la direction Nord-Sud, d'une longueur d'environ 3,5 km et d'une largeur de 1,5 km ; elle est relativement étroite et abritée par le cap de Nice à l'Ouest et par la presqu'île du Cap Ferrat à l'Est. La rade possède un seuil séparant un bassin peu profond au Nord (~10 m) d'une pente rapide qui plonge dans le canyon de Villefranche. La présence de ce canyon et l'absence de plateau continental induisent des profondeurs très grandes pour un site côtier, allant jusqu'à 200 m à l'entrée de la rade et 1 000 m à seulement 4 km de la côte. Cette particularité géographique et bathymétrique a des conséquences, tant hydrologiques qu'écologiques, notamment pour l'étude de phénomènes de production planctonique (Nival et al. 1975). L'hydrodynamisme de la rade de Villefranche-sur-Mer est influencé par le courant Ligure (Béthoux et al. 1988). Il existe depuis 1995 un suivi hebdomadaire des conditions hydrologiques de la rade au point B (43°41,10'N-7°18,94'E) qui appartient au réseau du Service d'Observation en Milieu Littoral (SOMLIT). Ainsi, chaque semaine, des relevés environnementaux sont réalisés et des paramètres environnementaux sont mesurés le long de la colonne d'eau comme la température, la salinité, le taux de chlorophylle a et d'autres paramètres communément employés en océanographie. La rade de Villefranche présente un profil caractéristique des eaux de la Mer Méditerranée, avec une stratification thermique de la colonne d'eau débutant généralement au mois d'avril pour se terminer au mois de novembre, limitant l'enrichissement en nutriments des couches superficielles.



Figure 6. Site d'échantillonnage au nord-ouest de la Mer Méditerranée (à gauche). Photo satellite de la rade de Villefranche-sur-Mer avec les positions du site d'échantillonnage B+ (identifiée par X) et du laboratoire océanographique (identifiée par O).

2. Stratégie d'échantillonnage

Notre campagne d'échantillonnage a été réalisée à bord du navire de recherche de la station océanologique (la *Sagitta*) au niveau du point B+ (profondeur du site d'environ 200 mètres ; 43°41'00"N, 7°18'44"E). Cette campagne s'est étalée pendant deux ans, et a compris 15 sorties d'échantillonnage. De février 2009 à février 2010, l'échantillonnage a été mensuel, puis saisonnier jusqu'à décembre 2010. Une stratégie d'échantillonnage couplant les prélèvements par filet à plancton et bouteille Niskin a été choisie dans le but d'analyser différentes facettes de la diversité des ciliés tintinnidés (Figure 7).

a. Trait de filet

Les échantillons de plancton ont été recueillis à chaque sortie au point B+ à l'aide d'un filet sur mesure de 53 μm de vide de maille, recommandé pour l'étude du microzooplancton. Ce filet avait un diamètre d'ouverture de 54 cm, était long de 280 cm, et muni d'un collecteur conçu pour concentrer la biomasse planctonique sans l'abîmer. Les prélèvements ont été effectués à la même profondeur de février 2009 à décembre 2010. Le filet à plancton, lesté d'un poids, a été plongé en oblique à 50 m de profondeur et traîné sur quelques dizaines de mètres autour du point B+. Le filet a été remonté en oblique. Le plancton présent dans le collecteur a été ensuite conservé dans un volume de 5 L d'eau de mer contenu dans un bac transparent jusqu'au retour à la station marine. De retour au laboratoire, une partie du plancton (500 mL) a été immédiatement fixée avec du Lugol à une concentration finale de 2%. Ce plancton fixé a servi à l'observation morphologique rapide des espèces présentes. Le reste de l'échantillon du trait de filet a été préparé selon la technique d'Utermöhl (1958). Un sous-échantillon de 50 mL a été versé dans un cylindre posé sur une chambre de sédimentation et laissé pendant 2 heures, le temps que le plancton sédimente au fond. Le cylindre a été enlevé en le glissant sur le côté et les observations réalisées ensuite au microscope inversé (Olympus IX51) équipé d'un appareil photo (Olympus DP71) afin de procéder à l'examen morphologique de la communauté planctonique. Les cellules identifiées comme des tintinnidés ont été prélevées individuellement à l'aide d'un fin capillaire fait sur mesure à partir d'une pipette pasteur. La cellule a été isolée par succion et placée successivement dans une série de gouttes d'eau de mer filtrée sur un filtre de 0,2 μm de diamètre de pore et stérilisée pour éviter les contaminations par d'autres organismes. Les cellules nettoyées ont été finalement introduites dans des tubes Eppendorf de 0,2 mL dans 30 μL d'éthanol pur. Les tubes ont été numérotés et conservés à 4°C ou à température ambiante dans le noir jusqu'à l'analyse moléculaire.

Au total, j'ai isolé plus de 877 cellules de tintinnidés pendant mon échantillonnage à Villefranche-sur-Mer ([Annexe 1](#)). Une partie a été utilisée pour caractériser la diversité génétique des marqueurs ribosomiques visés par les différentes analyses (voir partie IV.A). La quasi-totalité des spécimens d'*Undella claparedei* a été utilisée pour l'analyse populationnelle de la dernière partie des résultats (voir partie IV.D).

III. Matériel et Méthodes

Après rinçage, certains individus ont été déposés sur filtre de 0,2 μm et laissés à l'air libre pour sécher. Ces échantillons déshydratés ont pu être observés par microscopie électronique à balayage pour examiner la structure externe des loricae à haute résolution (voir photos de la [Figure 5](#) et de [la page de garde](#)).

b. Prélèvement d'eau à la bouteille Niskin

En parallèle des traits de filet, des prélèvements d'échantillons d'eau de mer ont été effectués au point B+ à cinq profondeurs différentes (5, 25, 50, 100, 160 mètres) à l'aide d'une bouteille Niskin de 12 L. L'eau de mer prélevée à chaque profondeur a été conservée dans des bidons à l'abri du soleil jusqu'au retour à la station marine. L'eau a été immédiatement préfiltrée à travers un filtre NITEX de 200 μm de maille afin d'exclure le mesoplancton (organismes supérieurs à 200 μm). Ensuite, chaque échantillon a été filtré indépendamment sur une rampe à filtration reliée à une pompe à vide. La totalité des 10 L a été filtrée au travers de filtres Millipore de 5 μm de diamètre de pore. Enfin, le filtre a été conservé dans des tubes numérotés de 2 mL dans l'éthanol pur jusqu'à l'analyse moléculaire.

75 échantillons d'eau de mer ont été collectés au total durant le suivi temporel dans la rade de Villefranche ([Annexe 2](#)).

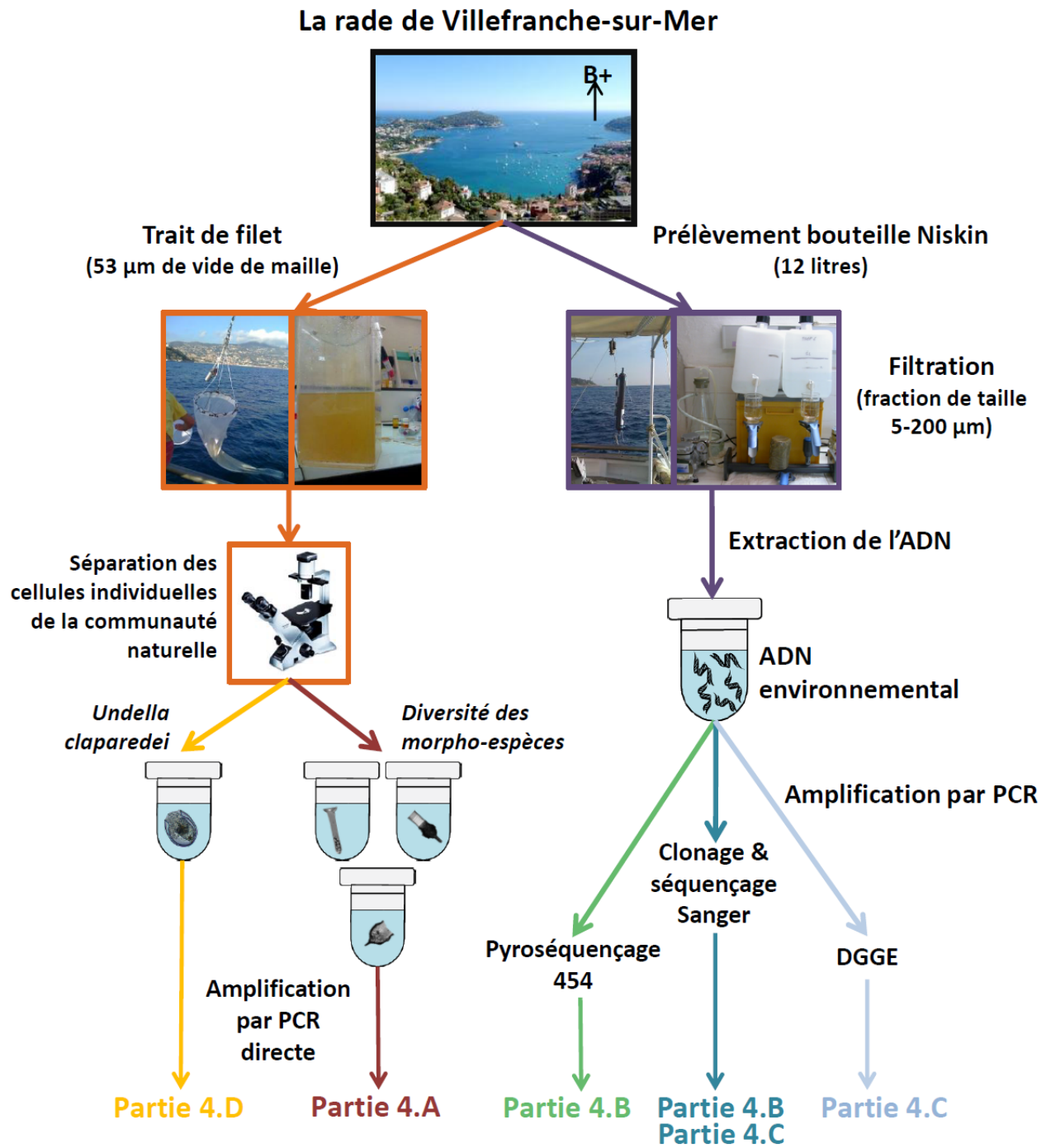


Figure 7. Schéma bilan de la stratégie d'échantillonnage et des méthodes d'analyse moléculaire utilisées lors de ce travail de thèse.

B. Méthodes d'analyse moléculaire

Les échantillons conservés (cellules et filtres) ont été transportés dans le laboratoire d'Orsay pour traitement moléculaire et analyses *in silico*.

1. Choix des marqueurs

Nous avons choisi comme marqueurs génétiques les gènes codant pour les ARNr 18S, 5.8S et une partie du 28S, ainsi que les ITS 1 et 2 (fragment total d'environ 2500 bp) pour réaliser des analyses de phylogénie moléculaire des tintinnidés à partir des cellules individuelles afin d'avoir un maximum de positions informatives. Dans le cas où le fragment total n'était pas amplifié avec succès, nous avons décidé de seulement cibler le gène codant pour l'ADNr 18S pour une seconde tentative (fragment total d'environ 1500 bp). Pour l'amplification, le clonage et le séquençage (Sanger) à partir d'ADN environnemental, un fragment d'ADN contenant l'ensemble des marqueurs génétiques cités ci-dessus a été visé (ADNr 18S + ITS 1 + 5.8S + ITS 2 + partie du 28S). Pour l'amplification par PCR et le pyroséquençage à partir d'ADN environnemental, des marqueurs plus petits ont été visés pour pouvoir être entièrement séquencés par cette technique, et très variables pour pouvoir différencier les séquences des différentes souches entre elles. Nous avons donc choisi d'amplifier une partie de l'ADNr 18S qui couvre la région hypervariable V4 (fragment d'environ 350 bp), ainsi que la région des ITS qui couvre l'ADNr 5.8S et des parties flanquantes de l'ITS 1 et 2 (fragment d'environ 260 bp). C'est également cette dernière région qui est visée dans l'analyse de *fingerprint* pour estimer la diversité des tintinnidés à l'aide de l'électrophorèse en gradient de gel dénaturant (*Denaturing Gradient Gel Electrophoresis*, DGGE). Pour l'étude de la diversité intra-spécifique d'*U. claparedei*, afin d'obtenir l'information phylogénétique contenue dans les séquences de l'ADNr 18S et de la région des ITS pour un maximum de cellules, j'ai choisi d'amplifier séparément les deux marqueurs (fragment de 1000 bp couvrant le 18S et un fragment de 800 bp couvrant la région des ITS (ITS 1 et 2, ADNr 5.8 S et 28S). En fonction de l'approche envisagée (cellules uniques, clonage, pyroséquençage, fingerprinting) des couples d'amorces différents ont donc été sélectionnés pour amplifier des fragments de gènes plus ou moins grands (de 260 à 2500 bp) (Table 1 et Figure 8). Nous avons en notre possession des jeux d'amorces pour amplifier les ADNr des eucaryotes en général. Des amorces ont été spécialement conçues dans cette étude pour amplifier spécifiquement des régions codant pour les ARNr et/ou les régions intergéniques correspondantes des tintinnidés (Table 1, Figure 8).

Table 1. Amorces utilisées pour les amplifications et le séquençage des gènes ARNr 18S et/ou ARNr 5.8S, 28S, ITS1 et ITS2.

Nom	d'amorce	Spécificité	sens	séquence (5' -> 3')	Référence
EK-42F	Eucaryotes	<i>forward</i>		CTCAARGAYTAAGCCATGCA	López-García et al 2001
EK-1498R	Eucaryotes	<i>reverse</i>		CACCTACGCAAACCTTGTTA	López-García et al 2003
EK-1379R	Eucaryotes	<i>reverse</i>		TGTGTACAAAGGGCAGGGAC	Wörheide et al 2002
EK-1520F	Eucaryotes	<i>forward</i>		GTAGGTGAACCTGCRG	López-García et al 2001
EK-28S564R	Eucaryotes	<i>reverse</i>		TGGTCCGTGTTTCRAGACG	López-García et al 2001
18S-Tin3F	Tintinnidés	<i>forward</i>		GCGGTATTTATTAGATAWCAGCC	cette étude
28S-TinR1	Tintinnidés	<i>reverse</i>		TGGTGCCTAGTATCAAAGT	cette étude
18S-Tin454F	Tintinnidés	<i>forward</i>		GTAGTGACAAGAAATAAYGG	cette étude
18S-Tin454R	Tintinnidés	<i>reverse</i>		CAACCCCTCARACCAACAG	cette étude
ITS-Tin454F	Tintinnidés	<i>forward</i>		CAATYTGTTGCAGRGCMAAGC	cette étude
ITS-Tin454R	Tintinnidés	<i>reverse</i>		AGCAATAGAAGGGCATCTA	cette étude
ITS-Tin454F-				CGCGCGCCGCGCCCCGCGCCCGTCCCGCCGCC	
Gcclamp	Tintinnidés	<i>forward</i>		CCCGCCCGCAATYTGTTGCAGRGCMAAGC	cette étude

2. À partir de cellules isolées, amplification par PCR directe des gènes d'ARNr 18S, 5.8S, 28S et des ITS

Les spécimens de tintinnidés fixés dans l'éthanol ont été sédimentés au fond des tubes par centrifugation pendant 5 min à 3000 rpm. L'éthanol a été ensuite évaporé dans un dessiccateur pendant 2 heures ou bien dans un SpeedVac pendant 40 min à température ambiante. Les cellules individuelles ont été directement resuspendues dans un volume final de réaction de PCR (*Polymerase Chain Reaction*) de 30 µL contenant 28,8 µL de tampon de réaction ExTaKaRa (TaKaRa, distribué par Lonza Cia., Levallois-Perret, France), 0,2 µL (5U/µL) de polymérase ExTaKaRa et 0,5 µL (10 pmoles par réaction) des amorces *forward* et *reverse* visant l'ADNr 18S, la région des ITS et le début du 28S des eucaryotes en général (voir Table 1, Figure 8). La réaction de PCR a été faite selon le programme suivant : une étape initiale de dénaturation de 3 minutes à 94 °C, suivi de 10 cycles en « touch-down » (dénaturation de 15 secondes à 94 °C, hybridation de 30 secondes avec une température décroissante de 1°C à chaque cycle de 65 °C à 55 °C, élongation de 2 minutes à 72 °C), puis de 20 cycles identiques avec une étape d'hybridation de 30 secondes à 55 °C, finalement suivi

III. Matériel et Méthodes

par une étape d'élongation de 7 minutes à 72 °C. Les produits de PCR révélant rarement des bandes visibles sur gel après cette première étape de PCR directement sur cellules individuelles, des PCRs nichées ou semi-nichées ont été réalisées. Aussi, pour éviter les possibles contaminations par d'autres eucaryotes (comme par exemple les parasites ou les proies digérées), des amorces spécifiques des tintinnidés situées à l'intérieur du fragment d'ADN amplifié ont été utilisées.

Une PCR nichée ou semi-nichée a été ensuite réalisée avec différents couples d'amorces en fonction des marqueurs visés selon l'étude (voir Table 1, Figure 8). Les PCRs ont été réalisées dans un volume final de 24 µL contenant 15,6 µL d'eau ultrapure, 4 µL de tampon de réaction 5X (Promega, USA), 0,5 µL de chaque amorce, 0,4 µL de dNTPs (stock à 10 mM), 0,2 µL de GoTaq (5 U/µL) (Promega, USA) et 3 µL du produit de la première PCR dilué 10 fois. Les réactions de PCR se sont déroulées en utilisant le même programme que décrit au dessus dans le cas de l'étude de phylogénie moléculaire (partie IV.A) et selon le programme suivant pour l'étude populationnelle d'*Undella claparedei* (partie IV.B): 50 cycles ont été réalisés (dénaturation de 15 secondes à 94 °C, hybridation de 30 secondes à 55 °C et élongation de 1 minute à 72 °C), précédés d'une étape de dénaturation de 3 minutes à 94 °C et suivis d'une étape finale d'élongation de 7 minutes à 72 °C. Les amplicons ont été entièrement séquencés par séquençage de type Sanger (BC Genomics) avec les amorces utilisées pour les PCRs et une amorce située plus à l'intérieur du fragment d'ADN amplifié pour les fragments supérieurs en taille à 1600 bp.

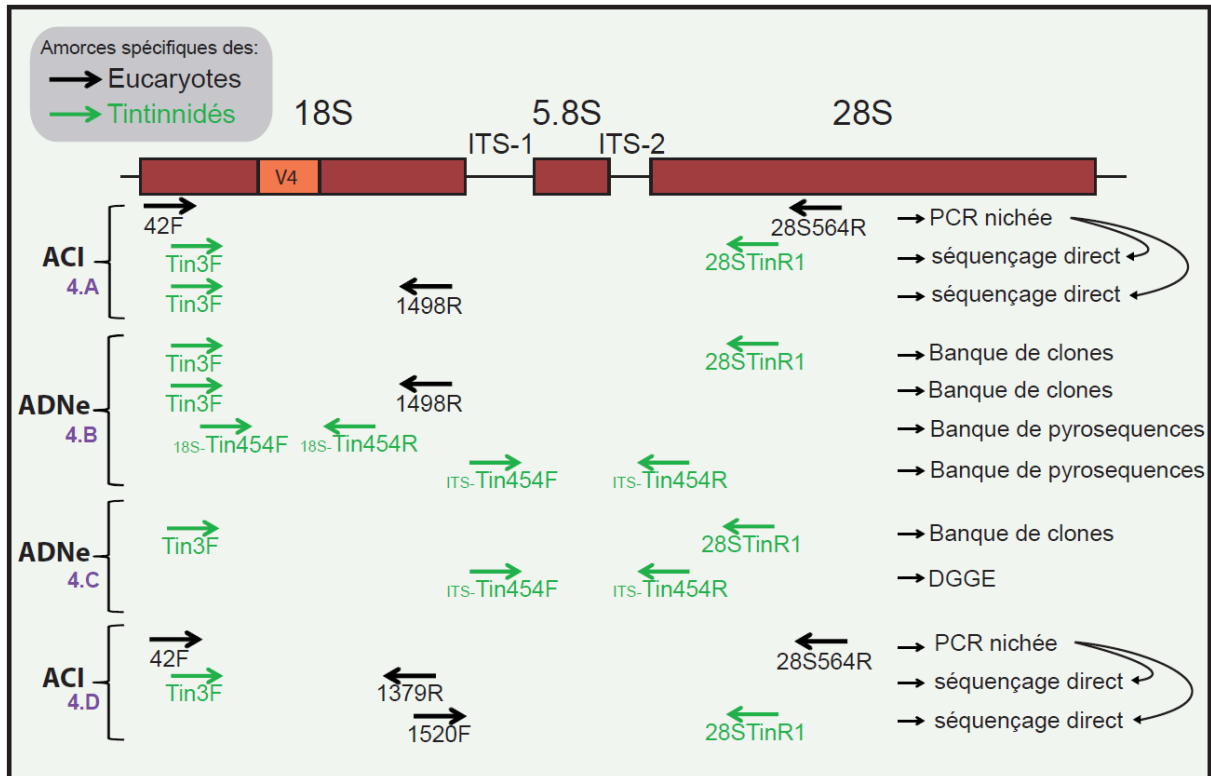


Figure 8. Représentation schématique de l'unité de transcription ribosomique (ADNr). Les régions codantes sont colorées en rouge. Les différents couples d'amorces utilisés dans les différentes études pour réaliser les PCRs sont indiqués par des flèches colorées en fonction de leur spécificité. Sur la gauche, les accolades indiquent les différents couples d'amorces utilisés dans les différentes parties de la thèse. Les fragments amplifiés par PCR directe sur cellules individuelles (ACI) sont soit réutilisés pour réaliser une PCR nichée, soit séquencés directement (indiqué à droite). Les fragments amplifiés à partir d'ADN environnemental (ADNe) sont ensuite soit clonés et séquencés pour construire des banques de séquences, soit pyroséquencés pour construire des banques de pyroséquences, soit analysés par DGGE (indiqué à droite).

3. Extraction de l'ADN environnemental

Les filtres Millipore de diamètre de pore de 5 µm ayant retenu la biomasse des communautés planctoniques de la gamme de taille comprise entre 5 et 200 µm ont été coupés en petits morceaux à l'aide d'un scalpel stérile. L'ADN a été purifié avec le kit PowerSoil DNA (MoBio, Carlsbad, CA, USA) selon les recommandations du fournisseur. Les acides nucléiques ont été resuspendus dans 100 µL de Tris-HCl à 10 mM, pH 8.

4. Clonage et séquençage des gènes d'ARNr et régions ITS

Nous avons utilisé différents couples d'amorces (Table 1, Figure 8) pour amplifier par PCR spécifiquement l'ADNr 18S (partie IV.B) et/ou l'ADNr 18S + 5.8S + ITS (partie IV.B et IV.C) des tintinnidés. Les PCR ont été réalisées dans un volume final de 24 μL contenant 15,6 μL d'eau ultrapure, 4 μL de tampon de réaction 5X (Promega, USA), 0,5 μL de chaque amorce à 20 μM (10 pmoles par réaction), 0,4 μL de dNTP (stock à 10 mM), 0,2 μL de GoTaq (5 U/ μL) (Promega, USA) et 2 μL d'ADN environnemental. La réaction de PCR a été faite selon le programme suivant : une première étape de dénaturation pendant 3 minutes à 94 °C suivie de 35 cycles comprenant chacun une étape de dénaturation de 15 secondes à 94 °C, une étape d'hybridation de 30 secondes à 52 °C et une étape d'élongation de 2 minutes à 72 °C, finalement suivie par une étape d'élongation de 15 minutes à 72 °C.

Les produits de PCR ont ensuite été clonés chez *Escherichia coli* avec le kit de clonage TopoTA (Invitrogen, Carlsbad, CA, USA) en suivant les recommandations du fournisseur. Les colonies d'*E. coli* blanches en milieu contenant du X-gal (ayant incorporé donc un insert dans le gène *lacZ*) ont été repiquées dans 50 μL de Tris à 10 mM. L'insert a été réamplifié à partir de ce produit avec les amorces du vecteur T7/M13R flanquant le site d'insertion suivant le même protocole de PCR que précédemment, exception faite de l'étape d'hybridation qui a été réalisée à 55°C.

Les fragments ont ensuite été séquencés partiellement (avec l'amorce *forward*) ou entièrement (avec les deux amorces *forward* et *reverse* plus une amorce interne, souvent 1498R) (Figure 8) par séquençage de type Sanger (BC Genomics).

5. Pyroséquençage des amplicons d'ADNr 18S

Nous avons utilisé deux couples d'amorces (voir Table 1, Figure 8) pour amplifier par PCR spécifiquement la région hypervariable V4 de l'ADNr 18S et une partie de la région des ITS (partie IV.B) des tintinnidés. Aux extrémités 5' des amorces ont été ajoutées des tags d'identification en multiplex (MID) différents pour chaque échantillon, qui permettent de regrouper les différents produits de séquençage lors du *run* de pyroséquençage et de les trier en fonction des échantillons lors de l'étape d'analyse *in silico*. Finalement, aux extrémités 5' des MID sont ajoutées des amorces de fusion A et B (respectivement pour les amorces

forward et *reverse*) nécessaires pour le séquençage. Les PCRs ont été réalisées dans un volume final de 25 µL de mélange de réaction Platinum HIFI (Invitrogen) contenant 19 µL d'eau ultrapure, 2,5 µL de tampon de réaction 10 X, 0,5 µL de dNTPs (stock à 10 mM), 0,75 µL de MgCl₂ pour une concentration finale de 1,5 mM, 0,75 µL de chaque amorce (15 pmoles par réaction), 0,125 µL (5U/µL) Taq Platinum et 2 µL d'ADN environnemental. La réaction de PCR a été faite selon le programme suivant : une première étape de dénaturation pendant 3 minutes à 94 °C suivie de 25 cycles comprenant chacun une étape de dénaturation de 15 secondes à 94 °C, une étape d'hybridation de 30 secondes à 52 °C et une étape d'élongation de 1 minute à 72 °C, finalement suivie par une étape d'élongation de 15 minutes à 72 °C. Seulement 25 cycles de PCR ont été effectués. Pour compenser la plus faible quantité d'ADN et limiter/moyenner les biais d'amplification liés à chaque réaction, 10 PCRs indépendantes ont été réalisées et les produits de PCR ont été regroupés. L'ADN a été purifié pour chaque échantillon et chaque fragment amplifié (V4 et ITS) et mélangé dans des proportions 1 : 1 : 1 : 1 avant d'être pyroséquéncé par un séquenceur 454 GS FLX utilisant la technologie Roche Titanium (Eurofins MWG Operon, Germany). Les milliers de séquences générées ont été ensuite analysées à l'aide des outils informatiques (voir ci-dessous).

6. DGGE (Denaturing Gel Gradient Electrophoresis)

Une région de l'ITS des tintinnidés a aussi été amplifiée à partir des échantillons d'ADN environnemental pour permettre de séparer les différents amplicons par migration sur gel à gradient dénaturant et d'obtenir ainsi un profil de la diversité par échantillon. Pour cela, la région ITS choisie (la même région courte d'environ 200 bp de l'ITS que celle étudié par pyroséquençage) a été amplifiée par PCR en utilisant des amorces spécifiques (voir Table 1, Figure 8). L'amorce *forward* contenait un *GCclamp* qui permet aux amplicons de rester en double brin pendant une bonne partie de la migration. Les réactions de PCR ont été faites dans un volume final de 20 µL de mélange de réaction Platinum HIFI comme décrit dans la partie III.B.5 (ci-dessus). La réaction de PCR a été faite selon un programme en « touch-down » : une étape initiale de dénaturation de 3 minutes à 94 °C, suivi de 10 cycles (dénaturation de 15 secondes à 94 °C, hybridation de 30 secondes avec une température décroissante de 1°C à chaque cycle de 60 °C à 50 °C, élongation de 90 secondes à 72 °C), puis de 20 cycles identiques avec une étape d'hybridation de 30 secondes à 50 °C, finalement suivis par une

étape d'élongation de 1 heure à 72 °C. Un système DGGE-2000 (CBS Scientific Company) a permis de réaliser l'électrophorèse des amplicons sur un gel de polyacrylamide à 8% présentant un gradient d'agent dénaturant (7 M d'urée et 40% de formamide) allant de 20% à 40%. La totalité du produit de PCR a été chargée pour chaque échantillon dans les puits du gel. Les produits de PCR ont migré dans le gel pendant 6 heures à 150 V et à 60 °C dans un bain de tampon Tris-acétate-EDTA (TAE) 0,5 X (1X: 40 mM Tris, 20 mM acide acétique, 1 mM EDTA, pH 8,0). À la suite de l'électrophorèse, le gel et l'ADN contenu dedans ont été plongés 45 min dans un bain de coloration (400 ml de TAE 0.5 X + 30 µL SYBR GOLD) et les gels ont été visualisés avec un transilluminateur UV et photographiés. Des bandes d'intérêt ont été découpées sous illumination UV, puis baignées dans 30 µL de Tris 10 mM (pH 8) à 30°C pendant une heure. Les fragments d'ADN ont été ensuite ré-amplifiés dans les mêmes conditions de PCR que précédemment en ajoutant par volume de réaction 4 µL du tampon de Tris (pH 8) avec l'ADN en suspension. Les fragments ont ensuite été séquencés entièrement (avec l'amorce *reverse*) par séquençage de type Sanger (BC Genomics). Les photographies des gels ont été ensuite traitées, normalisées en se basant sur la position du marqueur de taille, et analysées avec le logiciel Bionumerics® (AppliedMaths, Belgique). Les bandes ont été assignées manuellement. Une matrice de distance basée sur la présence/absence des bandes (coefficient Dice) a permis l'analyse comparée des profils et leurs groupements par ressemblance.

C. Analyses *in silico* des séquences

1. Contrôle de qualité des séquences Sanger et regroupement en UTOs

Les extrémités de basse qualité des séquences produites par la méthode Sanger ont été automatiquement éliminées grâce au logiciel *CodonCode aligner* (CodonCode Corporation ; www.codoncode.com). Dans le cas des fragments séquencés dans leur intégralité, les contigs ont été assemblés avec les paramètres par défaut de *CodonCode aligner*. Les chromatogrammes de séquençage ont été vérifiés manuellement pour corriger les positions ambiguës. Dans le cadre des analyses moléculaires par amplification, clonage et séquençage (Sanger), un fichier de données a été construit avec les séquences d'ADNr 18S + ITS générées par PCR directe sur cellules individuelles et les séquences des banques de clones. Le

Le jeu de séquences a été aligné avec le logiciel MAFFT (Katoh et al. 2002). Pour détecter des séquences chimériques, l'alignement a été analysé avec un ensemble d'algorithmes de détection d'événements de recombinaison implémenté dans le logiciel RDP v.4.14 (Recombination Detection Program, Martin et al. 2005). Si des événements de recombinaison étaient détectés entre des séquences attribuées à des espèces décrites de familles différentes, elles étaient considérées comme chimériques et éliminées du jeu de données. Les séquences ont été comparées par BLAST (Altschul et al. 1997) avec notre base de données de référence comprenant les séquences d'ADNr 18S et des régions ITS générées par PCR directe à partir des cellules individuelles des morpho-espèces de tintinnidés de référence (partie IV.A). Les UTOs ont été définies comme des groupes de séquences d'ADNr 18S seul ou avec la région des ITS partageant plus de 99% d'identité.

2. Reconstruction d'arbres phylogénétiques

Dans le cadre de l'étude de phylogénie moléculaire des tintinnidés (partie IV.A), un fichier de données a été construit avec les séquences d'ADNr 18S générées par PCR directe sur cellules individuelles, les séquences affiliées à l'ordre des Tintinnida disponibles sur GenBank (<http://www.ncbi.nlm.nih.gov/>) et des séquences de ciliés choreotrichidés ou oligotrichidés. Le jeu de séquences a été aligné avec le logiciel MAFFT (Katoh et al. 2002). Le choix des positions alignées sans ambiguïté s'est fait à la main avec l'ensemble de logiciels MUST (Philippe 1993), ou des méthodes automatisées telles que GBLOCKS (Castresana 2000) ou BMGE (Criscuolo et Grimaldo 2010) en utilisant les paramètres par défaut. Des analyses phylogénétiques ont ensuite été réalisées par maximum de vraisemblance (MV) avec le logiciel Treefinder (Jobb et al. 2004) ou par approche bayésienne (MB) avec le logiciel MrBayes (Huelsenbeck et Ronquist 2001). L'inférence phylogénétique par MV a été réalisée en appliquant un modèle GTR (General Time Reversible) d'évolution des séquences avec une loi gamma pour tenir compte des différentes vitesses d'évolution entre sites et 4 catégories de sites auxquelles s'ajoutent des sites invariables. Les valeurs des bootstraps, mesurant le support statistique des nœuds de l'arbre, ont été calculées en utilisant 1 000 pseudo-réplicats sous le même modèle d'évolution. L'inférence phylogénétique par MB a été réalisée avec 4 cycles indépendants et 1 000 000 d'itérations par cycle. Après échantillonnage des topologies obtenues dans le plateau des choix optimaux et élimination des 3 500 premiers

arbres, un arbre consensus a été reconstruit avec l'échantillonnage d'un arbre toutes les 1 000 itérations. Des comparaisons de topologies ont été faites en appliquant le test Approximately Unbiased (AU) (Shimodaira 2002) qui permet de tester statistiquement la vraisemblance d'hypothèses évolutives alternatives. Les arbres obtenus ont été visualisés avec le logiciel FIGTREE (Rambaut 2009).

Dans le cadre des études de diversité à partir d'ADN environnemental (parties IV.B et IV.C), un fichier de données a été construit avec les séquences générées par les différentes méthodes de séquençage. Des analyses phylogénétiques préliminaires ont été réalisées par une méthode de distance en « neighbor-joining » (NJ) avec l'ensemble de logiciels MUST. À partir des arbres préliminaires, une séquence représentative par UTO est sélectionnée. Les UTOs ont été définies comme des groupes de séquences partageant plus de 99% de similarité (voir parties IV.A et IV.B), valeur qui permet de discriminer efficacement les différentes espèces de tintinnidés rencontrés dans la rade. Ces séquences représentatives ont ensuite été réalignées avec les séquences de référence les plus proches, majoritairement générées par l'étude de la partie IV.A. Des analyses phylogénétiques par maximum de vraisemblance comme décrit ci-dessus ont été réalisées.

3. Filtrage et débruitage des pyroséquences

Dans le cadre de l'étude environnementale réalisée dans la partie IV.B, les séquences générées par pyroséquençage nécessitent des étapes de filtration et de débruitage afin de corriger les erreurs de séquençage systématiquement produites par les plateformes de pyroséquençage. Nous avons préparé un jeu de pyroséquences par échantillon (identifié par des MIDs spécifiques) et par marqueur moléculaire (ADNr 18S vs ITS). Les séquences présentant des sites ambigus (N) ou des erreurs de séquençage dans la partie des amorces et celles qui ne couvrent pas la totalité du marqueur amplifié ont été éliminées. Les séquences ont été comparées par BLAST (Altschul et al. 1997) avec notre base de données de référence comprenant les séquences d'ADNr 18S et des régions ITS générées par amplification directe par PCR à partir des cellules individuelles des morpho-espèces de tintinnidés de référence (partie IV.A). Les pyroséquences dont la première sortie présente moins de 97% de similarité avec les séquences de référence sont vérifiées par analyse phylogénétique rapide par NJ. Si elles branchent parmi les tintinnidés, elles ont été gardées, sinon, elles ont été exclues. La

possibilité de présence de séquences chimériques est vérifiée avec le programme ChimeraSlayer (Haas et al. 2011) implémenté dans Mothur (Schloss et al. 2009). Une matrice de distance a été générée par comparaison des séquences deux à deux (Pair-wise distance) à l'aide du module NDIST implémenté dans le programme AmpliconNoise (Quince et al. 2011). Les erreurs de séquençage ont été corrigées soit par les méthodes d'agglomération SLP (Huse et al. 2010) ou CD-HIT-OTU (Wu et al. 2011; <http://weizhonglab.ucsd.edu/metagenomic-analysis/server/cd-hit-otu/>), soit par la méthode probabiliste itérative AmpliconNoise (Quince et al. 2011). Les UTOs ont été définies comme des groupes de pyroséquences d'ADNr 18S ou d'ITS partageant plus de 99% d'identité.

4. Analyses statistiques des données génétiques intra-spécifiques

Dans le cadre de l'étude de diversité intra-spécifique d'*Undella claparedei* (partie IV.D), un fichier de données a été construit avec les séquences d'ADNr 18S et un autre avec les séquences de la région de l'ITS générées par PCR directe sur cellules individuelles. Les jeux de séquences ont été alignés avec MAFFT. L'analyse de variance moléculaire (AMOVA) avec le logiciel Arlequin (Excoffier et al. 2005; Excoffier et al. 1992) s'effectue pour évaluer la distribution de la diversité dans ou entre les populations définies soit par mois d'échantillonnage (si la population comprend plus de 10 individus) soit par condition de stratification de la colonne d'eau (stratifiée vs mélangée). Nous avons calculé l'indice de fixation F_{st} , qui mesure le degré de différenciation génétique des sous-populations par le calcul des variances standardisées des fréquences d'allèles des populations. La signification statistique se calcule pour les valeurs de F_{st} entre paires de populations pour tester l'hypothèse nulle d'un manque de différenciation génétique entre les populations et, par conséquent, la partition de la diversité génétique.

Les séquences concaténées de l'ADNr 18S et de la région des ITS obtenues par PCR directe sur les cellules d'*U. claparedei* ont été alignées avec MAFFT. La recombinaison entre les séquences a pour conséquence la génération de différents patrons de polymorphisme dans la population, que l'on peut comparer au patron de polymorphisme attendu sous l'hypothèse nulle de reproduction clonale. Ces indices de recombinaison sont détectables par différents tests statistiques : MAXCHI (Maynard Smith 1992), GENECONV (Sawyer 1989), qui

III. Matériel et Méthodes

défectent l'hétérogénéité du patron de polymorphisme aux alentours du point de recombinaison, et les tests de Mantel (appelés LDr2 et LD|D'|) qui évaluent la significativité de la relation entre le déséquilibre de liaison et la distance entre les sites polymorphes.

IV Résultats

A. Phylogénie moléculaire des ciliés tintinnidés marins

1. Contexte et objectifs

Le premier objectif général de ma thèse était de construire une phylogénie robuste des ciliés tintinnidés qui serve de référence pour le reste du travail, en mettant en relation la diversité génétique des ciliés avec la diversité morphologique décrite classiquement. Pour ce faire, nous avons isolé à partir d'échantillons de plancton prélevés à Villefranche-sur-Mer, à Marseille et à Banyuls-sur-Mer des individus de différentes morpho-espèces bien distribués au sein de la diversité des tintinnidés rencontrée en Mer Méditerranée. De plus, nous avons incorporé à cette étude des espèces vivant dans des milieux lacustres afin de tester si ces organismes branchent au sein des groupes marins et se sont adaptés indépendamment (multiples adaptations) aux milieux d'eau douce ou s'ils forment un groupe séparé dont l'ancêtre s'est adapté au milieu d'eau douce (un seul événement d'adaptation). Ces tintinnidés lacustres ont été récoltés dans des étangs de la région parisienne et dans un lac côtier à l'est de l'Espagne (Albufera de Valencia). À partir des cellules individuelles, nous avons amplifié et séquencé les gènes codant pour l'ARNr 18S de 62 individus représentant 29 espèces marines et 3 espèces d'eaux douces. De plus, pour 25 de ces individus représentant 17 espèces, nous avons également amplifié et séquencé les gènes codant pour le 5.8S et une partie du 28S, et les régions intergéniques correspondantes (ITS1 et ITS2). Le premier objectif spécifique de cette partie du travail était d'enrichir la base de données des séquences d'ADNr 18S avec des espèces pour lesquelles des données moléculaires n'étaient pas disponibles jusqu'à maintenant. Ensuite, l'objectif était de construire une phylogénie basée sur les séquences obtenues afin de corréliser l'histoire évolutive expliquée par les données moléculaires et la classification proposée par les taxonomistes sur les critères morphologiques de la lorica. Enfin, l'objectif final de cette partie du travail était d'offrir un cadre phylogénétique de référence pour les analyses moléculaires de la diversité à partir d'ADN environnemental.

Ce travail a donné lieu à la publication de l'article « Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora) » dans la revue *Protist* (Bachy, Gómez, López-García, Dolan and Moreira. *e-pub ahead of print* le 8 février 2012).

**2. Manuscrit de l'article 1 : Molecular Phylogeny of Tintinnid Ciliates
(Tintinnida, Ciliophora)**

ORIGINAL PAPER

Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora)

Charles Bachy^a, Fernando Gómez^b, Purificación López-García^a, John R. Dolan^c, and David Moreira^{a,1}

^aUnité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, Bâtiment 360, 91405 Orsay Cedex, France

^bInstituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, PO Box 22085, 46071 Valencia, Spain

^cUniversité Pierre et Marie Curie and Centre National de la Recherche Scientifique (CNRS), UMR 7093, Laboratoire d'Océanographie de Villefranche, Marine Microbial Ecology, Station Zoologique, B.P. 28, 06230 Villefranche-sur-Mer, France

Submitted October 6, 2011; Accepted January 4, 2012
Monitoring Editor: Hervé Philippe

We investigated the phylogeny of tintinnids (Ciliophora, Tintinnida) with 62 new SSU-rDNA sequences from single cells of 32 marine and freshwater species in 20 genera, including the first SSU-rDNA sequences for *Amphorides*, *Climacocylis*, *Codonaria*, *Cyttarocylis*, *Parundella*, *Petalotricha*, *Undella* and *Xystonella*, and 23 ITS sequences of 17 species in 15 genera. SSU-rDNA phylogenies suggested a basal position for *Eutintinnus*, distant to other Tintinnidae. We propose Eutintinnidae fam. nov. for this divergent genus, keeping the family Tintinnidae for *Amphorellopsis*, *Amphorides* and *Steenstrupiella*. *Tintinnopsis* species branched in at least two separate groups and, unexpectedly, *Climacocylis* branched among *Tintinnopsis* sensu stricto species. *Tintinnopsis* does not belong to the family Codonellidae, which is restricted to *Codonella*, *Codonaria*, and also *Dictyocysta* (formerly in the family Dictyocystidae). The oceanic genus *Undella* branched close to an undescribed freshwater species. *Metacylis*, *Rhabdonella* and *Cyttarocylis* formed a well supported clade with several *Tintinnopsis* species at a basal position. *Petalotricha ampulla* and *Cyttarocylis cassis* SSU-rDNA and ITS sequences were identical or almost identical. Therefore, we propose *Cyttarocylis ampulla* comb. nov. for them. Intensive use of single-cell isolation and sequencing revealed unexpected complexity in the evolutionary history of these relatively well-studied ciliates. Notably, the diversity of freshwater forms suggests multiple marine-freshwater invasions.

© 2012 Elsevier GmbH. All rights reserved.

Key words: Alveolates; rDNA phylogeny; loricate ciliates; single-cell PCR.

¹Corresponding author; fax +33169154697.

e-mail charles.bachy@u-psud.fr (C. Bachy), fernando.gomez@fitoplancton.com (F. Gómez), puri.lopez@u-psud.fr (P. López-García), dolan@obs-vlfr.fr (J.R. Dolan), david.moreira@u-psud.fr (D. Moreira).

Abbreviations: AU, Approximately Unbiased; bp, base pairs; BI, Bayesian Inference; BV, bootstrap value; ITS, Internal Transcribed Spacer; ML, maximum likelihood; PCR, polymerase chain reaction; rDNA, ribosomal deoxyribonucleic acid; s.l., sensu lato; s.s., sensu stricto; SSU, small subunit.

Introduction

Microzooplankton, composed largely of ciliates and heterotrophic dinoflagellates, plays a major role in the transfer of energy and matter through the pelagic food web by consuming a majority of the primary production (Calbet and Landry 2004). Planktonic ciliates are ubiquitous and much attention has been given to their role as primary consumers of pico- and nanosized autotrophs and heterotrophs, as well as nutrient regenerators and as important food sources for metazoan zooplankton and fish larvae (Dolan et al. 2002; Pierce and Turner 1992). They also represent one of the most morphologically diverse groups in the plankton. Commonly, planktonic ciliates are divided into loricate (tintinnids) and aloricate forms. The identification of aloricate ciliates requires slide mounting, cytological staining (i.e., protargol silver staining) and examination under high magnification (Montagnes and Lynn 1987). By contrast, by examining untreated specimens in plankton settling chambers using an inverted microscope, tintinnids can be easily differentiated from other species based on their secreted shell, the lorica. These loricae range in size from ca. 20 μm up to several hundred micrometers, and can be hyaline or agglutinated with mineral or biological particles attached.

The existing classification of tintinnids is entirely based on the characteristics of the loricae, including differences in the size, general shape, ornamentation, fine-scale surface structures and the presence or absence of agglutinated particles. There are numerous detailed descriptions at the species level, especially with the conspectus of Kofoid and Campbell (1929) in which 627 species were distinguished based on characteristics of the lorica. The first attempt to reconstruct tintinnid phylogeny was also lorica-based and considered the agglutinated lorica as an ancestral character (Kofoid and Campbell 1939). However, reliance on lorica characteristics in tintinnid taxonomy is problematic. In fact, some lorica characteristics, typically its length, have long been known to be quite variable (e.g. Gold and Morales 1975; Laackmann 1908). In natural populations, significant lorica plasticity has been suspected in a variety of species (e.g. Boltovskoy et al. 1990; Davis 1981; Santoferrara and Alder 2009; Wasik and Mikolajczyk 1994). For example, a morphometric study of the genus *Cymatocylis* (Williams et al. 1994) found that 45 lorica morphologies described as separate species could only be reliably sorted into 5 distinct morphotypes. However, the real limitations of lorica-based taxonomy

and phylogeny were revealed using cultures with the demonstration that a single species of *Favella* can show lorica characteristic thought to be typical of different genera (Laval-Peuto 1977, 1981, 1983). Unfortunately, for the overwhelming majority of species, there is no information on other morphological characters such as the ciliary pattern or the kinetome. The structure of the kinetome is known for only twenty species, too few for a rigorous analysis and insufficient for the demonstration of clear apomorphies within the Tintinnida (Agatha 2010a; Agatha and Strüder-Kypke 2007; Choi et al. 1992; Foissner and Wilbert 1979).

Ciliates of the subclass Choreotrichia Small & Lynn, 1985 are characterized by an adoral zone of polykinetids, used in locomotion and feeding, which forms a closed membranellar zone (Lynn 2008). Within this subclass, the presence of a lorica distinguishes the order Tintinnida Kofoid & Campbell, 1929 from the order Choreotrichida (strobiliids). In the last decade, advances in molecular phylogeny have allowed evaluation of these diagnostic criteria. Most of the available sequences corresponded to small subunit rRNA gene (SSU-rDNA) sequences. The first studies were based on species available in culture or from pooling cells collected in field samples (Snoeyenbos-West et al. 2002; Strüder-Kypke and Lynn 2003). Nevertheless, most species are difficult to culture or to find in abundance in natural samples. Fortunately, techniques such as single-cell PCR and sequencing are especially useful for tintinnids and other ciliates that usually have numerous copies of highly expressed genes in their macronuclear genomes (Prescott 1994).

There are more than 50 accepted genera of extant tintinnids (Lynn 2008). However, publicly available sequences are restricted only to species of 12 genera, with a strong bias towards coastal species of temperate waters (Duff et al. 2008; Gao et al. 2009; Li et al. 2009; Snoeyenbos-West et al. 2002; Strüder-Kypke and Lynn 2003, 2008). Nevertheless, tintinnid diversity is highest in tropical or sub-tropical waters (Dolan et al. 2006; Gómez 2007a). Analyses of the relatively small number of sequences have confirmed the monophyly of Tintinnida and supported some higher order groups of families but also have suggested the paraphyly of certain genera (Strüder-Kypke and Lynn 2003, 2008). This suggested that the use of lorica morphology to form natural groups or to reconstruct the evolutionary history of tintinnids is questionable.

Insufficient taxon sampling, namely the lack of molecular information for numerous tintinnid genera, has greatly hindered advances in determining the classification and evolutionary history of

tintinnids. We have addressed this issue by providing a comprehensive molecular phylogeny of Tintinnida and establishing diagnostic characters for a classification supported by new 62 SSU-rDNA sequences of 32 species from marine and freshwater environments. SSU-rDNA sequences of members of the genera *Amphorides* Strand, *Climacocylis* Jörgensen, *Codonaria* Kofoid & Campbell, *Cyttarocylis* Fol, *Parundella* Jörgensen, *Petalotricha* Kent, *Undella* Daday, *Xystonella* Brandt are determined for the first time. We also provide 25 ITS sequences of 17 species in 15 genera in order to compare the phylogeny of Tintinnida using another ribosomal molecular marker.

Results

Specimens Isolated for SSU-rDNA Sequencing

Over the two-year period of our study, a total of 62 SSU-rDNA sequences from 32 morphologically-distinguished species from marine and freshwater environments were determined. All specimens were individually identified, photographed and isolated under the microscope. These species are representative of a wide range of loricae morphologies, lengths, shapes, number of apertures and type of agglomeration (Fig. 1; Supplementary Fig. S1). In some cases, for a single species we included specimens collected in different locations and dates. In addition, ITS and 5.8 rRNA gene sequences were obtained for 17 species. All isolates for which we obtained either the SSU-rDNA sequence or both the SSU-rDNA and ITS sequences are listed in Supplementary Table S1. A detailed description of the isolated specimens classified by family is available in the Supplementary Materials.

SSU-rDNA Phylogeny

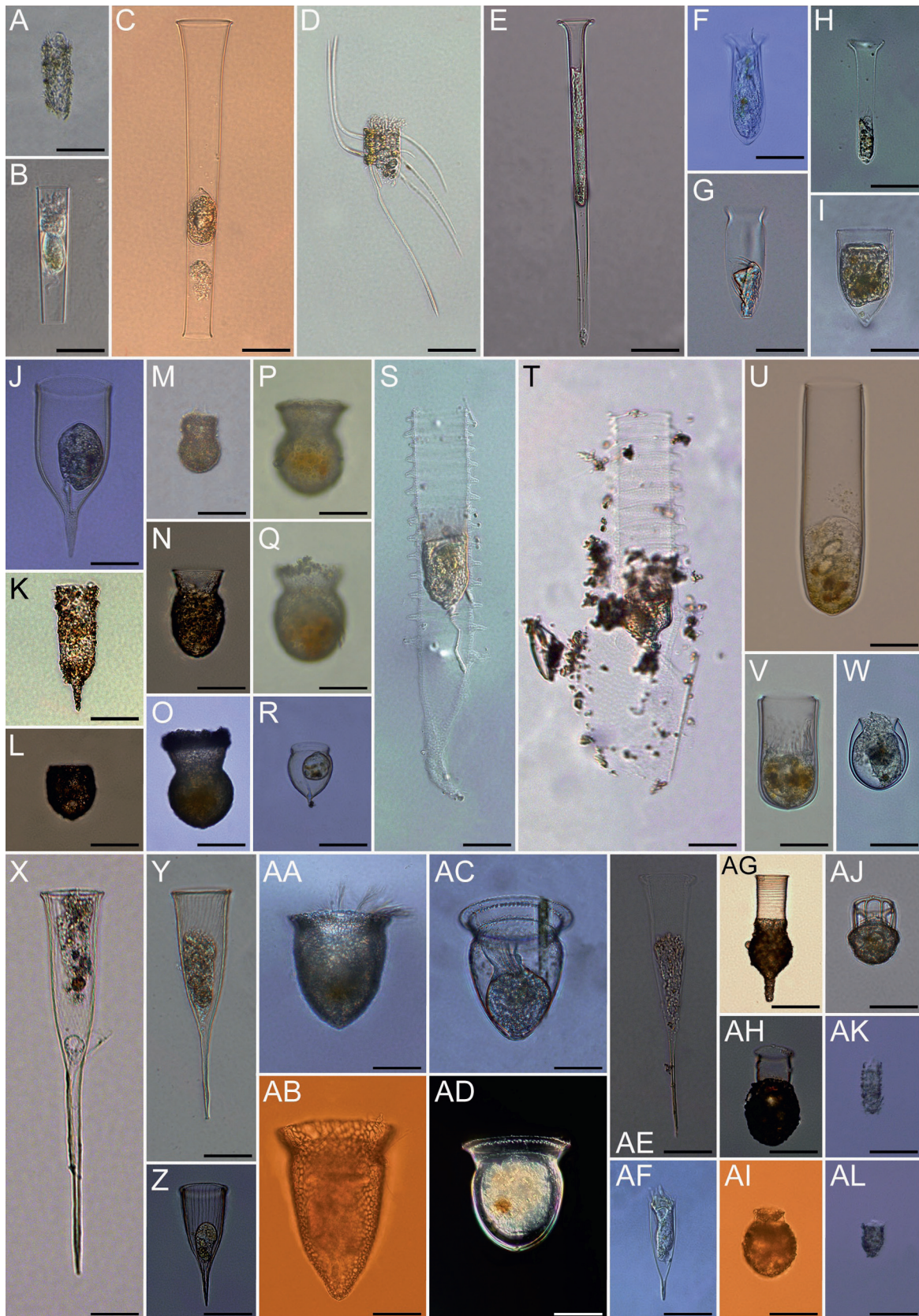
Sequence identity between the tintinnid SSU rDNA sequences ranged from 90% to 100%. In particular, the sequences of *Codonella*, *Codonellopsis*, *Stenosemella* and *Dictyocysta* species were very similar (98%-100%). By contrast, the sequences of the congeneric species of *Eutintinnus* exhibited higher dissimilarity (ranging from 95%-97% nucleotide identity). This implies a higher difference between the *Eutintinnus* species than between other species belonging to different genera. We also found possible evidence of intra-specific genetic variation in the SSU-rDNA locus of the species *Amphorides quadrilineata*. The distance

between the two isolates FG1141 and FG293 (from two different Mediterranean sites) was almost 3%, while genetic variation was not found in other tintinnid species (e.g. *Tintinnopsis cylindrica* or *Codonellopsis morchella*).

There were no differences among the maximum likelihood (ML) and Bayesian inference (BI) tree topologies for all well-supported nodes in the tintinnid SSU-rDNA phylogeny. Moreover, the use of different sequence alignment and masking methods did not affect these well-supported nodes. Oligotrichs and aloricate choreotrichs branched as sister groups of a large monophyletic group containing all tintinnid sequences with the exception of *Tintinnidium* species (Fig. 2). In fact, in all our analyses, *Tintinnidium* species branched close to the Strombidiidae, although with weak statistical support (ML bootstrap value -BV- 60%, Bayesian posterior probability -PP- 0.67). To check the robustness of this result, we carried out an Approximately Unbiased (AU) test on the SSU-rDNA data set to assess the monophyly of tintinnid species. The test could not significantly reject the constrained topology where all the tintinnids *sensu stricto* (with a lorica) were forced to be monophyletic ($p = 0.22$) (Supplementary Table S2).

In addition to the divergent clade of *Tintinnidium*, the species of the genus *Eutintinnus* formed a basal branching group within the tintinnid core. This clade is divided into two sister groups, the first composed of *Eutintinnus apertus*, *E. pectinis* and an unidentified species of this genus, and the second of *Eutintinnus fraknoi* and *E. tubulosus* (Fig. 2). The genera *Salpingella*, *Amphorellopsis*, *Amphorides*, and *Steenstrupiella* branched together in a clade with a strong support (BV 100%). In the classical taxonomical schemes, these genera are placed in the family Tintinnidae. Within this clade, our sequence identified as *Salpingella acuminata* branched in a basal position, forming a paraphyletic group with a second sequence retrieved from GenBank under the same name *S. acuminata* (EU399536). The sequences of *Amphorellopsis*, *Amphorides*, and *Steenstrupiella* formed a well supported group (BV 100%) with a long basal branch, indicating an acceleration of evolutionary rate before the diversification of these three genera. Our five sequences of *Amphorides quadrilineata* showed different positions. Three of them formed a group with *Steenstrupiella* whereas, surprisingly, the sequences of our isolates *A. quadrilineata* FG1141 and FG249 formed a group with *Amphorellopsis acuta*. We were unable to clearly find morphological differences between the five *A. quadrilineata* specimens (Fig. 1G and H).

4 C. Bachy et al.



Under the same morphology these specimens clearly formed two cryptic phylotypes.

A clade with a strong support (BV 100%) was formed exclusively of species of *Favella*, with the exception of a GenBank sequence named *Favella taraikaensis* (FJ196073), which branched close to *Metacyclis* spp. The sequences of our isolates *Favella azorica* FG1106 and *F. azorica* FG1111 were identical between them and to that of *F. campanula* (Schmit) Jörgensen, (GenBank FJ422984), while the isolate *F. adriatica* FG1102 was more distantly related to them. All these specimens were collected in the same location, Bay of Villefranche, and in an interval of one week (Supplementary Table S1).

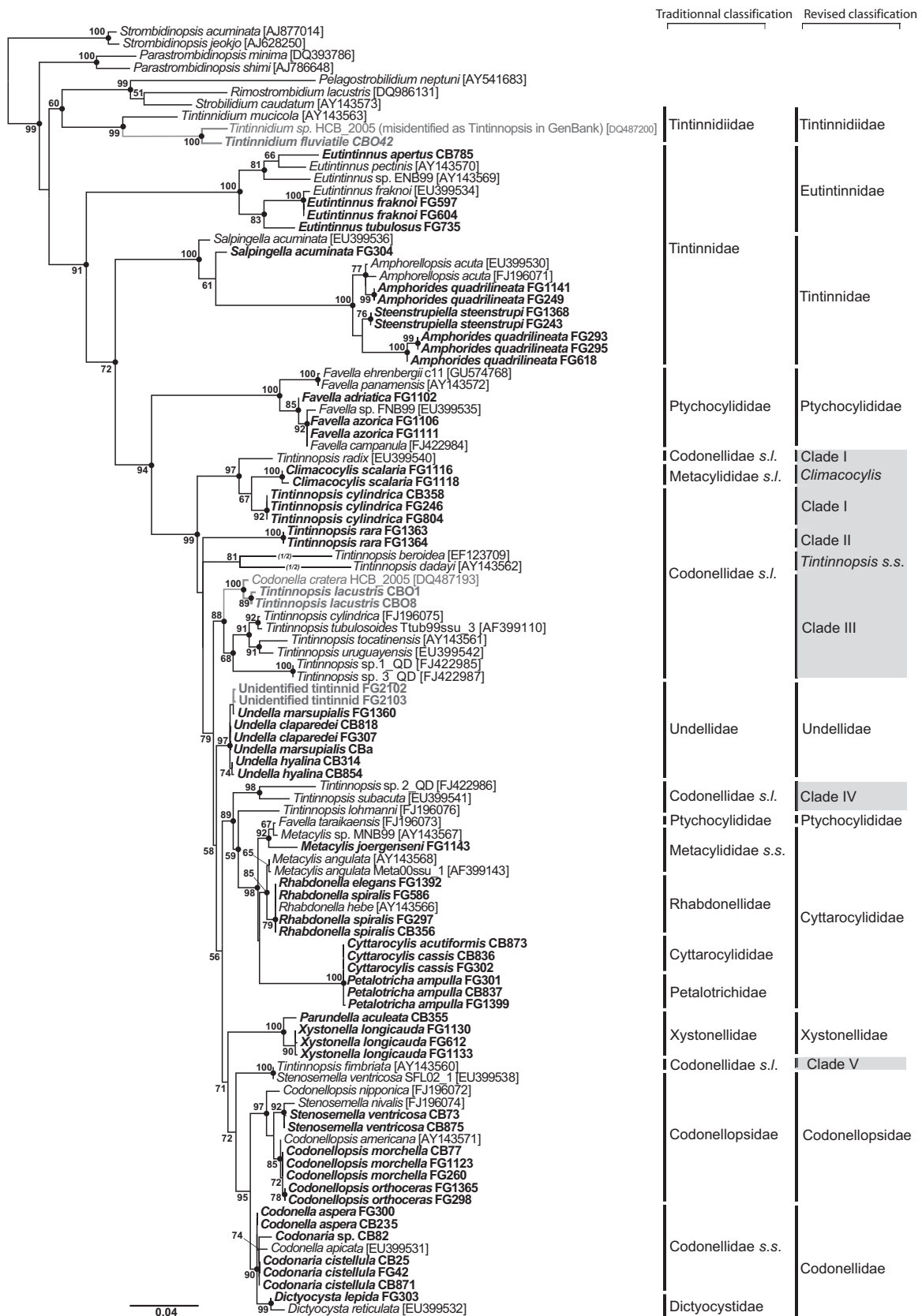
The rest of species emerged within a strongly supported group (BV 99%, PP 1), within which different sub-clades grouped species with high support but the relationships among them remained unresolved. Notably, the sequences of the genus *Tintinnopsis* were scattered among several sub-clades. A sequence retrieved from GenBank under the name *Tintinnopsis beroidea* Stein (EF123709), if correctly identified, corresponded to the *Tintinnopsis* type species (Li et al. 2009). Most other *Tintinnopsis* species were grouped in several clades branching in the same region of the tree as the type species. Nevertheless, there were two groups distantly related to the group of the type species. One was composed of *T. subacuta* Jörgensen and *T. lohmanni* Laackmann in addition to sequences of *Favella*, *Metacyclis* *Rhabdonella*, *Cyttarocyclus* and *Petalotricha* species. The second one was represented only by a sequence named *Tintinnopsis fimbriata* Meunier (Strüder-Kypke and Lynn 2003), which was 99% identical to a *Stenosemella ventricosa* sequence, suggesting

the possible misidentification of the sequence submitted to GenBank. Interestingly, the sequences of *Climacocylis* branched in the main group of sequences of *Tintinnopsis* (BV 97%, Fig. 2). This was unexpected because the lorica morphology of *Climacocylis* is radically different from *Tintinnopsis* (see Fig. 1K and S) and no taxonomic scheme has ever proposed a close relationship between these genera.

As all the *Tintinnopsis* species did not cluster together and appeared in different subclades in the tree, AU tests were carried out on the SSU-rDNA data set to assess the polyphyly of *Tintinnopsis* spp. (Supplementary Table S2). The test significantly rejected the constraint that all *Tintinnopsis* species formed a monophyletic clade ($p=0$), contradicting the lorica-based morphological classification. However, as long as we did not consider *Tintinnopsis fimbriata* AY143560, *Tintinnopsis subacuta* EU399541, *Tintinnopsis* sp. 2 FJ422986 and *Tintinnopsis lohmanni* FJ196076, the AU test could not reject the constrained topology where the 11 other *Tintinnopsis* spp. were forced to be monophyletic ($p=0.16$). We could not discard the possibility that the fast evolving sequences of the type species *Tintinnopsis beroidea* and of *Tintinnopsis dadayi* (see the long branches in Fig. 2) formed a monophyletic group because of a long-branch attraction artefact (Bergsten 2005).

A well-supported clade was composed of the species of *Undella* (BV 97%). The sequence difference between *Undella marsupialis*, *U. claparadei* and *U. marsupialis* were extremely low (>99.9% identity). Unexpectedly, two sequences of two specimens of an unidentified species branched in this clade. These specimens were collected in a freshwater lake and showed a lorica with

Figure 1. Light micrographs of specimens of *Tintinnida* collected for single-cell PCR analysis **A-AL**. See Table S1 for collection dates, locations and accession numbers. **A.** *Tintinnidium fluviatile* isolate CBO42. **B.** *Eutintinnus tubulosus* isolate FG735. **C.** *E. fraknoi* isolate FG604. **D.** *E. apertus* isolate CB785. **E.** *Salpingella acuminata* isolate FG304. **F.** *Amphorides quadrilineata* isolate FG1141. **G.** *Amphorides quadrilineata* isolate FG293. **H.** *Steenstrupiella steenstrupi* isolate FG1368. **I.** *Favella azorica* isolate FG1111. **J.** *F. adriatica* isolate FG1102. **K.** *Tintinnopsis cylindrica* isolate CB358. **L.** *Tintinnopsis rara* isolate FG1364. **M.** *Tintinnopsis lacustris* isolate CBO1. **N.** *Codonella aspera* isolate FG300. **O.** *Codonaria cistellula* isolate FG42. **P.** *Codonaria cistellula* isolate CB871. **Q.** *Codonaria* sp. isolate CB82. **R.** *Metacyclis joergenseni* isolate FG1143. **S.** *Climacocylis scalaria* isolate FG1118. **T.** *Climacocylis scalaria* isolate FG1116. **U.** *Undella hyalina* isolate CB854. **V.** *U. marsupialis* isolate CBa. **W.** *U. claparadei* isolate CB818. **X.** *Rhabdonella spiralis* isolate CB356. **Y.** *R. spiralis* isolate FG586. **Z.** *R. elegans* isolate FG1392. **AA.** *Cyttarocyclus cassis* isolate FG302. **AB.** *C. acutiformis* isolate FG873. **AC.** *Petalotricha ampulla* isolate FG301. **AD.** *Petalotricha ampulla* isolate CB837. **AE.** *Xystonella longicauda* isolate FG1133. **AF.** *Parundella aculeata* isolate CB355. **AG.** *Codonellopsis orthoceras* isolate FG298. **AH.** *Codonellopsis morchella* isolate FG260. **AI.** *Stenosemella ventricosa* isolate CB875. **AJ.** *Dictyocysta lepida* isolate FG303. **AK.** Unidentified tintinnid isolate FG2102. **AL.** Unidentified tintinnid isolate FG2103. Scale bar 50 μ m. Pictures of additional tintinnid cells from some of these species for which we obtained sequences are available in Supplementary Figure S1.



agglomerated particles (Fig. 1AK and AL). In sharp contrast, the species of *Undella* are typically oceanic and their loricae are smooth (Fig. 1, U-W). The identity of this freshwater taxon needs further research.

As mentioned above, another clade with relatively strong support (BV 89%, PP 0.93) was composed of sequences from the genera *Tintinnopsis* (*T. subacuta* and *T. lohmanii*), *Metacylis*, *Rhabdonella*, *Cyttarocyliis* and *Petalotricha*. Beyond *Tintinnopsis*, the other genera have a long history of re-classification as they have been ascribed to several different families. The sequences of *Metacylis joergenseni*, *Metacylis* sp. MNB99 and *Metacylis angulata* branched in two different sub-clades, whereas the sequences of *Rhabdonella elegans*, *R. spiralis* and *R. hebe* were identical. Similarly, despite the morphological differences between specimens (Fig. 1, AA-AD), the three sequences of *Petalotricha ampulla* were identical or almost identical to the three sequences of *Cyttarocyliis cassis* and *C. acutiformis*. Consequently, we can consider that these six sequences corresponded to a single genus, or even a single species, in contrast with classical taxonomical schemes where the genera *Petalotricha* and *Cyttarocyliis* belong to the two different families Petalotrichidae and Cyttarocylididae (Lynn 2008).

Another clade (BV 100%) was composed of *Parundella* and *Xystonella*, placed together in the family Xystonellidae. The most distal group was composed of a variety of sequences of the genera *Stenosemella*, *Codonellopsis*, *Codonella*, *Codonaria* and *Dictyocysta*. The loricae of these genera are totally or partly agglutinated with mineral or biological particles. The highly supported clustering of these genera (BV 95%) provided a strong evidence for the close relationships between the family Codonellopsidae, Codonellidae *sensu stricto* and Dictyocystidae.

ITS+5.8S-rDNA and Concatenated ITS+5.8S-rDNA+SSU-rDNA Analyses

For the specimens analyzed in this study, the ITS and 5.8S rRNA gene sequences were less conserved than the SSU-rDNA and diverged twice as much (Supplementary Fig. S2), with relatively

high regularity ($r^2 = 0.66$). ITS+5.8S-rDNA phylogenies were quite congruent with the corresponding SSU-rDNA phylogenies, at least for the strongly supported nodes (Supplementary Fig. S3). The only noteworthy exception was *Tintinnidium fluviatile*, which branched at a basal position with some non-tintinnid sequences in the complete SSU-rDNA phylogeny (Fig. 2) but as sister of a *Salpingella*+*Amphorides* clade in the ITS+5.8S-rDNA phylogeny (Supplementary Fig. S3). Given this overall congruence between the SSU-rDNA and the ITS+5.8S-rDNA topologies, a concatenated alignment of the SSU-rDNA, ITS and 5.8S-rDNA sequences was compiled in order to increase the phylogenetic signal to evaluate the phylogenetic relationships among Tintinnida (Fig. 3). According to this concatenated analysis, tintinnids were divided into three clades: the first one only contained the two sequences of *Eutintinnus fraknoi* (BV 100, PP 1); the second one (weakly supported: BV 70%, PP 0.75) grouped *Tintinnidium fluviatile*, *Salpingella acuminata* and *Amphorides quadrilineata*; and the last one (strongly supported: BV 100%, PP 1) corresponded to the large complex of species also observed in the SSU-rDNA phylogeny. This complex was divided into two sister groups. The first sub-clade (BV 67%, PP 0.8) included *Xystonella longicauda*, *Undella claparedei*, *U. hyalina*, *Tintinnopsis lacustris*, *Rhabdonella elegans*, *R. spiralis*, *Cyttarocyliis cassis* and *Petalotricha ampulla*. The second sub-clade (BV 97%, PP 1) grouped, on the one hand, *Codonella aspera*, *Codonaria cistellula* and *Codonaria* sp. CB82 with, on the other hand, *Codonellopsis morchella* and *Stenosemella ventricosa*. Once again, our sequences formed well-supported lineages, but despite of the supplementary amount of nucleotide positions the internal relationships among those lineages remained poorly resolved.

Discussion

Reconciling Tintinnid Molecular Phylogeny and Morphological Taxonomy

Until recently, the phylogenetic robustness of the lorica-based tintinnid taxonomy was difficult to

Figure 2. Maximum likelihood rooted phylogenetic tree of choreotrich SSU-rDNA sequences, based on 1,240 aligned positions. Names in bold represent sequences obtained in this study. Traditional and revised taxonomic groups are shown on the right; light grey background indicates incertae sedis groups. Numbers at nodes are bootstrap values (values under 50% are omitted). Bayesian posterior probabilities higher than 0.90 are indicated by filled circles. Accession numbers are provided between brackets. Grey lines and names indicate the freshwater lineages. The scale bar represents the number of substitutions for a unit branch length.

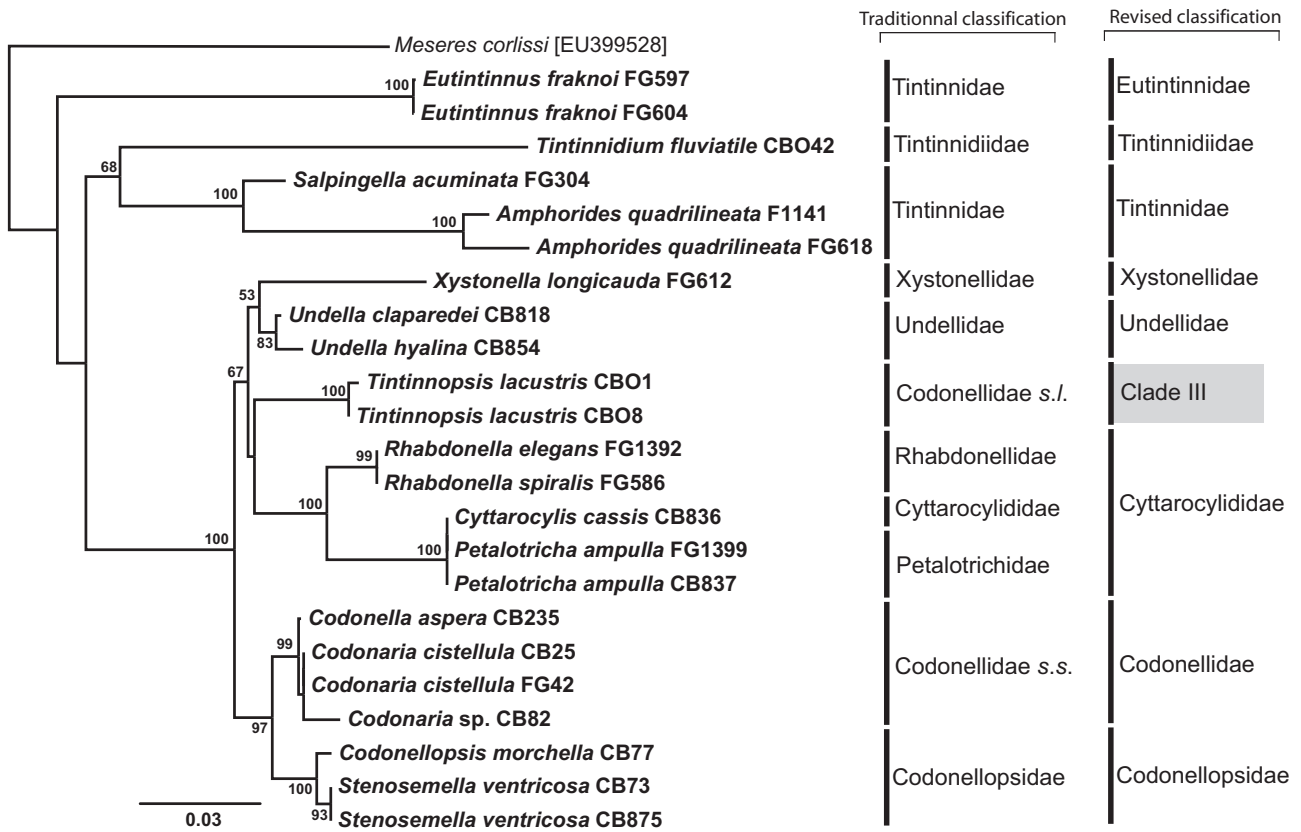


Figure 3. Maximum likelihood rooted phylogenetic tree of tintinnid SSU-rDNA, 5.8S-rDNA and ITS sequences, based on 1,682 aligned positions. Traditional and revised taxonomic groups are shown on the right; light grey background indicates incertae sedis groups. Numbers at nodes are bootstrap values (values under 50% are omitted). Bayesian posterior probabilities higher than 0.90 are indicated by filled circles. Accession numbers are provided between brackets. The scale bar represents the number of substitutions for a unit branch length.

test. This is now possible thanks to the development of molecular phylogeny. With this aim, we assembled a comprehensive data set of tintinnid sequences, including 62 SSU-rDNA and 25 ITS+5.8S-rDNA sequences covering 13 of the 15 families currently defined by morphological classifications (Lynn 2008).

To describe the evolutionary history of tintinnids, one essential issue is to define the position of the root within this ciliate order. Previous studies based on lorica morphology (Kofoed and Campbell 1929, 1939), infraciliature pattern (Laval-Peuto and Brownlee 1986) and SSU-rDNA phylogeny (Duff et al. 2008; Strüder-Kypke and Lynn 2003) suggested that *Tintinnidium* spp. were located at the base of the tintinnid tree. To investigate the most basal lineage of the tintinnid tree, we used a variety of other choreotrichs and oligotrichs as outgroup taxa. However, our results showed little confidence in the exact position of the tintinnid root in both the ML and Bayesian trees (Fig. 2), and also in our

concatenated rDNA analysis (Fig. 3). To check whether this uncertainty was due to the use of a too narrow set of outgroup sequences, we carried out phylogenetic analyses including a diversity of additional outgroup sequences from different groups within the large ciliate class Spirotrichea (Supplementary Figs S4 and S5). As in our first analysis, these new phylogenies failed to place *Tintinnidium* spp. robustly. Nevertheless, they showed a different position for the Eutintinnidae clade, which was close to the base of Tintinnida in the initial tree but emerged either as sister of the *Salpingella-Amphorellopsis-Amphorides-Steenstrupiella* group (BV < 50%) or at the next branch after this group (BV 64%). At any rate, statistical support for the basal part of the tintinnid tree remained weak. Therefore, it appeared clear that supplementary sequence information, in particular sequences from other conserved markers, is needed to assess the phylogenetic position of *Tintinnidium* spp. Likewise,

further work will be needed to robustly infer the phylogenetic position of the families Strombilidiidae and Parastrombidinopsidae, which tend to group with the Tintinnidiidae rather than with the Strombidinopsidae.

Previous studies employing sequence data reported the paraphyly of the genus *Tintinnopsis* (Li et al. 2009; McManus and Katz 2009; Snoeyenbos-West et al. 2002). Our results with a richer taxonomic sampling support this view. It is often difficult to observe details of the loricae of *Tintinnopsis* because the agglomerated particles often conceal diagnostic characteristics of the lorica. This likely explains that under the genus *Tintinnopsis* have traditionally been described species that, according to molecular phylogeny analyses, should belong to different genera. Moreover, annotation errors and possible misidentifications may lead to incorrect interpretations. For example, the GenBank sequence DQ487200 annotated as *Tintinnopsis* sp. seems to correspond to a *Tintinnidium* species in the published article (Duff et al. 2008). Similarly, in the case of *Tintinnopsis fimbriata*, an unambiguous identification seems not possible from the picture of the specimen shown (Strüder-Kypke and Lynn 2003). Not surprisingly, our molecular analysis reveals that the phylogeny of the genus *Tintinnopsis* is even more complex than previously thought. For example, we have found a freshwater genus that shares the external appearance of *Tintinnopsis*, with an entirely agglomerated lorica but that clearly branched with the oceanic genus *Undella*, with a smooth lorica. *Climacocylis scalaria*, with a distinctive hyaline lorica, represented a similar case since it branched very close to *T. cylindrica* and *T. radix* within one of the *Tintinnopsis* clades. Moreover a potential monophyletic grouping of the majority of *Tintinnopsis* sequences from GenBank and those added in this study cannot be excluded, as attested by our AU tests (see above). Taking all this into account, it seems clear that the taxonomic value of the “agglomeration of particles” or “agglutinated lorica” as a diagnostic character at the family level is not supported by the molecular data. With no doubt, the current classification of the genus *Tintinnopsis* needs a complete revision in order to establish a taxonomic scheme compatible with molecular phylogeny.

In contrast with the poorly informative *Tintinnopsis* lorica, other species that have distinctive and easily visible lorica morphologies are ideal to investigate the phylogenetic value of the lorica characters. Our results reveal that the species of the genus *Eutintinnus* form a cohesive monophyletic group in a basal position in the tintinnid phylogeny

(Fig. 2). Compared to other tintinnids, the lorica of *Eutintinnus* is relatively simple, just a smooth tube lacking ornamentations (horns, annuli, collar, spines...) except for the toothed oral rim of *E. pectinis*. In most *Eutintinnus* species the diameter of the oral and aboral aperture are quite similar while in other tintinnids the aboral end is closed or, if open, has a very small diameter compared to the oral diameter. This morphological character together with the support of the molecular data argues in favor of the erection of a new family for *Eutintinnus* (see Taxonomic Appendix).

An additional problem concerns the consistency of lorica morphology. For example, studies of the life cycle of some *Favella* species have revealed important changes in lorica morphology during cell development, with individuals often exhibiting morphologies previously considered to belong to separate genera, such as *Coxiella* (Laval-Peuto 1981). These observations are also supported by molecular data. In fact, Kim et al. (2010) used cultures and observed lorica polymorphism for specimens of *Favella ehrenbergi* that shared identical SSU-rDNA sequences. Similarly, our Mediterranean specimens of *Favella azorica*, without an aboral pedicel, showed identical SSU-rDNA and, moreover, these sequences were identical to those of *F. campanula*, with an aboral pedicel, isolated from China (Gao et al. 2009). The phylogenetic position of the *F. taraikaensis* Hada sequence obtained by Li et al. (2009), which unexpectedly branched with *Metacyclis* sp., is also questionable. Additional sequences of this taxon are required to ascertain the monophyly of the genus *Favella*. At present, *Favella* is the only representative of the Ptychocylididae for which sequences are available so that the monophyletic character of this family also remains to be tested.

The studies of the life cycle of tintinnids are nearly restricted to common coastal species that are maintained in temporary cultures. As a consequence, very little is known about the life cycle and changes in the lorica morphology for rare and oceanic species. Single-cell molecular analyses offer an opportunity to study these species. Thus, our work using this approach has revealed conflicting cases comparable to those cited for *Favella*. For example, in the case of the genera *Petalotricha* and *Cyttarocylis*, our sequences of *P. ampulla* (Fig. 1, AC-AD) branched in the same clade as those of *Cyttarocylis acutiformis* (Fig. 1, AB) and *C. cassis* (Fig. 1, AA). These six *Petalotricha* and *Cyttarocylis* specimens collected at different dates and places showed identical or almost identical SSU-rDNA sequences, and

even the more rapidly evolving ITS and 5.8S-rDNA sequences were identical. Such genetic identity of the two distinct forms, one with the lorica wall finely sculptured with polygonal reticuli (*Cyttarocylis* morphotype) and the other with a smooth lorica with irregular small fenestrae in the subnuchal area (*Petalotricha* morphotype), suggests that they may just represent different stages of lorica development in a single genus or even in a single species. This is unexpected if we consider that in all previous taxonomical schemes, *Petalotricha* and *Cyttarocylis* have been classified into two separate families based on the conspicuous differences in the lorica structure (Fol 1881; Kent 1882; Kofoid and Campbell 1929; Marshall 1969). *Cyttarocylis* Fol, 1881 has the priority over *Petalotricha* Kent, 1882 and, consequently, we propose a new combination to place *Petalotricha ampulla* under the genus *Cyttarocylis* (see Taxonomic Appendix).

For the genus *Undella*, numerous species have been described based on minute variations of the lorica (Daday 1887; Kofoid and Campbell 1929). Our phylogenetic analysis shows a very close relationship between the clearly distinct morphotypes of *Undella hyalina*, *U. marsupialis* and *U. claparedei*. It also supports that the genus *Proplectella* Kofoid & Campbell, 1929, which grouped the small and round-shaped Undellidae with an inner collar (*U. claparedei* morphotype; Fig. 1W), could be considered a junior synonym, as already supposed by Balech (1975).

Simple morphological characters, such as the lorica length, seem to have low taxonomic value in the case of genera such as *Rhabdonella*. Our *Rhabdonella* spp. specimens showed lorica with lengths between 100 μm and 400 μm (Fig. 1, X-Z). Although described as different species, our specimens of *R. spiralis* and *R. elegans* shared identical SSU-rDNA, ITS and 5.8S-rDNA sequences. In the case of *Codonaria cistelulla*, our specimens showed different degrees of collar agglutination (Fig. 1, O-P) but, again, their SSU-rDNA sequences were also identical.

The monograph of Kofoid and Campbell (1929) catalogued 697 species, raising to species status many forms previously described as varieties, often based on slight lorica differences. In this study, we have found cases of little or none genetic difference between species with distinct loricae belonging to the genera *Undella*, *Rhabdonella* or *Cyttarocylis*. This suggests that the number of species have most likely been excessively inflated. On the other hand, *Tintinnopsis* may require a split into several genera, even families. The species *Amphorides quadrilineata* is an exception. Our sequences branched in

two distinct clades that appear to hide a cryptic speciation.

Marine-Freshwater Transitions

As observed in many eukaryotic lineages (Logares et al. 2009), there are genera composed of marine and freshwater species. In those cases, the phylogenies do not support a deep ancient divergence between members of marine and freshwater habitats but a number of more or less recent independent colonizations. If our tintinnid phylogenetic tree topology is correct, at least three transitions between marine and freshwater/brackish environments appear to have occurred during the evolutionary history of these ciliates. Both parsimony and maximum likelihood character optimization suggest that the ancestor of tintinnids *sensu stricto* was a marine species, as it is the case for most other choreotrich ciliates, and, consequently, that the three colonizations inferred occurred from marine ancestors to freshwater environments (Supplementary Fig. S6). Until now, few tintinnid species living in freshwaters have been described. Our results suggest that freshwater colonizations are not restricted to families characteristic of coastal waters but also include at least one representative of typical open water forms (Undellidae). This phenomenon appears therefore to be phylogenetically diverse within the Tintinnida.

The generic affiliation of the freshwater species *Codonella cratera* is probably incorrect, as inferred by cytological and ciliature features (Agatha 2010a, b; Agatha and Strüder-Kypke 2007; Laval-Peuto and Brownlee 1986). In our SSU-rDNA tree, a freshwater lineage strongly grouping *Codonella cratera* (Duff et al. 2008) with *Tintinnopsis lacustris* branches as sister clade of marine *Tintinnopsis* spp. This confirms that *C. cratera* should be removed from the genus *Codonella* to become a *Tintinnopsis* species.

Classification into Families

Our results revealed that the genera traditionally ascribed to the family Tintinnidae formed at least three distantly related clades, suggesting that they may define different families. The clearest example is *Eutintinnus* spp. that forms a separate cohesive monophyletic group (BV 100%, PP 1) distant to other members of the Tintinnidae (Fig. 2). From a morphological point of view, *Eutintinnus* differs from the other Tintinnidae in the wide aboral opening and the lack of longitudinal aboral fins. By contrast, the very rarely reported type genus *Tintinnus* is characterized by a homogeneous wall with a closed

aboral end (Kofoid and Campbell 1939). Therefore, both the morphological and phylogenetic molecular data converge to support the exclusion of *Eutintinnus* from the family Tintinnidae. For this reason, we propose in the taxonomical appendix the erection of Eutintinnidae fam. nov. to host the members of the genus *Eutintinnus*. This new family is characterized by a hyaline lorica, cylindrical or nearly so, clearly opened at the both ends. Some species of *Eutintinnus* are characterized by the association with diatoms (*Chaetoceros*, *Hemiaulax*) (Gómez 2007b).

While *Eutintinnus*, with simple lorica morphology, tends to branch in a basal position in the tintinnid phylogeny, the genus *Dictyocysta*, with a peculiar reticulate lorica, branches in the most distal position (Fig. 2). In the SSU-rDNA phylogeny, our sequences of *D. lepida* and that of *D. reticulata* (Strüder-Kypke and Lynn 2008) branch with high support (BV 90%, PP 1) with species of the genera *Codonaria* and *Codonella*, type of family Codonellidae. In the classical taxonomic schemes, *Dictyocysta* has been placed in the family Dictyocystidae. However, the very close phylogenetic relationship between *Dictyocysta* and members of the Codonellidae suggests the placement of *Dictyocysta* in this family. This is supported by cytological characteristics shared by these genera: the presence of a lorica sac and closing apparatus (see review in Agatha 2010b).

Tintinnopsis has been classified in the family Codonellidae. However, the *Tintinnopsis* sequences form, at least, two separate groups and are never related to Codonellidae sensu stricto (Fig. 2). The group of *Tintinnopsis subacuta* and *T. lohmanni* appeared in a clade (BV 89%) that also contained the genera *Metacylis*, *Rhabdonella* and *Cyttarocyliis*, which belong to three separate families. This strongly suggests that *T. subacuta* and *T. lohmanni* do not belong to the genus *Tintinnopsis*. Detailed morphological studies are required for these species before proposing the transfer into another existing genus or the erection of a new genus. Concerning the entire clade to which *T. subacuta* and *T. lohmanni* belong, it also groups the families Metacyliidae, Rhabdonellidae and Cyttarocylididae. However, their branching pattern does not support any particular relationship among these families, indicating that the three families could be merged into a single one. In so doing, the family Cyttarocylididae would have the priority to name the entire family.

Most *Tintinnopsis* species, including the type species *T. beroidea*, formed a paraphyletic group with several distinct lineages. Unexpectedly,

Climacocyliis (family Metacyliidae) appears closely related to several of the *Tintinnopsis* species, far from *Metacylis*, the type of the family. Consequently, *Climacocyliis* should be excluded from the Metacyliidae. However, we have been unable to find common morphological characters that could support the placement of *Tintinnopsis* and *Climacocyliis* under a single family or the tentative transfer of *Tintinnopsis* species such as *T. radix* or *T. cylindrica* into the genus *Climacocyliis*. Given the observed complexity, the assignment of *Tintinnopsis* and *Climacocyliis* to any existing family requires further research.

Methods

Sampling and cell isolation: Our primary sampling site was the Bay of Villefranche-sur-Mer in the NW Mediterranean Sea. Additional samples were collected from 2 other coastal sites in the NW Mediterranean, Marseille and Banyuls-sur-Mer, and 2 sampling stations in the open Mediterranean, one near the center of the Western basin and another in the Central basin. We also collected samples in freshwater lakes (see below). Sampling in the Bay of Villefranche sur Mer, Ligurian Sea, was performed at a long-term monitoring site called 'Point B' (43°41'10"N, 7°19'00"E; water column depth ~80 m). Sampling in double oblique angle was done with a custom-made conical phytoplankton net (53 µm mesh size, 54 cm diameter and 280 cm length). Additional samples were collected at discrete depths by using 12-L Niskin bottles, and concentrated by screening water through 20 µm Nylon mesh (Nitex Co.). The material was examined in Utermöhl chambers and individual tintinnids were isolated and photographed using an inverted microscope (Olympus IX51) equipped with an Olympus DP71 digital camera. Based on microscopic observations, species designations were made employing characteristics of lorica morphology. The main taxonomic works consulted included Jörgensen (1924), Kofoid and Campbell (1929, 1939) and Balech (1959). We followed the classification into families and genera proposed by Lynn (2008). The identified and photographed specimens were micropipetted individually with a fine capillary into a clean chamber and washed several times in serial drops of 0.2-µm filtered and sterilized seawater. The isolated washed cells were placed in 0.2 ml Eppendorf tubes filled with several drops of absolute ethanol. The tubes were stored at 4 °C or at room temperature and in darkness until the molecular analysis could be performed. Other specimens were collected from the end of the pier (depth 3 m) of the Station Marine d'Endoume, Marseille (43°16'48"N, 5°20'57"E). A strainer with netting of 20 µm mesh-size was used to collect the organisms, with a filtered volume ranging between 10 and 100 liters according to the concentration of particles. The concentrated material was processed in a similar manner as described above, using a Nikon Eclipse TE200 inverted microscope equipped with a Nikon Coolpix E995 camera. Tintinnids were also collected from surface waters of the harbour of Banyuls-sur-Mer (42°28'50"N, 3°08'09"E) following the same procedure. The specimens were observed with an Olympus inverted microscope (Olympus IX51) and photographed with an Olympus DP71 digital camera. Specimens from open waters of the Mediterranean specimens were isolated from samples collected during the BOUM (Biogeochemistry from the

Oligotrophic to the Ultra-oligotrophic Mediterranean) cruise in the Mediterranean Sea in June-July 2008. Material was from surface water sampling at the permanent stations of the Western Basin, 'Station A' (39°6'N, 5°30'E), and the Central Ionian Sea, 'Station B' (34°8'N, 18°45'E) (see Christaki et al. 2011 for details). Ten litres were collected from the surface with a bucket and filtered by using a strainer of 20- μ m netting aperture. The retained material was fixed with absolute ethanol to a final concentration of 50% concentrated seawater sample and 50% ethanol. The fixed sample was subsequently examined and processed as described above. Tintinnids were also collected from freshwater lakes in the Chevreuse Valley (Étang du Perray, 48°41'49"N, 1°51'37"E; Étang de Pourras, 48°42'52"N, 1°50'39"E; Étang des Vallées 48°41'20"N, 1°54'59"E), near Paris, France, and from La Albufera (39°21'32"N, 1°38'22"W), a large shallow lake near Valencia, Spain. The specimens were observed, isolated and processed as described above.

PCR amplification of small subunit rRNA genes, internal transcribed spacers ITS1 and ITS2, and 5.8S rRNA genes: The specimens fixed in ethanol were centrifuged gently for 5 min at 3,000 rpm. Ethanol was then evaporated in a vacuum desiccator and single cells were resuspended directly in 25 μ l of Ex TaKaRa buffer (TaKaRa, distributed by Lonza Cia., Levallois-Perret, France). PCR reactions were done in a volume of 30 μ l reaction mix containing 10-20 pmol of the forward eukaryotic-specific SSU-rDNA primers EK-42F (5'-CTCAARGAYTAAGCCATGCA-3') and the reverse eukaryotic large subunit rDNA EK-28S564R (5'-TGGTCCGTGTTTCRAGACG-3') (López-García et al. 2001). The PCR reactions were performed under the following conditions: 2 min denaturation at 94 °C; 10 cycles of 'touch-down' PCR (denaturation at 94 °C for 15 s; a 30 s annealing step at decreasing temperature from 65 down to 55 °C employing a 1 °C decrease with each cycle, extension at 72 °C for 2 min); 20 additional cycles at 55 °C annealing temperature; and a final elongation step of 7 min at 72 °C. Internal primers were designed for this study to specifically target Tintinnida ribosomal genes and to avoid any possible PCR contamination by other eukaryotes (e.g. ingested preys or parasites). A nested PCR reaction was then carried out using 3 μ l of the first PCR reaction diluted by 1/10 in a GoTaq (Promega, Lyon, France) polymerase reaction mix containing the tintinnid-specific primers 18S-Tin3F (5'-GCGGTATTTATTAGATAWCAGCC-3') and 28S-TinR1 (5'-TGGTGCACCTAGTATCAAAGT-3') and similar PCR conditions as described above. At this step, if amplicon bands were not visible on agarose gels, a third, semi-nested PCR was carried out using the eukaryotic-specific SSU-rDNA primer EK-1498R (5'-TAACAATACAGGGCATCCAT-3') and keeping the forward primer TIN3F. Negative controls without template DNA were used at all amplification steps. Amplicons of the expected size (~2,000 bp after the nested PCR and ~1,500 bp after the semi-nested one) were then entirely sequenced using the primers 18S-Tin3F, EK-1498R and 28S-TinR1 with an automated 96-capillary sequencer ABI PRISM 3730xl (Beckman Coulter Genomics, Takeley, U.K.). Sequences were deposited in GenBank with accession numbers JQ408154-JQ408215 (see also Supplementary Table S1).

Phylogenetic analyses: We carried out a multiple alignment with our sequences and publicly available complete or nearly complete (>1,400 bp) SSU-rDNA sequences of spirotrich ciliates using the profile alignment option of MAFFT (Katoh et al. 2002) and MUSCLE 3.6 (Edgar 2004). We kept in our data set the 37 tintinnid sequences from specimens that were associated with micrographs in the references given in GenBank to minimize the inclusion of mis-identified species. As outgroup, we used 17 sequences belonging to hypotrich,

stichotrich, oligotrich (Strombidiidae) and choreotrich ciliates (Strobilidiidae and Strombidinopsidae). Partial sequences too short to be included in our data set were eliminated (e.g. *Eutintinnus* sp., EU399533, 1309 bp) to keep the maximum of sequence information. Pairwise sequence comparisons (percentage of sequence identity) were calculated with the package ClustalX (Larkin et al. 2007). The resulting alignment was manually inspected using the program ED of the MUST package (Philippe 1993). Ambiguously aligned regions and gaps were excluded from phylogenetic analyses. A second sequence data set was constructed through the same procedure that included the sequences of tintinnids for which we succeeded to amplify the SSU-rDNA, the ITS1 and ITS2 and the 5.8S rRNA genes. Conserved alignment sites were trimmed using GBLOCKS (Castresana 2000) or BMGE (Crisuolo and Gribaldo 2010) with default parameters. Then, we carried out Maximum Likelihood (ML) phylogenetic analyses with the program TREEFINDER (Jobb et al. 2004) applying a GTR + Γ + I model of nucleotide substitution with a Γ -shaped distribution of substitution rates with four rate categories. This model was selected using the model selection tool implemented in TREEFINDER. Bootstrap values were calculated using 1,000 pseudoreplicates with the same substitution model. The Bayesian Inference (BI) analyses were carried out with the program MrBayes (Huelsenbeck and Ronquist 2001), with two independent runs and 1,000,000 generations per run. After checking convergence (maximum difference between all bipartitions <0.01) and eliminating the first 3,500 trees (burn-in), a consensus tree was constructed sampling every 100 trees. Comparison of different tree topologies was carried out by applying the Approximately Unbiased (AU) test (Shimodaira 2002) implemented in TREEFINDER. Finally, the most parsimonious pattern of marine-freshwater transitions was inferred for the set of tintinnid species using MESQUITE 2.75 (Maddison and Maddison 2011).

Acknowledgements

We acknowledge financial support from the French CNRS and the ANR Biodiversity program (ANR BDIV 07 004-02 'Aquaparadox'). F.G. is supported by the contract JCI-2010-08492 of the Ministerio Español de Ciencia y Tecnología.

Appendix A. Taxonomic appendix

Based on the information obtained from molecular phylogenetic analyses, we propose the following classification of the tintinnid genera for which molecular data are available.

- Subclass Choreotrichia Small & Lynn, 1985.
- Order Tintinnida Kofoid & Campbell, 1929.

Family Tintinnidiidae:

Tintinnidium

Family Eutintinnidae fam. nov. Bachy et al.:

Eutintinnus

Family Tintinnidae:

Salpingella, *Amphorellopsis*, *Amphorides*,

Steenstrupiella

Family Ptychocylididae

Favella

Family Undellidae:

Undella

Family Cyttarocylididae:

Metacylis, *Rhabdonella*, *Cyttarocylis*
(=*Petalotricha*)

Family Xystellonellidae:

Xystonella, *Parundella*

Family Codonellopsidae:

Codonellopsis, *Stenosemella*

Family Codonellidae (including Dictyocystidae):

Codonella, *Codonaria*, *Dictyocysta*

Incertae sedis:

Tintinnopsis sensu stricto (type species),
Climacocylis, *Tintinnopsis* clades I/II/III/IV/V

New Taxonomical Proposals

Combinatio nova:

Cyttarocylis ampulla (Kent) C. Bachy, J.R. Dolan and P. López-García, comb. nov.

Basionym: *Petalotricha ampulla* Kent 1882. A manual of the infusoria: including a description of all known flagellate, ciliate, and tentaculiferous Protozoa, British and foreign, and an account of the organization and affinities of the sponges. Vol. II. D. Bogue, London. p. 627, figs 1-2.

Synonyms: *Tintinnus ampulla* Fol (1881), *Petalotricha major* Jörgensen (1924), *Petalotricha serrata* Kofoid & Campbell (1929).

Familia nova:

- Subclass Choreotrichia Small & Lynn, 1985.
- Order Tintinnida Kofoid & Campbell, 1929.
- Family Eutintinnidae fam. nov. (=Tintinnidae, pro parte).
- Diagnosis: The family is characterized by a hyaline and smooth lorica, cylindrical or nearly so, clearly opened at the both ends. The family differs from Tintinnidae in the wide aboral opening and the lack of longitudinal aboral fins. Free-living, marine. Some species are associated with an ectosymbiotic diatom.
- Type genus: *Eutintinnus* Kofoid & Campbell, 1939.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.protis.2012.01.001.

References

- Agatha S** (2010a) Redescription of *Tintinnopsis parvula* Jörgensen, 1912 (Ciliophora: Spirotrichea: Tintinnina), including a novel lorica matrix. *Acta Protozool* **49**:213–234
- Agatha S** (2010b) A light and scanning electron microscopic study of the closing apparatus in tintinnid ciliates (Ciliophora, Spirotricha, Tintinnina): A forgotten synapomorphy. *J Eukaryot Microbiol* **57**:297–307
- Agatha S, Strüder-Kypke MC** (2007) Phylogeny of the order Choreotrichida (Ciliophora, Spirotricha, Oligotricha) as inferred from morphology, ultrastructure, ontogenesis, and SSrRNA gene sequences. *Eur J Protistol* **43**:37–63
- Balech E** (1959) Tintinninea del Mediterraneo. *Trabajos del Instituto Español de Oceanografía* **28**:1–88
- Balech E** (1975) La familia Undellidae (Protozoa, Ciliophora Tintinnina). *Physis Sec A* **34**:377–398
- Bergsten J** (2005) A review of long-branch attraction. *Cladistics* **21**:163–193
- Boltovskoy D, Dinofrio EO, Alder VA** (1990) Intraspecific variability in Antarctic tintinnids: the *Cymatocylis affinis/convallaria* species group. *J Plankton Res* **12**:403–413
- Calbet A, Landry MR** (2004) Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnol Oceanogr* **49**:51–57
- Castresana J** (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**:540–552
- Choi JK, Coats DW, Brownlee DC, Small EB** (1992) Morphology and infraciliature of three species of *Eutintinnus* (Ciliophora: Tintinnina) with guidelines for interpreting protargol-stained tintinnine ciliates. *J Eukaryot Microbiol* **39**:80–92
- Christaki U, Van Wambeke F, Lefevre D, Lagaria A, Prieur L, Pujó-Pay M, Grattepanche JD, Colombet J, Psarra S, Dolan JR, Sime-Ngando T, Conan P, Weinbauer MG, Moutin T** (2011) Microbial food webs and metabolic state across oligotrophic waters of the Mediterranean Sea during summer. *Biogeosciences* **8**:1839–1852
- Criscuolo A, Gribaldo S** (2010) BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**:210
- Daday Ev** (1887) Monographie der Familie der Tintinnodeen. *Mitteilungen aus der Zool. Station zu Neapel* **7**:473–591
- Davis CC** (1981) Variations of lorica shape in the genus *Ptychocylis* (Protozoa: Tintinnina) in relation to species identification. *J Plankton Res* **3**:433–443
- Dolan JR, Jacquet S, Torrétón J-P** (2006) Comparing taxonomic and morphological biodiversity of tintinnids (planktonic ciliates) of New Caledonia. *Limnol Oceanogr* **51**:950–958
- Dolan JR, Claustre H, Carlotti F, Plounevez S, Moutin T** (2002) Microzooplankton diversity: relationships of tintinnid ciliates with resources, competitors and predators from the Atlantic Coast of Morocco to the Eastern Mediterranean. *Deep-Sea Res part I* **49**:1217–1232

- Duff RJ, Ball H, Lavrentyev PJ** (2008) Application of combined morphological-molecular approaches to the identification of planktonic protists from environmental samples. *J Eukaryot Microbiol* **55**:306–312
- Edgar RC** (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **19**:113
- Foissner W, Wilbert N** (1979) Morphologie, Infraciliatur und Ökologie der limnischen Tintinnina: *Tintinnidium fluviatile* Stein, *Tintinnidium pusillum* Entz, *Tintinnopsis cylindrata* Daday und *Codonella cratera* (Leidy) (Ciliophora, Polyhymenophora). *J Protozool* **26**:90–103
- FoH** (1881) Contribution to the knowledge of the family Tintinnodea. *Annals Mag Nat Hist* **12**:73–88
- Gao S, Gong J, Lynn D, Lin X, Song W** (2009) An updated phylogeny of oligotrich and choreotrich ciliates (Protozoa, Ciliophora, Spirotrichea) with representative taxa collected from Chinese coastal waters. *Syst Biodivers* **7**:235–242
- Gold K, Morales EA** (1975) Seasonal changes in lorica sizes and the species of Tintinnida in the New York Bight. *J Eukaryot Microbiol* **22**:520–528
- Gómez F** (2007a) Trends on the distribution of ciliates in the open Pacific Ocean. *Acta Oecol* **32**:188–202
- Gómez F** (2007b) On the consortium of the tintinnid *Eutintinnus* and the diatom *Chaetoceros* in the Pacific Ocean. *Mar Biol* **151**:1899–1906
- Huelsenbeck JP, Ronquist F** (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755
- Jobb G, von Haeseler A, Strimmer K** (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* **4**:18
- Jørgensen** (1924) Mediterranean Tintinnids. Report on the Danish Oceanographical Expeditions 1908–10 to the Mediterranean and Adjacent Seas, Vol 2 (Biology), J.3., 1–110
- Katoh K, Misawa K, Kuma K, Miyata T** (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066
- Kent WS** (1882) A Manual of the Infusoria, Including a Description of All Known Flagellate, Ciliate, and Tentaculiferous Protozoa, British and Foreign and an Account of the Organization and Affinities of the Sponges. David Bogue, London
- Kim SY, Yang EJ, Gong J, Choi JK** (2010) Redescription of *Flavella ehrenbergii* (Claparède and Lachmann, 1858) Jørgensen, 1924 (Ciliophora: Choreotrichia), with phylogenetic analyses based on small subunit rRNA gene sequences. *J Eukaryot Microbiol* **57**:460–467
- Kofoed CA, Campbell AS** (1929) A Conspectus of the Marine and Freshwater Ciliata Belonging to the Suborder Tintinnoinea, with Descriptions of New Species Principally from the Agassiz Expedition to the Eastern Tropical Pacific 1904–1905. *Univ Calif Publ Zool*
- Kofoed CA, Campbell AS** (1939) The Ciliata: the Tintinnoinea. *Bull Mus Comp Zool Harvard*
- Laackmann H** (1908) Ungeschlechtliche und geschlechtliche Fortpflanzung der Tintinnen. *Wissensch Meeresunters, Abteilung Kiel* **10**:13–38+Plates 1–3
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG** (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948
- Laval-Peuto M** (1977) Reconstruction d'une lorica de forme *Coxiella* par le trophonte nu de *Favella ehrenbergii* (Ciliata, Tintinnina). *C R Acad Sci* **284**:547–550
- Laval-Peuto M** (1981) Construction of the lorica in Ciliata Tintinnina. In vivo study of *Favella ehrenbergii*: Variability of the phenotypes during the cycle, biology, statistics, biometry. *Protistologica* **17**:249–272
- Laval-Peuto M** (1983) Sexual reproduction in *Favella ehrenbergii* (Ciliophora, Tintinnina) taxonomical implications. *Protistologica* **19**:503–512
- Laval-Peuto M, Brownlee DC** (1986) Identification and systematics of the Tintinnina (Ciliophora): evaluation and suggestions for improvement. Institut océanographique, Paris, France
- Li Z, Yi Z, Yang J, Gong J, Clamp JC, Al-Rasheid KAS, Al-Arifi S, Al-Khedhairi AA, Song W** (2009) Phylogenetic investigation on five genera of tintinnid ciliates (Ciliophora, Choreotrichia), based on the small subunit ribosomal RNA gene sequences. *Progr Nat Sci* **19**:1097–1101
- Logares R, Brate J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K** (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* **17**:414–422
- López-García P, Rodríguez-Valera F, Pedros-Alio C, Moreira D** (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**:603–607
- Lynn DH** (2008) The Ciliated Protozoa. Characterization, Classification, and Guide to the Literature. 3rd ed Springer, Dordrecht
- Maddison W, Maddison DR** (2011) Mesquite: a modular system for evolutionary analysis. Version 2.75. <http://mesquiteproject.org>
- Marshall SM** (1969) Protozoa, order Tintinnia. Conseil International pour l'exploration de la Mer, Fiches d'Identification de Zooplancton:fiches 117–127
- McManus GB, Katz LA** (2009) Molecular and morphological methods for identifying plankton: what makes a successful marriage? *J Plankton Res* **31**:1119–1129
- Montagnes DJS, Lynn DH** (1987) A quantitative Protargol stain (QPS) for ciliates: Method description and test of its quantitative nature. *Mar Microb Food Webs* **2**:83–93
- Philippe H** (1993) MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* **21**:5264–5272
- Pierce RW, Turner E** (1992) Ecology of planktonic ciliates in marine food webs. *Rev Aquat Sci* **6**:139–181
- Prescott DM** (1994) The DNA of ciliated protozoa. *Microbiol Mol Biol Rev* **58**:233–267
- Santoferrara LF, Alder VA** (2009) Morphological variability, spatial distribution and abundance of *Helicostomella* species (Ciliophora: Tintinnina) in relation to environmental factors (Argentine shelf; 40–55 S). *Sci Mar* **73**:701–716

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* **51**:492–508

Small EB, Lynn DH (1985) Phylum Ciliophora Doflein, 1901. In Lee JJ, Hutner SH, Bovee EC (eds) *An Illustrated Guide to the Protozoa*. Society of Protozoologists. Allen Press, Lawrence, Kansas, pp 393–575

Snoeyenbos-West OLO, Salcedo T, McManus GB, Katz LA (2002) Insights into the diversity of choreotrich and oligotrich ciliates (Class: Spirotrichea) based on genealogical analyses of multiple loci. *Int J Syst Evol Microbiol* **52**: 1901–1913

Strüder-Kypke MC, Lynn DH (2003) Sequence analyses of the small subunit rRNA gene confirm the paraphyly of oligotrich ciliates sensu lato and support the monophyly of the subclasses

Oligotrichia and Choreotrichia (Ciliophora, Spirotrichea). *J Zool* **260**:87–97

Strüder-Kypke MC, Lynn DH (2008) Morphological versus molecular data - Phylogeny of tintinnid ciliates (Ciliophora, Choreotrichia) inferred from small subunit rRNA gene sequences. *Denisia* **23**:417–424

Wasik A, Mikołajczyk E (1994) Annual cycle of tintinnids in Admiralty Bay with an emphasis on seasonal variability in *Cymatocylis affinis/convallaria* lorica morphology. *J Plankt Res* **16**:1–8

Williams R, McCall H, Pierce RW, Turner JT (1994) Automatic categorisation of five species of *Cymatocylis* (Protozoa, Tintinnida) by artificial neural network. *Mar Ecol Prog Ser* **107**:263–272

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Molecular phylogeny of tintinnid ciliates (Tintinnida, Ciliophora)

*Charles Bachy, Fernando Gómez, Purificación López-García, John R. Dolan, David
Moreira*

SUPPLEMENTARY MATERIALS

Detailed description of the tintinnid specimens isolated in this work, classified into families according to Lynn (2008).

Family Tintinnidiidae

This family was represented in GenBank exclusively by two sequences of the genus *Tintinnidium* Kent. The lorica of *Tintinnidium* is soft and gelatinous, tends to be irregular but generally tubular. We isolated an elongated specimen of *Tintinnidium fluviatile* Stein with an aboral end closed and flattened (Fig. 1, A) from a freshwater lake. The lorica was 75 µm long and the oral aperture was 30 µm in diameter.

Family Tintinnidae

Genus *Eutintinnus* Kofoid & Campbell (Fig. 1, B-D)

This genus is characterized by a lorica in the form of a truncated cone or cylinder open at both ends; the wall is smooth and hyaline. The sequences of *Eutintinnus fraknoi* Daday, *E. pectinis* Kofoid & Campbell, and *Eutintinnus* sp. were available in GenBank. We determined sequences for three species of *Eutintinnus*. *Eutintinnus tubulosus* Ostenfeld (Fig. 1, B) sequences were retrieved from two conjugating specimens. The cells were 140 µm long. The oral and aboral apertures were 40 µm and 25 µm in diameter, respectively. We obtained two sequences of two specimens of *E. fraknoi*, a species with loricae characterized by the slight flaring of both ends (Fig. 1, C). Both specimens had large loricae (300±40 µm long) with oral and aboral apertures of 60 µm and 30 µm, respectively. *Eutintinnus apertus* Kofoid & Campbell is a relatively small species (60 µm long) with an oral aperture diameter of 25 µm, and an aboral region contracting abruptly (10 µm diameter). A colony of three cells of the diatom *Chaetoceros dadayi* Pavillard was attached to the outer part of the lorica in our specimen (Fig. 1, D).

Genera *Salpingella* Jörgensen, *Amphorides* and *Steenstrupiella* Kofoid & Campbell (Figs. 1, E-H)

These genera are characterized by an elongated and hyaline lorica of a nail- to vase-like shape and flaring collar and an aboral end slightly open or closed. Sequences attributed to *Salpingella acuminata* (Claparède & Lachmann) Jörgensen and *Amphorellopsis acuta* Schmidt were the only representatives in GenBank identified at species level. We determined a SSU-rDNA sequence of *S. acuminata*, sequences of the types of the genera *Amphorides* and

Steenstrupiella, with 5 sequences of *A. quadrilineata* Claparède & Lachmann Strand and 2 sequences of *Steenstrupiella steenstrupi* (Claparède & Lachmann) Kofoid & Campbell.

The lorica of *Salpingella acuminata* collected in harbor of Banyuls was nail-shaped and quite long (350 μm long, 38 μm oral aperture) and showed vertical fins at the aboral end (Fig. 1, E). Our observations did not permit verification that the small aboral end was open, as commonly reported within this genus.

Amphorides quadrilineata is one of the most ubiquitous tintinnid species in the French Mediterranean coasts. We obtained five sequences from five specimens collected in several places (Table 1). The loricae were vase-shaped, with flaring collars and truncated aboral ends. The dimensions of the five cells were identical, 115 μm long and with oral aperture of 45 μm (Fig. 1, F-G).

Steenstrupiella steenstrupi were isolated from ethanol-fixed samples from the open Mediterranean Sea (Table 1). The lorica resembled a peg (Fig. 1, H), shorter than *Salpingella* and narrower than *Amphorides*. The specimens were longer than *A. quadrilineata* ones (165 μm long), while the oral aperture diameter was similar in both species (45 μm in diameter).

Family Ptychocyliidae

This family was represented in GenBank exclusively by the genus *Favella* Jörgensen including the type *F. ehrenbergi* Claparède & Lachmann and four other species. The lorica of *Favella* is hyaline and forms a cylindrical bowl-like with an aboral constriction. We identified two morphologically different species that we identified as *Favella azorica* (Cleve) Jörgensen and *Favella adriatica* Imhof on the basis of lorica size and shape. The figure 1-I represented two bell-shaped specimens of *F. azorica* (100 μm long, 50 μm oral aperture). From the same sampling site in Villefranche, we sequenced the SSU-rDNA from a bigger specimen of *F. adriatica* (190 μm long, 80 μm oral aperture) with a pedicel (Fig. 1, J).

Family Codonellidae

The tintinnids of this family are characterized by tubular, bowl- or cup-shaped loricae, usually reticulated. The lorica is typically heavily covered of agglutinated particles, which often make difficult to observe its details. Sequences from these species, common in estuarine and coastal environments, dominated the public sequence data bases. The family represents about 50% of

the tintinnid sequences available in GenBank, especially the genus *Tintinnopsis* Stein with up to 13 species, and two species of *Codonella* Haeckel.

Our specimens of the genus *Tintinnopsis* were collected in coastal waters, and no specimens were observed in samples from the open waters. The specimens of *T. cylindrica* Daday (130-150 μm long, 42-45 μm oral aperture) were identified on the basis of the elongated and narrow lorica with a distinct aboral horn (Fig. 1, K). We obtained two SSU-rDNA sequences of a smaller and bowl-shaped species of 55 μm long and 43 μm of oral diameter, which we attributed to the species *T. rara* Kofoid & Campbell (Fig. 1, L). The genus *Tintinnopsis* is also present in freshwater environments. Two specimens of *T. lacustris* Entz (Fig. 1, M) were isolated from two ponds near Paris (Table 1). The lorica was characterized by a cylindrical bowl-like anterior part and a bowl-like posterior part (60 μm long, 37 μm oral aperture).

The genera *Codonella* and *Codonaria* differ from *Tintinnopsis* in the shorter and wider loricae divided by a constriction into a rounded bowl and an upper collar, both entirely agglomerated with particles. We obtained the SSU-rDNA sequences for the species *Codonella aspera* Kofoid & Campbell from two specimens collected in coastal waters (Table 1). When compared to other congeneric species, *C. aspera* was thinner and with a less flaring collar (90 μm long, 55 μm oral aperture; Fig. 1, N). *Codonaria* is similar to *Codonella* in all respects except the presence of a suboral cone always superposed above collar proper. Three specimens of the type species, *Codonaria cistelulla* (Fol) Kofoid & Campbell, were isolated from both Villefranche and Banyuls in winter. These specimens showed similar lorica shape (neck deeply constricted, bowl globular) and dimensions (100 μm long, 70 μm oral aperture), but the suboral part was more or less agglomerated with particles (Fig. 1, O-P). Finally, another specimen (100 μm long, 70 μm oral aperture; Fig. 1, Q) resembled *C. cistelulla*, but it might represent an incompletely developed specimen and therefore more difficult to identify. To avoid potential misnaming, we named it as *Codonaria* sp.

Family Metacylidae

Members of this family have tubular- or goblet-shaped loricae, with an oral region showing rings or a spiral structure. Some specimens may have a horn, or a pointed aboral end of the loricae. Two sequences attributed to the *Metacylis angulata* Lackey & Balech and an unidentified *Metacylis* sp. were available in GenBank. We determined the sequence of *Metacylis joergenseni* (Cleve) Kofoid & Campbell collected from Banyuls during the

summer. This species was characterized by a pointed aboral end, with a collar narrower and shorter than the bowl-shaped lorica (66 μm long, 40 μm oral aperture). The collar showed 3 rings or annuli (Fig. 1, R).

From the Bay of Villefranche, we obtained sequences of two specimens of *Climacocylis scalaria* (Brandt) Jørgensen. The lorica was hyaline, alveolated, characteristically spiraled and quite large compared to most other tintinnids, also with a large aboral region. The two specimens were of similar dimensions (410-420 μm long, 55 μm oral aperture). One specimen showed a lorica more spiralled and a smaller alveolated aboral end (Fig. 1, S) than the other one (Fig. 1, T).

Family Undellidae

The members of this family are characterized by cylindrical, bowl- or flask-shaped loricae. The aboral end may be flattened, pointed or hemispherical. The wall is hyaline with thick lamellae and often with a suboral thickening. We determined SSU-rDNA sequences for members of this family. One of the specimens corresponded to the type species, *Undella hyalina* Daday (Fig. 1, U). This species was characterized by an upper part cylindrical and a lower part slightly angled in the aboral end. Its lorica, which reached 240 μm , was the longest of the genus.

We also successfully sequenced SSU-rDNAs from two additional species: *U. marsupialis* Brandt (Fig. 1, V) and *U. claparedei* (Entz) Daday (Fig. 1, W) are smaller than the type species. *Undella marsupialis*, morphologically similar to the type, was one half the length of *U. hyalina* (120 μm long) but both species have a similar oral diameter of 60 μm . Two specimens of *U. claparedei* with a bowl-shaped lorica and a suboral thickening (75 μm long, 40 μm oral aperture) were isolated from coastal waters.

Family Rhabdonellidae

This family is characterized by a chalice-shaped lorica, with longitudinal ridges, in some species and often with very long tapered aboral end or horn termed a pedicel. The sequence of *Rhabdonella hebe* Cleve (#AY143566) was the only representative of this family in GenBank. We provided three SSU-rDNA sequences of the type species, *Rhabdonella spiralis* (Fol) Brandt, and one SSU-rDNA sequence of *Rhabdonella elegans* Jørgensen. *Rhabdonella spiralis* had a long vase-shaped lorica with a long aboral pedicel, and covered by vertical grooves from the oral lip to the tip of pedicel. The specimens collected in the Bay of Villefranche and from the port of Banyuls were quite similar in dimensions (405 \pm 5 μm long,

60 µm oral aperture) (Fig. 1, X). A specimen of a variant form collected from the open Mediterranean Sea (Western Basin), identified as *R. spiralis*, showed a lorica shorter (260 µm) than other specimens but with a similar oral diameter (60 µm) (Fig. 1, Y). *Rhabdonella elegans* Jörgensen differed from the type species in the shorter lorica and aboral end (130 µm long, 50 µm oral aperture). We obtained the SSU-rDNA sequence of one *R. elegans* specimen (Fig. 1, Z).

Family Cyttarocylididae

We obtained the SSU-rDNA sequences of representatives of the family Cyttarocylididae. *Cyttarocylis* is characterized by a bell-shaped lorica with a well marked flaring collar set off by constriction and a wall usually reticulated. Our specimens were commonly found in December in the French Mediterranean coastal sites. We determined SSU-rDNA sequences of two species: the type *Cyttarocylis cassis* (Haeckel) Fol and *Cyttarocylis acutiformis* Kofoid & Campbell. *Cyttarocylis cassis* (130 µm long, 110 µm oral aperture) showed a low collar slightly flaring, an aboral end slightly pointed, and the wall strongly reticulated (Fig. 1, AA). *Cyttarocylis acutiformis* Kofoid & Campbell was longer and larger (190 µm long, 130 µm oral aperture) and it possessed a more flaring collar and a more pointed aboral end than *C. eucecryphalus*. Its lorica was also reticulated (Fig. 1, AB).

Family Petalotrichidae

The family Petalotrichidae is characterized by a bell-shaped lorica with well marked flaring collar set off by constriction. We sequenced cells identified as *Petalotricha ampulla* (Fol) Kent, with a lorica in the shape of an ovoid bowl with an oral constriction showing a distinct outer margin (circum-oral shelf). The lorica differed from that of the *Cyttarocylis* spp. by the absence of reticulations in the wall but was similar in oral dimensions (125 µm diameter). Our specimen from Banyuls had a longer bowl (135 µm) with a slightly pointed aboral end (Fig. 1, AC). In contrast, the two Villefranche specimens were slightly shorter (120 µm long) and globular (Fig. 1, AD).

Family Xystonellidae

The family Xystonellidae is characterised by an elongated lorica, more or less cylindrical, often with a pedicel. We obtained SSU-rDNA sequences of three specimens of *Xystonella longicauda* (Brandt) Kofoid & Campbell (Fig. 1, AE). This species had an elongated lorica, hyaline, conical in the upper part and an aboral pedicel forming a lance (280 µm long, 55 µm

oral aperture). The oral rim was slightly thickened and the lorica structure showed an irregular and conspicuous meshwork between wall layers (Fig. 1, AE).

We determined the SSU-rDNA sequence of the genus *Parundella*. The species of the sequence of *P. aculeata* Jörgensen was characterized by its relatively small size (105 µm long, 27 µm oral aperture), a cylindrical upper part then conical contracting gradually towards a short fine pedicel (Fig. 1, AF).

Family Codonellopsidae

The morphology of *Codonellopsis* Jörgensen is highly distinctive. Loricae are divided into a short rounded agglomerated bowl and a hyaline collar with spiral or annular structures. They are typically 80-130 µm long and 25-30 µm large (oral aperture). The type species, *C. orthoceras* (Haeckel) Jörgensen, is characterized by the presence in the aboral end of a stout pedicel and a globular bowl with a neck below a long collar. We determined SSU-rDNA sequences of two specimens of the type species. The loricae, agglomerated with particles, were 125-130 µm long (of which 45-50 µm corresponded to the long collar) with an oral aperture of 30 µm (Fig. 1, AG). We also determined SSU-rDNA sequences for two specimens of *Codonellopsis morchella* Cleve which showed an ovate agglomerated bowl (with mineral particles) with a flattened aboral end, and a long collar slightly shorter than the bowl. The loricae were 90-100 µm long, (of which 20-30 µm corresponded to the collar) with an oral diameter of 30 µm (Fig. AH).

When compared to other genera, *Stenosemella* Jörgensen is characterized by a shorter collar. We determined SSU-rDNA sequences for two specimens of the type species, *Stenosemella ventricosa* (Claparède & Lachmann) Jörgensen. The loricae were 70 µm long with an oral aperture of 30 µm. They showed the specific low collar (10 µm) with 2 spiral turns. The outer wall was densely covered with agglomerated mineral particles (Fig. 1, AI).

Family Dictyocystidae

We isolated an individual identified as *Dictyocysta lepida* (Ehrenberg) Kofoid & Campbell. The lorica was cup-shaped, rounded aborally and agglomerated with biogenic particles (68 µm long, 40 µm oral aperture). This species was easily distinguishable by the cylindrical collar with beams surrounding large windows in one row (Fig. 1, AJ).

Unidentified tintinnids

We obtained SSU-rDNA sequences from two specimens of tintinnid species that we were unable to identify at the species or genus level (Fig. 1, AK-AL). Both specimens were collected from a sample from the shallow freshwater lake, La Albufera, near Valencia, Spain. The loricae were flask-shaped with flattened aboral ends. The overall shape resembled that of *Undella* spp., however, an entirely agglomerated lorica with mineral particles with no obvious secondary structure would place them in the genus *Tintinnopsis*. Up to ten specimens were observed from the same sample. The loricae were 41 ± 6 μm long and the cylindrical oral apertures were 22 ± 2 μm in diameter. We named this species as "unidentified tintinnid" until further studies will elucidate whether it constitutes a new species or a morphotype of a known species.

TABLE S1. List of new tintinnid SSU-rDNA and ITS sequences used for phylogenetic analyses. Accession numbers, geographical origin and collection dates, corresponding photographs and rDNA regions sequenced are provided.

Taxa	GenBank	Geographical origin (date)	Figure	SSU	ITS+5.8S
<i>Tintinnidium fluviatile</i> CBO42		Chevreuse Valley (2 May 2011)	1A	X	X
<i>Eutintinnus tubulosus</i> FG735		Banyuls (30 Apr 2009)	1B	X	X
<i>Eutintinnus fraknoi</i> FG597		BOUM cruise station A (14 Jul 2008)	1C	X	X
<i>Eutintinnus fraknoi</i> FG604		BOUM cruise station B (4 Jul 2008)	1C	X	X
<i>Eutintinnus apertus</i> CB785		Villefranche (1 Dec 2010)	1D	X	
<i>Salpingella acuminata</i> FG304		Banyuls (30 Nov 2008)	1E	X	X
<i>Amphorides quadrilineata</i> FG1141		Villefranche (3 Oct 2009)	1F	X	X
<i>Amphorides quadrilineata</i> FG249		Marseille (21 Aug 2008)	1F	X	
<i>Amphorides quadrilineata</i> FG618		BOUM cruise station A (14 Jul 2008)	1G	X	X
<i>Amphorides quadrilineata</i> FG293		Banyuls (3 Dec 2008)	1G	X	
<i>Amphorides quadrilineata</i> FG295		Banyuls (5 Dec 2008)	1G	X	
<i>Steenstrupiella steenstrupi</i> FG243		Marseille (9 Feb 2008)	1H	X	
<i>Steenstrupiella steenstrupi</i> FG1368		Villefranche (18 Oct 2009)	1H	X	
<i>Favella azorica</i> FG1106		Villefranche (25 Sep 2009)	1I	X	
<i>Favella azorica</i> FG1111		Villefranche (29 Sep 2009)	1I	X	
<i>Favella adriatica</i> FG1102		Villefranche (3 Oct 2009)	1J	X	
<i>Tintinnopsis cylindrica</i> FG245		Marseille (1 Aug 2008)	1K	X	
<i>Tintinnopsis cylindrica</i> CB358		Villefranche (24 Sep 2009)	1K	X	
<i>Tintinnopsis cylindrica</i> FG804		Banyuls (19 Aug 2009)	1K	X	
<i>Tintinnopsis rara</i> FG1363		Villefranche (17 Oct 2009)	1L	X	
<i>Tintinnopsis rara</i> FG1364		Villefranche (17 Oct 2009)	1L	X	
<i>Tintinnopsis lacustris</i> CBO1		Chevreuse Valley (15 Jun 2010)	1M	X	X
<i>Tintinnopsis lacustris</i> CBO8		Chevreuse Valley (15 Jun 2010)	1M	X	X
<i>Codonella aspera</i> FG300		Banyuls (10 Nov 2008)	1N	X	
<i>Codonella aspera</i> CB235		Villefranche (16 April 2009)	1N	X	X
<i>Codonaria cistellula</i> FG42		Banyuls (26 Feb 2009)	1O	X	X
<i>Codonaria cistellula</i> CB871		Villefranche (2 Dec 2010)	1P	X	
<i>Codonaria cistellula</i> CB25		Villefranche (18 Feb 2009)	1P	X	X
<i>Codonaria</i> sp. CB82		Villefranche (20 Feb 2009)	1Q	X	X
<i>Metacyclis joergenseni</i> FG1143		Banyuls (3 Aug 2009)	1R	X	
<i>Climacocylis scalaria</i> FG1118		Villefranche (29 Sept 2009)	1S	X	
<i>Climacocylis scalaria</i> FG1116		Villefranche (26 Sept 2009)	1T	X	
<i>Undella hyalina</i> CB854		Villefranche (2 Dec 2010)	1U	X	X

<i>Undella hyalina</i> CB314	Villefranche (25 Jun 2009)	1U	X	
<i>Undella marsupialis</i> CBa	Villefranche (3 Dec 2008)	1V	X	
<i>Undella marsupialis</i> FG1360	Villefranche (12 Nov 2009)	1V	X	
<i>Undella claparedei</i> CB818	Villefranche (1 Dec 2010)	1W	X	X
<i>Undella claparedei</i> FG307	Banyuls (7 Dec 2008)	1W	X	X
<i>Rhabdonella spiralis</i> CB356	Villefranche (24 Sep 2009)	1X	X	
<i>Rhabdonella spiralis</i> FG297	Banyuls (21 Nov 2008)	1X	X	
<i>Rhabdonella spiralis</i> FG586	BOUM cruise station A (14 Jul 2008)	1Y	X	X
<i>Rhabdonella elegans</i> FG1392	Villefranche (29 Oct 2009)	1Z	X	X
<i>Cyttarocylis cassis</i> FG302	Banyuls (3 Dec 2008)	1AA	X	
<i>Cyttarocylis cassis</i> CB836	Villefranche (2 Dec 2010)	1AA	X	X
<i>Cyttarocylis acutiformis</i> CB873	Villefranche (2 Dec 2010)	1AB	X	
<i>Petalotricha ampulla</i> FG301	Banyuls (3 Dec 2008)	1AC	X	
<i>Petalotricha ampulla</i> FG1399	Villefranche (10 Dec 2009)	1AD	X	X
<i>Petalotricha ampulla</i> CB837	Villefranche (2 Dec 2010)	1AD	X	X
<i>Xystonella longicauda</i> FG612	BOUM cruise station B (4 Jul 2008)	1AE	X	X
<i>Xystonella longicauda</i> FG1133	Villefranche (1 Oct 2009)	1AE	X	
<i>Xystonella longicauda</i> FG1130	Villefranche (27 Sep 2009)	1AE	X	
<i>Parundella aculeata</i> CB355	Villefranche (24 Sep 2009)	1AF	X	
<i>Codonellopsis orthoceras</i> FG298	Banyuls (17 Nov 2008)	1AG	X	
<i>Codonellopsis orthoceras</i> FG1365	Villefranche (17 Oct 2009)	1AG	X	
<i>Codonellopsis morchella</i> FG260	Marseille (11 Jan 2008)	1AH	X	
<i>Codonellopsis morchella</i> CB77	Villefranche (20 Feb 2009)	1AH	X	X
<i>Codonellopsis morchella</i> FG1123	Villefranche (26 Sep 2009)	1AH	X	
<i>Stenosemella ventricosa</i> CB73	Villefranche (20 Feb 2009)	1AI	X	X
<i>Stenosemella ventricosa</i> CB875	Villefranche (2 Dec 2010)	1AI	X	X
<i>Dictyocysta lepida</i> FG303	Banyuls (30 Nov 2008)	1AJ	X	
Unidentified tintinnid FG2102	Albufera lake (23 Mar 2011)	1AK	X	
Unidentified tintinnid FG2103	Albufera lake(23 Mar 2011)	1AL	X	

Table S2. Approximately Unbiased test results.

Topology constraints	Log-likelihood	AU p value
Unconstrained	-10437.88	0.78
Tintinnid <i>sensu stricto</i> (with lorica) monophyletic	-10439.85	0.22
Unconstrained	-10437.88	0.87
<i>Tintinnopsis</i> monophyletic (without considering <i>Tintinnopsis rara</i> FG1363 & FG1364, <i>Tintinnopsis fimbriata</i> AY143560, <i>Tintinnopsis subacuta</i> EU399541, <i>Tintinnopsis</i> sp. 2 FJ422986, <i>Tintinnopsis lohmanni</i> FJ196076)	-10443.75	0.24
<i>Tintinnopsis</i> monophyletic (without considering <i>Tintinnopsis fimbriata</i> AY143560, <i>Tintinnopsis subacuta</i> EU399541, <i>Tintinnopsis</i> sp. 2 FJ422986, <i>Tintinnopsis lohmanni</i> FJ196076)	-10446.33	0.16
<i>Tintinnopsis</i> monophyletic	-10486.91	0

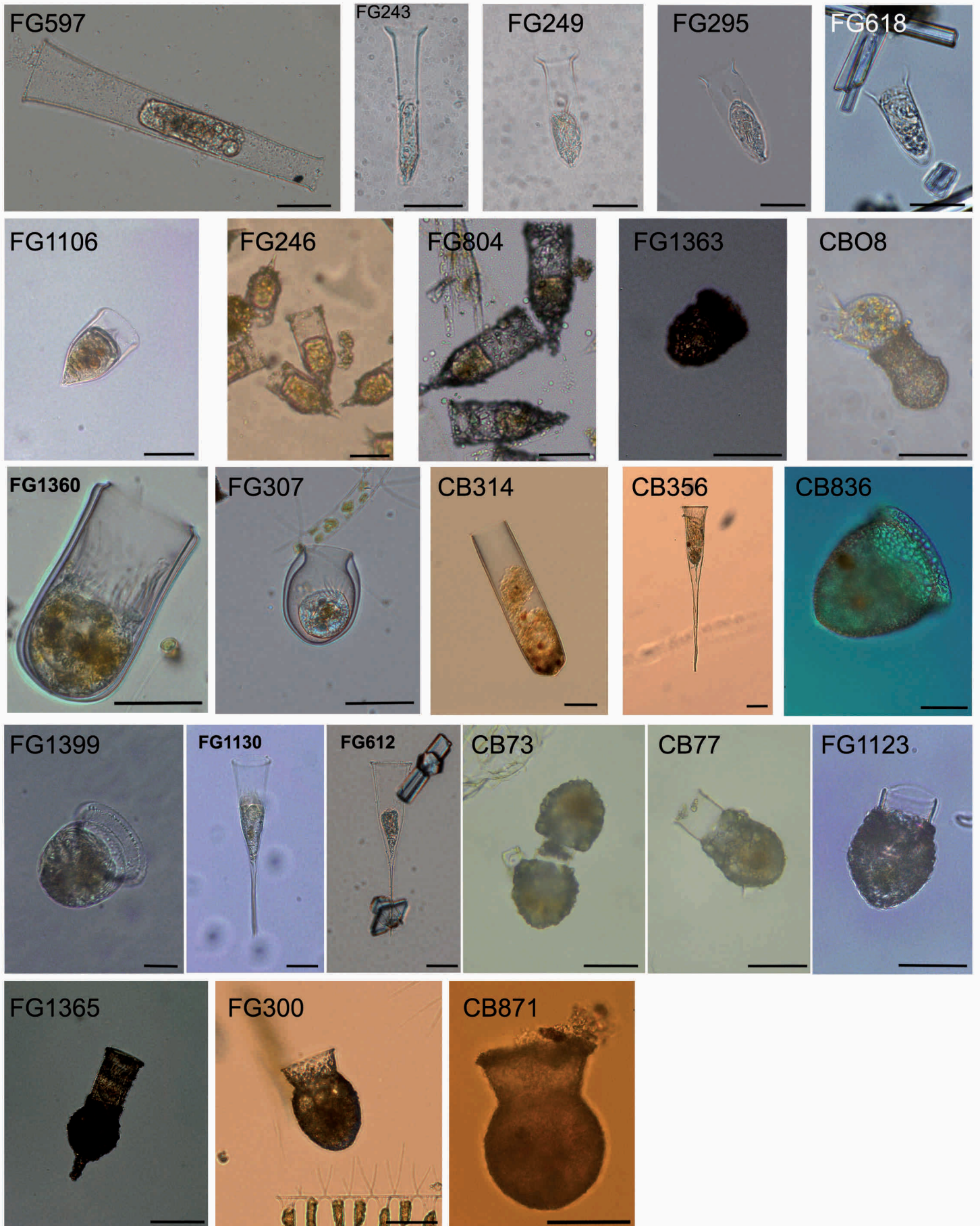


Figure S1. Light micrographs of additional tintinnid cells from species not shown in Fig. 1 but for which we obtained sequences. Scale bars 50 μm. Name of each isolate is indicated on its picture.

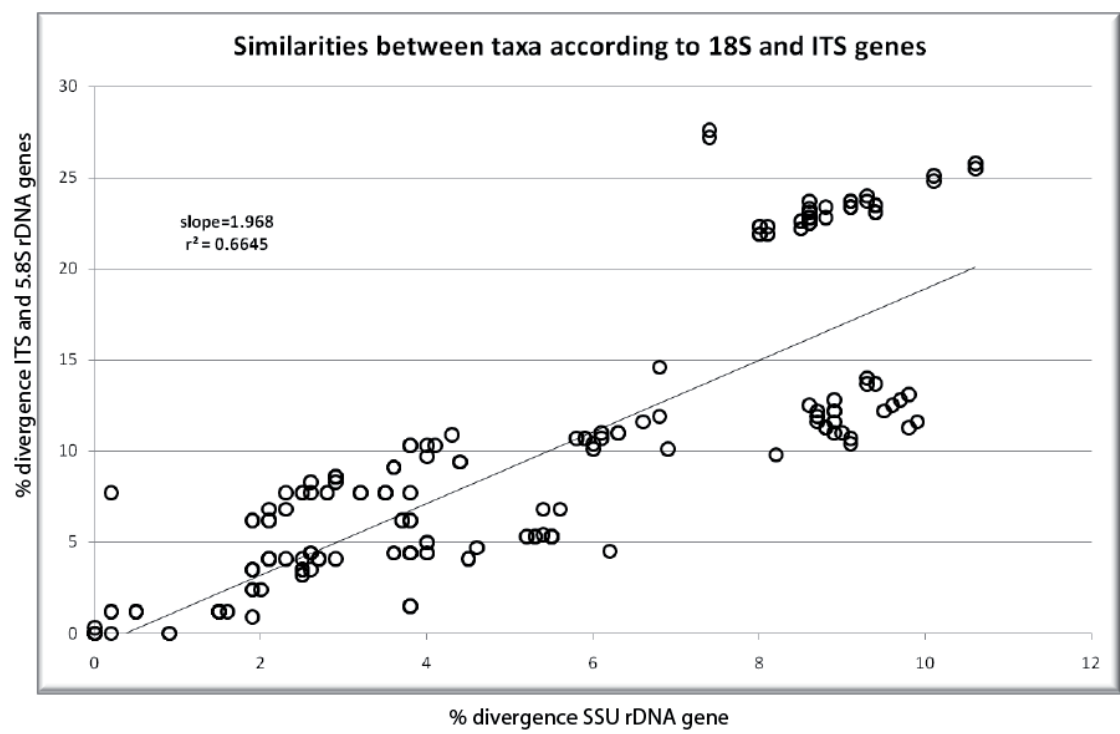


Figure S2. Comparison of sequence divergence between the SSU-rDNA and ITS+5.8S-rDNA markers.

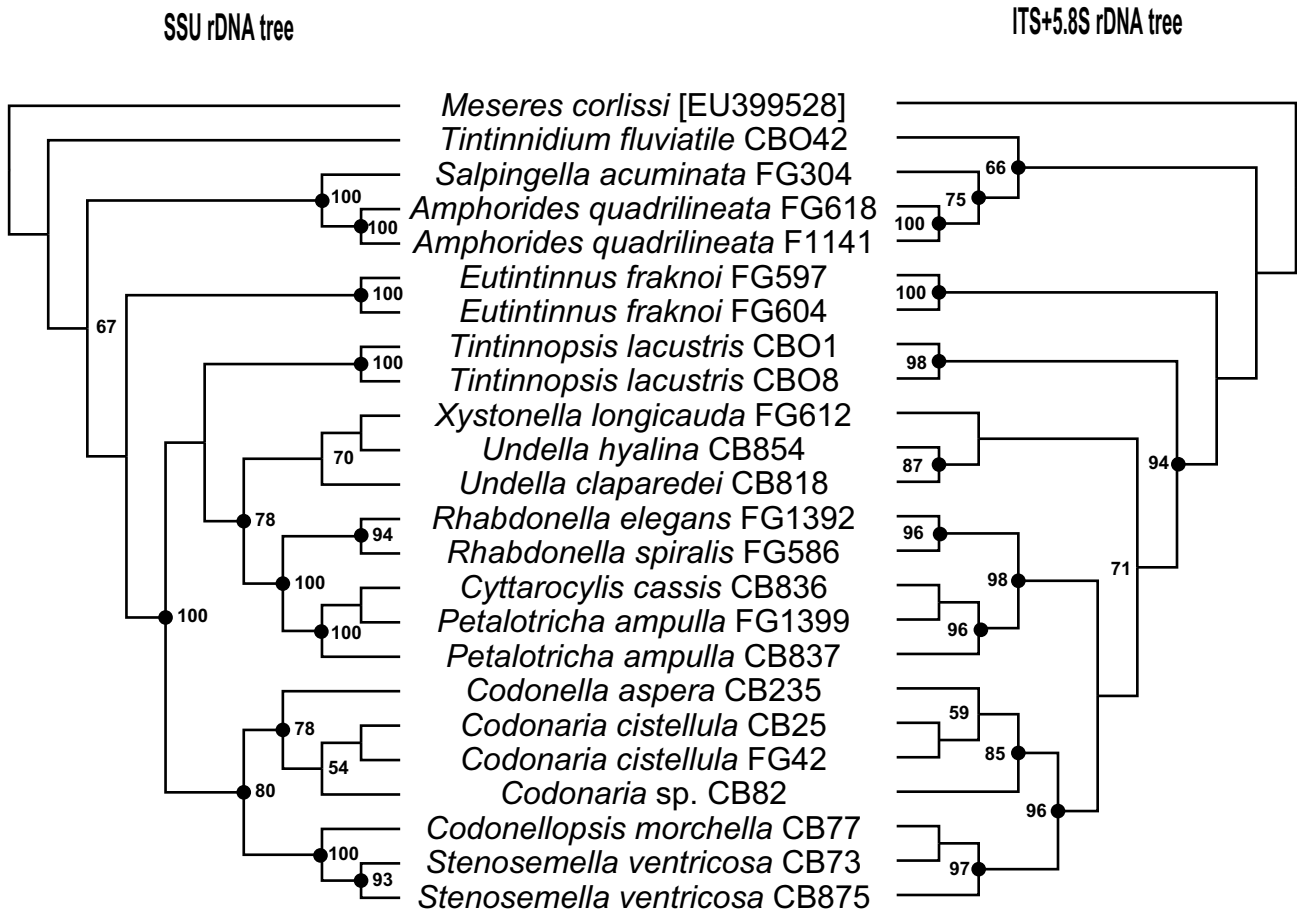


Figure S3. Comparison between maximum likelihood phylogenetic trees reconstructed from the SSU-rDNA data set (left) and the combined ITS and 5.8S-rDNA data set (right). Sequence data sets contained 1,328 (SSU-rDNA) and 354 (ITS+5.8S-rDNA) aligned positions. Numbers at nodes are bootstrap values (values under 50% are omitted). Bayesian posterior probabilities higher than 0.90 are indicated by filled circles.

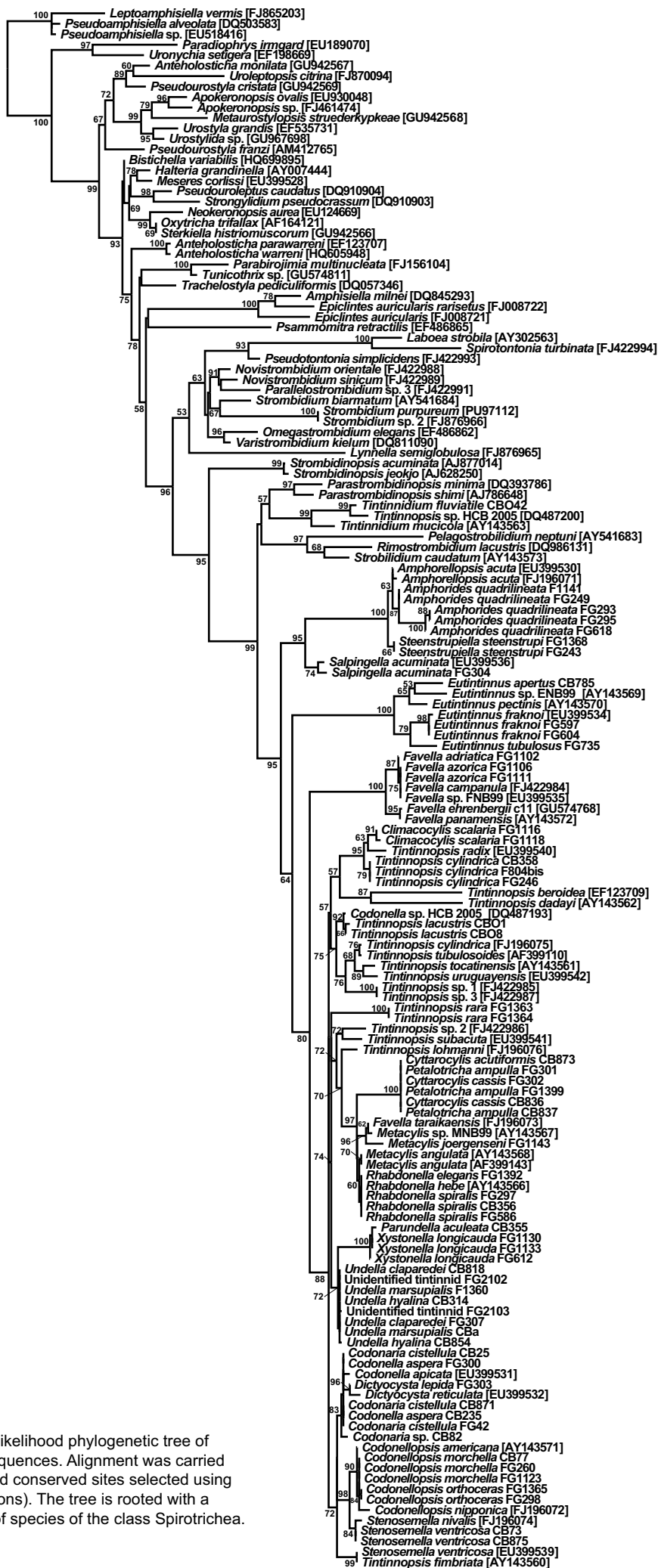


Figure S4. Maximum likelihood phylogenetic tree of tintinnid SSU rDNA sequences. Alignment was carried out using MUSCLE and conserved sites selected using GBLOCKS (999 positions). The tree is rooted with a variety of sequences of species of the class Spirotrichea.

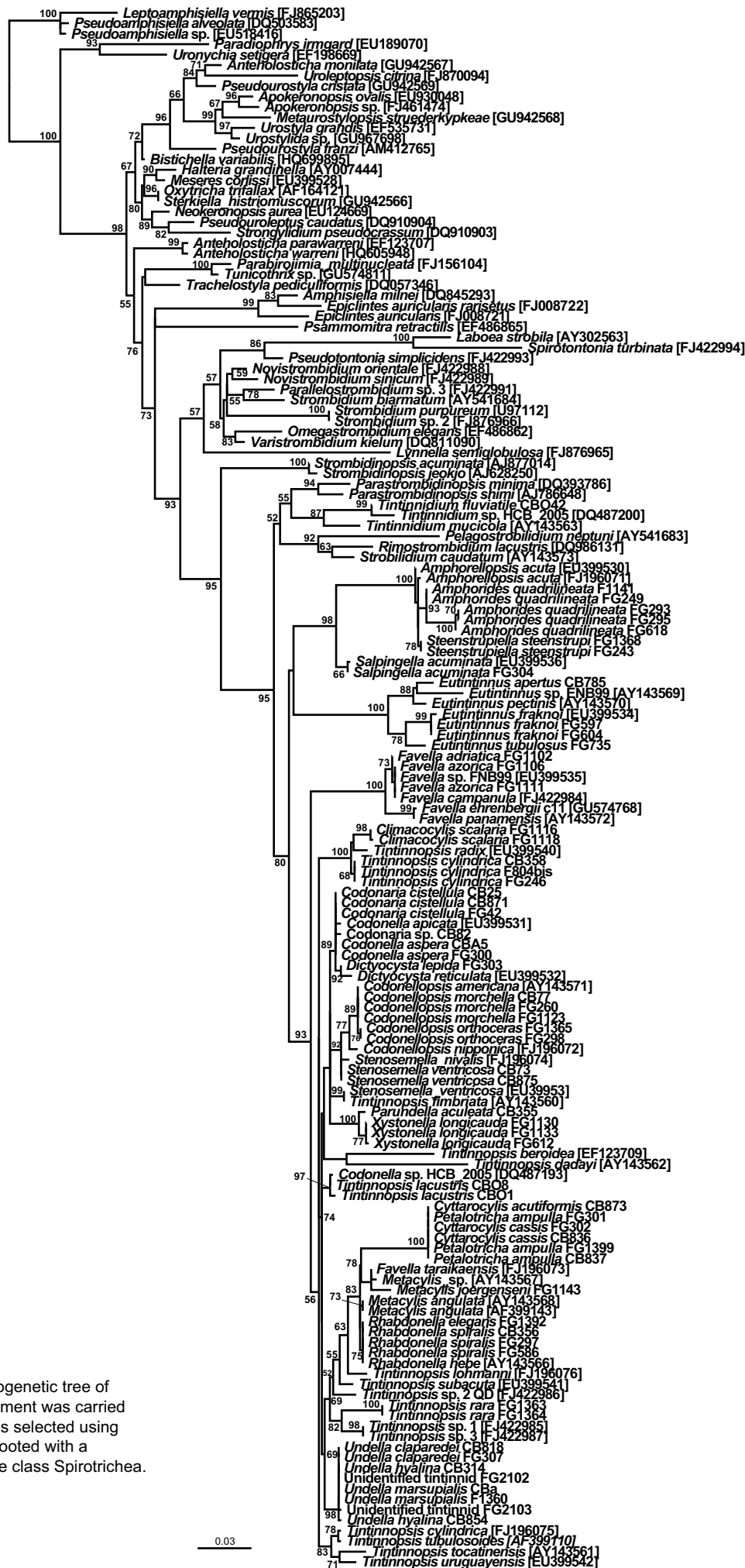


Figure S5. Maximum likelihood phylogenetic tree of tintinnid SSU rDNA sequences. Alignment was carried out using MAFFT and conserved sites selected using BMGE (1007 positions). The tree is rooted with a variety of sequences of species of the class Spirotrichea.



Figure S6. Parsimony optimization of the habitat character in Tintinnida, based on the combined tree shown in Fig. 2. Freshwater species are coloured in blue.

3. Synthèse et éléments de discussion

Nous avons cherché à caractériser au mieux la diversité des tintinnidés en nous intéressant à de nombreux individus de diverses morphologies (morpho-espèces). Cette étude basée sur des cellules uniques isolées nous a permis de réévaluer une partie de la classification des tintinnidés à différentes échelles taxonomiques.

D'abord à l'échelle de l'ordre des Tintinnida, les résultats de ce chapitre ont montré que la position exacte de la racine des tintinnidés dans nos différentes analyses phylogénétiques par maximum de vraisemblance ou par approche bayésienne était faiblement soutenue. En effet, la position basale des espèces du genre *Tintinnidium*, reconnues comme présentant des caractères ancestraux de par la morphologie de la lorica (Kofoid et Campbell 1929, 1939) et la disposition de l'infaciliature (Laval-Peuto et Brownlee 1986), n'est pas soutenue de façon robuste. Dans nos phylogénies moléculaires basées sur l'ADNr 18S, ces espèces groupent avec des espèces de choreotrichidés sans loricae comme *Strobilidium caudatum* ou *Pelagostrobilidium neptuni*. Pour évaluer de façon claire la position phylogénétique de *Tintinnidium* spp., il est nécessaire d'obtenir de l'information génétique supplémentaire, notamment en séquençant d'autres marqueurs moléculaires conservés au sein des ciliés. À l'exception de la position ambiguë du genre *Tintinnidium*, nous avons confirmé pour l'ensemble des autres espèces présentant les caractéristiques morphologiques des tintinnidés, qu'elles forment un groupe monophylétique bien supporté (Snoeyenbos-West et al. 2002; Strüder-Kypke et Lynn 2003, 2008).

Ensuite, à l'échelle des familles, les phylogénies moléculaires construites à l'aide des séquences d'ADNr 18S seules ou avec d'autres marqueurs (ADNr 5.8S + ITS) ont montré qu'une réévaluation de la classification de certaines familles était nécessaire. Nous avons élevé au rang de nouvelle famille le genre *Eutintinnus* à la fois sur des critères morphologiques (ouverture de la lorica aux deux extrémités) et sur des critères moléculaires, les séquences formant un groupe monophylétique distinct des autres espèces faisant *a priori* partie de la même famille des Tintinnidae. Inversement, certaines familles, que l'on supposait éloignées sur des critères morphologiques, forment des groupes monophylétiques bien soutenus. Les familles des Codonellidae *sensu stricto* et la famille des Dictyocystidae ont ainsi été fusionnées en une seule. Ces réorganisations taxonomiques visent à clarifier les liens de parentés existants entre les différents genres et la classification classique.

Enfin, à l'échelle de la diversité intra-générique et intra-spécifique, certains tintinnidés que l'on pensait appartenir à différentes espèces, différents genres et même différentes familles à cause de leur loricae très distinctes, semblent génétiquement appartenir à la même espèce. En effet, les espèces *Petalotricha ampulla* (Petalotrichidae) et *Cyttarocyliis* spp. (Cyttarocylididae) partagent des séquences d'ADNr 18S et des séquences de région des ITS identiques. Nous avons donc proposé la nouvelle combinaison *C. ampulla*. Ce terme de nouvelle combinaison (*combinatio nova*) est utilisé pour désigner une espèce transférée dans un autre genre, tandis que la précédente nomenclature (ici *Petalotricha ampulla*) devient le synonyme originel, ou basonyme. Aussi, nous avons découvert que les espèces *Rhabdonella spiralis* et *R. elegans*, pourtant très différentes en longueur, partagent également les mêmes séquences génétiques (ADNr 18S et ITS). Ces observations nous font supposer que les différentes formes que peuvent présenter les loricae peuvent être en réalité des écotypes (caractéristiques adaptées à un environnement particulier) ou des étapes différentes de développement. Ces cas conflictuels entre l'information moléculaire et morphologique sont à comparer avec le cas bien étudié des loricae des genres *Favella* et *Coxliella* qui présentaient des morphologies très distinctes, mais qui ont montré en culture être des loricae polymorphes d'une seule et même espèce (Kim et al. 2010; Laval-Peuto 1977). Les quelques espèces nouvelles de *Tintinnopsis* confirment que chez ce genre décrit comme paraphylétique lors des études précédentes de phylogénie moléculaire (Li et al. 2009; Snoeyenbos-West et al. 2002), la lorica est un caractère très peu informatif. Les précédentes études, et la nôtre, nous montrent que sous le genre *Tintinnopsis* sont décrites de nombreuses espèces qui devraient probablement appartenir à différents genres. L'hypothèse d'une observation difficile des détails caractéristiques des loricae des *Tintinnopsis* cachés par l'agglomération importante de particules est souvent avancée (McManus et Katz 2009). Un test AU soutient que l'hypothèse alternative d'une origine commune exclusive à tous les *Tintinnopsis* spp. n'est pas vraisemblable (p-value=0). En revanche, lorsque l'on exclut de l'analyse *Tintinnopsis fimbriata* AY143560, *Tintinnopsis sub-acuta* EU399541, *Tintinnopsis* sp. 2 FJ422986 et *Tintinnopsis lohmanni* FJ196076, l'hypothèse d'origine exclusive des onze *Tintinnopsis* spp. restantes devient vraisemblable (p-value=0,16). L'histoire de ce genre apparaît encore plus compliquée que prévu. D'abord, nous avons observé une espèce vivant en eau douce qui présente une lorica entièrement agglomérée semblable aux *Tintinnopsis*, mais qui branche clairement avec un genre marin à lorica hyaline, *Undella*. Ensuite, l'espèce *Clymacocylis*

scalaria qui présente une lorica très caractéristique (hyaline, relativement grande, alvéolée) branche dans notre phylogénie basée sur l'ADNr 18S au sein d'un clade de *Tintinnopsis* spp.

Ainsi, ce travail de phylogénie moléculaire a permis de réorganiser une partie de la classification des tintinnidés aux échelles macroévolutive et microévolutive. Il est toutefois notable qu'une partie des espèces groupent en différents clades bien soutenus, mais les relations entre ces clades restent non résolues. Afin d'améliorer la résolution des relations de parenté entre les différentes familles, il s'avère donc nécessaire dans l'avenir d'amplifier d'autres marqueurs moléculaires de choix pour augmenter l'information phylogénétique. Les approches d'amplification des marqueurs génétiques par PCR sur cellules uniques, suivant une étape d'amplification complète du génome (WGA), devraient faciliter l'obtention de suffisamment de matériel génétique pour générer des séquences pour un plus grand nombre de marqueurs. S'ajoute à cette limite de l'information génétique le fait qu'il existe bien d'autres espèces de tintinnidés dans les eaux marines du monde que celles que nous avons récoltées. Afin d'enrichir l'échantillonnage taxonomique des tintinnidés, nous encourageons à continuer ces études de diversité moléculaire par PCR directe sur cellules individuelles de tintinnidés pour caractériser génétiquement la diversité la plus large possible à l'échelle mondiale.

Malgré certains décalages observés entre la phylogénie moléculaire et la taxonomie classique, de nombreuses espèces présentant des morphologies variées ont été liées de façon évidente avec les séquences d'ADNr 18S, et parfois des régions des ITS, correspondantes. Nous pourrions donc ensuite, en comparant les séquences entre elles, déterminer un seuil de pourcentage de similarité de séquences permettant de discriminer la diversité intra-spécifique de la diversité inter-spécifique chez les tintinnidés.

Enfin, un cadre phylogénétique enrichi a ainsi été créé pour servir de base de travail la plus complète possible pour la suite de mes travaux de recherche. Par comparaison avec cette base de données, les séquences d'ADNr 18S obtenues par amplification à partir d'ADN environnemental à l'aide d'amorces spécifiques des tintinnidés pourront être identifiées le plus justement possible jusqu'à l'échelle de l'espèce.

B. Analyse comparative des études moléculaires environnementales de diversité et des études morphologiques classiques chez les ciliés tintinnidés

1. Contexte et objectifs

Comme nous l'avons vu dans l'introduction, l'amplification, le clonage et le séquençage Sanger des gènes codant pour l'ADNr 18S est devenu l'approche standard pour décrire la composition des communautés de protistes dans les échantillons environnementaux. L'apparition récente de nouvelles technologies de séquençage massif, en particulier le pyroséquençage 454, a bouleversé cette approche non seulement par le très grand nombre de séquences obtenues à moindre coût, mais par l'économie de l'étape de clonage supposée associée à différents biais. Le très grand nombre de séquences générées par cette technologie a permis l'identification des séquences appartenant à une diversité microbienne rare extrêmement diverse et dont le rôle écologique reste à définir (Pedrós-Alió 2007). Toutefois, des discussions sont en cours pour savoir si cette diversité microbienne "rare", ou du moins une partie de cette diversité, existe vraiment ou si elle est le fruit d'erreurs de séquençage. Le fait de travailler sur un groupe de protistes planctoniques pour lequel il est possible de réaliser une expertise taxonomique détaillée au microscope optique m'offre une occasion unique d'évaluer l'efficacité de ces études de diversité moléculaire. En effet, il existe peu de comparaisons entre les descriptions d'une communauté microbienne naturelle réalisées à l'aide des études de diversité moléculaire environnementales conventionnelles (amplification, clonage, séquençage) ou récentes (amplification, pyroséquençage) et la description de ces mêmes communautés sur un examen morphologique classique.

L'objectif central de cette partie était avant tout de valider ou non, par l'inspection morphologique, les études environnementales de diversité moléculaire basées sur l'amplification de l'ADNr 18S pour décrire la diversité des tintinnidés sur les mêmes échantillons. Ensuite, un objectif plus concret de l'étude a été de savoir parmi les différentes méthodes qui existent pour établir des UTOs à partir des séquences d'ADNr 18S générées par pyroséquençage, lesquelles donnent des résultats cohérents. Un troisième objectif était d'évaluer les différences entre les méthodes de pyroséquençage et de clonage / séquençage Sanger pour décrire la diversité des tintinnidés.

Pour atteindre ces buts, nous avons comparé ces différentes approches (morphologique, clonage/séquençage d'amplicons, pyroséquençage d'amplicons) sur deux échantillons marins

collectés en deux sites distincts dans la Méditerranée : la rade de Villefranche-sur-Mer (mer Ligure) et la mer Ionienne. J'ai utilisé deux marqueurs pour l'approche de pyroséquençage, en amplifiant une région hypervariable de l'ADNr 18S (V4) et la région des ITS, tandis que j'ai utilisé comme marqueur l'ADNr 18S sur pratiquement toute sa longueur dans l'approche d'amplification, clonage et séquençage Sanger.

2. Manuscrit de l'article 2 : Accuracy of protist diversity assessments: morphology compared to cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study

ISME Journal, under review

Accuracy of protist diversity assessments: morphology compared to cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study

Charles Bachy¹, John R. Dolan², Purificación López-García¹, Philippe Deschamps¹, David Moreira^{1*}

¹*Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, bâtiment 360, 91405 Orsay Cedex, France;* ²*Université Pierre et Marie Curie, CNRS UMR 7093, Laboratoire d'Océanographie de Villefranche, Marine Microbial Ecology, Station Zoologique, 06230 Villefranche-sur-Mer, France*

Correspondence: D. Moreira, Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, bâtiment 360, 91405 Orsay Cedex, France.

E-mail: david.moreira@u-psud.fr

Until recently, Sanger sequencing and construction of gene libraries of PCR-amplified 18S rRNA genes was the standard approach used to describe microbial community composition in environmental samples. Now, the use of deep-sequencing technologies, and in particular 454 pyrosequencing, has displaced that approach not only because of the high number of reads obtained but also because it sidesteps cloning and its potential associated biases. Deep 18S rDNA pyrosequencing has revealed a vast diversity of infrequent sequences, which has led to proposals of the existence of an extremely diverse microbial "rare biosphere" as a new paradigm in protistan ecology. While rare microbes no doubt exist, critical views suggest that many rare sequences may be artifacts. However, information about how diversity revealed by molecular methods relates to that revealed by classical morphology approaches is practically non-existent. To address this issue, we assessed diversity using different approaches in tintinnid ciliates, a species-rich group in which species can be easily distinguished morphologically. We estimated diversity in the same samples employing morphological observations as well as both classical cloning and sequencing and pyrosequencing of two different markers, the 18S rDNA and the ITS regions. Our study focused on two Mediterranean marine samples with different patterns of tintinnid diversity. We applied a variety of computational approaches currently used to analyze pyrosequence reads. We found that both molecular approaches were efficient in detecting the tintinnid species observed using a microscope and revealed similar phylogenetic structures of the tintinnid community at the species level. However, depending on the method used to analyze the pyrosequencing results, we observed discrepancies with the morphology-based assessments, up to several orders of magnitude. In several cases, the inferred number of operational taxonomic units (OTUs) largely exceeded the total number of tintinnid cells in the samples. Such inflation of the OTU numbers corresponded to "rare biosphere" composed largely of artefacts supporting the view that a careful and rigorous analysis of pyrosequencing datasets, including data denoising and sequence clustering with well-adjusted parameters, is necessary to accurately describe microbial biodiversity using this molecular approach. Combining approaches might be a prerequisite for protistan diversity surveys to avoid misleading interpretations of ecological data.

Keywords: amplicon; biodiversity; denoising; plankton; pyrosequencing; rare biosphere

Introduction

During the last decade, studies of the diversity of unicellular eukaryotes, or protists, have been deeply impacted by the advent of cultivation-independent molecular methods. Initially, this approach was based on analysis of 18S rRNA genes using PCR amplification, cloning and Sanger sequencing. These approaches revealed a wide and unsuspected diversity in oceans (Guillou et al 2008, López-García et al 2001, Massana and Pedrós-Alió 2008, Moon-van der Staay et al 2001, Not et al 2007a, Not et al 2007b). More recently, the development of 454 pyrosequencing, has permitted 'deep sequencing', producing thousands of sequence reads from a single sample and revolutionizing the study of natural microbial communities (Margulies et al 2005, Sogin et al 2006). In particular, this has revealed the existence of large collections of lineages represented by only a few sequences, which have been interpreted as evidence of a rare microbial biosphere (Pedrós-Alió 2007). Furthermore, very high estimates of eukaryotic species richness have been established in this way, notably in marine waters with several thousands of different OTUs detected in biomass from a few liters (Brown et al 2009, Edgcomb et al 2011). 18S rDNA amplicon pyrosequencing has become an accepted approach for protist diversity studies.

The sequence length obtained by pyrosequencing, originally of ~100 bp, now exceeds 400 bp (Titanium chemistry) or, very recently, even 700 bp (FLX+ technology), allowing for more precise taxonomic attribution of reads. However, multiple sources of bias have been identified from pyrosequencing errors (Huse et al 2007, Kunin et al 2010), poor alignment quality (Schloss 2010, Sun et al 2009), inappropriate clustering approaches (Huse et al 2010) or inconsistent results from different targeted 18S rDNA regions (Stoeck et al 2010). These biases can have profound effects on diversity estimation. Algorithms to remove pyrosequencing errors have been developed to reduce the overestimation of species richness. This is the case of AmpliconNoise (Quince et al 2011), based on flowgram clustering, or the single-linkage preclustering (SLP) approach (Huse et al 2010), based on sequence similarity analysis. Nevertheless, only a few studies have looked at the effects induced by different *in silico* analytical approaches when assessing species diversity in natural samples using pyrosequencing reads (Bik et al 2012).

In spite of a rich and long history of taxonomic description of microbial eukaryotes, accurate comparative studies between morphological and molecular approaches in natural

environments are still rare. One exception is the recent study of the overall diversity of eukaryotes in a freshwater lake, which demonstrated the efficiency of 18S rDNA pyrosequencing in detecting rare species but revealed strong incongruencies among taxa with regard to their relative abundances (Medinger et al 2010). However, environmental 18S rDNA sequence data require knowledge concerning how genotypic and phenotypic diversity relate in order to be reliable and informative (Caron et al 2009). Consequently, we need to know how diversity defined by classical morphology approaches relates with that based on molecular methods, whether conventional (cloning / Sanger sequencing) or recent NGS technologies (e.g., pyrosequencing). Employing the different approaches on the same samples, using morphology-based species assessments as a control, could be a robust approach to judge the reality of microbial rare biosphere raised by pyrosequencing amplification of 18S rDNA fragments. To profitably compare these approaches requires application on a biological model group which is abundant, diverse and easily distinguishable by morphology.

The species-rich order Tintinnida represents one of the most morphologically diverse groups of planktonic protists. Tintinnids are commonly found in marine surface waters of the globe, in concentrations generally ranging from 10^1 to 10^4 cells per liter (Pierce and Turner 1993). By examining untreated specimens in plankton settling chambers using an inverted microscope, they can be easily identified based on their species-specific secreted shell, the lorica (e.g., Dolan et al 2009). Several studies based on tintinnid single-cell identification, 18S rRNA gene and ITS PCR amplification and sequencing have recently been carried out, providing a comprehensive phylogenetic framework allowing the identification of environmental sequences at the species-level (Bachy et al 2012, Saccà et al 2012, Santoferrara et al 2012, Snoeyenbos-West et al 2002, Strüder-Kypke and Lynn 2008). So, applying a taxon-specific approach focused on marine tintinnid ciliates we assessed diversity in aliquots of the same samples using microscopic analysis (species identified based on classical morphological descriptions) and both cloning / Sanger sequencing and pyrosequencing analyses of 18S rDNA and ITS regions. We also examined different pyrosequencing data processing approaches to estimate species richness.

Materials and methods

Sample collection

Seawater samples from two sites in Mediterranean Sea were collected with Niskin bottles at the deep chlorophyll maximum (DCM) depth: the Bay of Villefranche-sur-Mer (Ligurian Sea, sample VilleFr-43) at the site 'Point B+' (43°41'00"N, 7°18'44"E; 50 m depth) on November 18th 2009; and the Ionian Sea (sample Ioni7; 36°28'58"N, 15°38'58"E; 77 m depth) on October 1st, 2010, during the Marmara2010 cruise. A total of 20 liters were collected in the coastal station of Villefranche and 168 liters in the Ionian Sea offshore station. Plankton samples were immediately after sampling pre-filtered through 200 µm Nitex screen and then subdivided into two equal volumes. For the VilleFr-43 sample, one subsample was immediately preserved with Lugol iodine solution (2% final concentration) and then concentrated via serial gravity sedimentation (i.e., 10 l to 2 l then 200 ml) to a final volume of 20 ml for taxonomic identification by morphological observation (see below). The second subsample was filtered onto 5 µm filter to concentrate all biomass. The filter was conserved in absolute ethanol at -20°C following a protocol already tested for plankton samples (Bachy et al 2011, López-García et al 2003). For the Ioni-7 sample, the two subsamples were concentrated by gravity filtering through a 20-µm Nitex filter to final volumes of ca. 4 ml. One subsample was fixed with Lugol's solution (Sigma, USA) and the other subsample was fixed in ethanol (80% final concentration) for further processing. The cells fixed in ethanol were then filtered onto 5 µm to concentrate and immediately processed for DNA extraction (see below).

Morphological observation

For each subsample fixed in lugol solution, the entire concentrate volume was examined to exhaustion at 200x magnification using an inverted microscope (Olympus IX51) in 2 ml aliquots settled in sedimentation chambers. Assignment of every single cell of tintinnid species was determined based on lorica morphology following standard taxonomic monographs (Kofoid and Campbell 1929, Kofoid and Campbell 1939, Marshall 1969). Empty loricae were included in the enumerations.

Generation of clone libraries and sequencing

Total DNA was purified from filters cut in small pieces using the PowerSoil DNA Isolation Kit (MoBio, USA). Nucleic acids were resuspended in 10 mM Tris-HCl, pH 8. Near full length 18S rDNA fragments were amplified by PCR using the forward tintinnid-specific primer 18S-Tin3F (5'-GCGGTATTTATTAGATAWCAGCC-3'; Bachy et al 2012) and reverse eukaryotic-specific 1498R (5'-CACCTACGCAAACCTTGTTA-3'; López-García et al 2003). PCR was carried out under the following conditions: 35 cycles (denaturation at 94 °C for 15 s, annealing at 52 °C for 30 s, extension at 72 °C for 2 min) preceded by 3 min denaturation at 94 °C, and followed by 15 min extension at 72 °C. We also targeted longer DNA fragments covering near full length 18S rRNA gene, internal transcribed spacers 1 and 2, 5.8S rDNA gene and partial long-subunit rDNA fragment under similar PCR conditions as described above. These fragments were amplified with the forward tintinnid-specific primer 18S-Tin3F and the reverse tintinnid-specific primer 28S-TinR1 (5'-TGGTGCACTAGTATCAAAGT-3'). This primer set yielded an amplicon size of ~2200 bp. Clone libraries were constructed using the Topo TA cloning system (Invitrogen) following the instructions provided by the manufacturer. Positive inserts of expected size were selected from each library and Sanger-sequenced with the forward primer. We generated a total of 200 high-quality partial sequences (>700 bp) for each sample, covering the 18S-rDNA regions also targeted by primers used for pyrosequencing analysis. These clone sequences served for a preliminary phylogenetic analysis and the identification of operational taxonomic units (OTUs, defined here as clusters of sequences having $\geq 99\%$ identity). For each library, we completely sequenced at least one clone per OTU to obtain complete sequences representative of the whole taxonomic diversity found, preferentially the ones which also encompassed the ITS, 5.8S and LSU rDNA. 116 ciliate sequences (among which 100 were tintinnid sequences) formed a reference dataset as phylogenetic framework to check the attribution of environmental clones. From all the OTUs detected by our stringent criterion ($\geq 99\%$ sequence identity), 25 and 17 OTUs corresponded to tintinnid ciliates in the coastal and offshore samples, respectively. 2 OTUs represented by one sequence each were excluded as non-tintinnids from the VilleFr-43 sample. These sequences formed with the previous reference dataset the “18S rDNA” dataset (100 tintinnid reference sequences and 42 clone sequences attributed to tintinnids). The sequences for which the ITS region was also sequenced, formed the “ITS” dataset, comprising a smaller number of 43 sequences (22 tintinnid reference sequences and 21 clone sequences attributed to tintinnids). The 42 complete sequences

identified as tintinnid were submitted to GenBank (accession numbers XXXXXX-XXXXXX).

Clustering of clone sequences into OTUs

To affiliate clone sequences to different taxa, they were compared by BLAST against the reference dataset. A fasta file with all the high-quality partial 18S rDNA clone sequences was created and submitted to the module NDIST of AmpliconNoise (Quince et al 2011) to generate a distance matrix of all pairwise sequence combinations by Needleman-Wunsch pairwise alignment (with default parameters). A first distance matrix was generated with an alignment of all available positions (~840 bp), and a second one with only the region covering positions targeted in the 18S rDNA pyrosequence read analysis (see below). The resulting matrices were used as input for Mothur (Schloss et al 2009) in order to group clone sequences by average linkage clustering at different sequence identity levels.

Generation of amplicon libraries and pyrosequencing

Two sets of primers were designed for this study to specifically target variable regions of the tintinnid 18S rRNA gene and ITS region. The primers were adapted for 454 using the manufacturer recommendations, with the forward and reverse fusion-primer (A/B) and also a 10 bp molecular identifier (MID) tag (to distinguish each of the two samples) inserted between the adapter A and the specific primer. The primers target 18S rDNAs from all tintinnid families sequenced to date. The first set of primers amplified a 330-340 bp fragment of the 18S rRNA gene including the hyper-variable V4 region while the second one amplified a 250-260 bp fragment of the ITS region including partial ITS1, complete 5.8S and partial ITS2. PCR reactions were carried out using 2 µl of environmental DNA in a volume of 20 µl Platinum HIFI reaction mix (Invitrogen) containing the tintinnid-specific 18S-rDNA primers Tin454-18SFw (5'-GTAGTGACAAGAAATAAYGG-3') and Tin454-18SRev (5'-CAACCCCTCARACCAACAG-3'). To amplify the ITS region, PCR reactions were carried out using the tintinnid-specific ITS primers Tin454-ITSFw (5'-CAATYTGTTGCAGRGCGMAAGC-3') and Tin454-ITSRev (5'-AGCAATAGAAGGGCATCTA-3'). The PCR reactions were done under the following conditions: 25 cycles (denaturation at 94 °C for 15 s, annealing at 52 °C for 30 s, extension at 72 °C for 1 min) preceded by 3 min denaturation at 94 °C, and followed by 15 min extension at 72 °C. We carried out only 25 PCR cycles in order to minimize the generation of

recombinant PCR products and compensated for the lower yield by pooling the products of ten independent PCRs. DNA was purified for each amplicon from each sample and mixed in 1:1:1:1 proportions before being 454-pyrosequenced from the A adapter on a 454 GS FLX sequencer using Titanium Roche protocols and supplies (Eurofins MWG Operon, Germany). Sequences are available at the NCBI Short Read Archive (SRA000000).

Low-quality and phylogenetic filtering of pyrosequencing reads

As a first step of filtering, only pyrosequencing reads having exact forward and reverse primer sequences and with no ambiguous nucleotides were kept. The remaining reads were then trimmed after eliminating the primer sequences. Then, they were compared by BLAST (Altschul et al 1997) against the “18S rDNA” and “ITS” datasets (see above) to determine their preliminary taxonomic affiliation. Sequences with a similarity <98% with *bona fide* tintinnid sequences were extracted and their affiliation to the Tintinnida was checked by phylogenetic analysis. These reads were aligned using MAFFT (Katoh et al 2002) and an approximately maximum likelihood (ML) tree was constructed using FastTree (Price et al 2009) with choreotrich and oligotrich sequences as outgroup for the “18S rDNA” dataset and with spirotrich sequences as outgroup for the “ITS” dataset. All 18S rDNA and ITS Ionian Sea reads did belong to tintinnid species, whereas 804 18S rDNA Villefranche reads appeared to belong to strombiliid species, a sister clade of tintinnids, and were therefore discarded. No chimera was detected with the Chimera Slayer algorithm (Haas et al 2011) implemented in Mothur (Schloss 2010), using the “18S rDNA” and “ITS” curated datasets as references. Within each sample, we randomly selected the number of pyrosequence reads contained in the smaller dataset (10,504 and 8,734 reads in VilleFr-43 and Ioni-7, respectively), to avoid potential biased comparisons due to variable dataset sizes (Gihring et al 2012). The trimmed, phylogenetically checked, and size-equivalent pyrosequence read libraries served as start point for subsequent analyses. With the corresponding reads identifiers, we also prepared flowgram datasets for AmpliconNoise, which needs them as input files.

Clustering of pyrosequence reads into OTUs

We applied several ways of clustering the pyrosequencing reads. First, with the quality- and taxonomy-filtered pyrosequence read datasets (see above), we used MAFFT (Katoh et al 2002) with default parameters for multiple sequence alignment and distance matrix generation. We then used Mothur (Schloss et al 2009) to estimate OTUs richness using the

average linkage clustering option. Another clustering process was to apply the NDIST module of AmpliconNoise (Quince et al 2011) to generate the distance matrix of all Needleman-Wunsch pairwise sequence alignments (with default parameters). Then, the average linkage option in the FCluster module of AmpliconNoise was used to cluster reads into OTUs. As an alternative, we also applied the clustering procedure implemented in Usearch (Edgar 2010) with default parameters at different identity levels.

The first steps of filtering (see above) were done to diminish the per nucleotide-error rate of sequence data (estimated at ~0.25% for pyrosequencing, see (Huse et al 2007)). We then used different additional approaches to correct the remaining PCR and pyrosequencing errors. We first analyzed the filtered pyrosequence reads with the single-linkage preclustering algorithm as detailed in Huse et al 2010, with a pairwise distance less than 0.005 used as threshold (equivalent to a single nucleotide difference in the 18S and ITS regions). We then used Mothur to create OTUs using the average linkage option to calculate richness. Alternatively, we ran AmpliconNoise with parameters identified as efficient to minimize the number of spurious OTUs for pyrosequencing (-r 0.005 for FCluster, -s 60.0 -c 0.005 for PyroNoiseM and -s 10.0 -c 0.08 for SeqNoise) (Quince et al 2011). The sequences passing this filter were clustered using pairwise alignments and average-linkage clustering as described above. In addition, we ran CD-HIT-OTU, a recent clustering algorithm for accurate and rapid identification of microbial composition for 18S rDNA pyrosequence data, with default parameters (including -e 0.0025) on the OTU finder web server (<http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/cd-hit-otu/>). It has to be noted that before the last step of OTU clustering with this algorithm, the clusters of only 1 or 2 sequences are considered as noise and therefore removed.

For OTUs inferred using all the different methods, rarefaction analyses were computed with Mothur (Schloss et al 2009).

Phylogenetic analyses

All representative clone sequences and 18S rDNA reads selected after filtering of pyrosequences by AmpliconNoise, together with their most similar sequences identified by BLAST (Altschul et al 1997) in the reference dataset, were aligned using MAFFT. The multiple sequence alignment was then manually edited using the program ED from the MUST

package (Philippe 1993). A maximum likelihood (ML) tree was constructed with FastTree for all these sequences in order to choose one sequence (the longest one) as representative for each OTU for further phylogenetic analyses. The selected dataset was then analyzed by ML using TREEFINDER (Jobb et al 2004) applying a general time reversible (GTR) model of sequence evolution with a 4-rate categories Gamma law to accommodate for among-site rate variation. ML bootstrap proportions were inferred using 1000 replicates.

Results

We studied the tintinnid ciliate community in two DCM marine plankton samples from a coastal (Villefranche: VilleFr-43 sample) and offshore (Ionian Sea: Ioni-7 sample) location. Each sample was divided in two equal parts (2 x 10 l for Villefranche samples and 2 x 84 l for Ionian DCM waters). One part was used for morphological identification of tintinnid cells using a microscope. The second one was used for both 18S rRNA gene library construction and Sanger sequencing and direct pyrosequencing of 18S rRNA gene and ITS amplicons. For these molecular analyses, we designed tintinnid-specific primers allowing the PCR amplification of the targeted regions (see Materials and methods).

Morphology and sequence datasets

In plankton subsamples subjected to microscopic examination, we identified 673 tintinnid cells (465 in Villefranche and 208 in the Ionian Sea). From plankton subsamples devoted to molecular analyses, after DNA extraction and PCR amplification, we generated partial 18S rDNA sequences of 400 clones (200 clones from each locality) using Sanger sequencing and a total of 76,206 pyrosequence reads (termed as "pyroreads") for both samples and molecular markers. Quality and phylogenetic filtering of these pyroreads yielded 50,817 sequences (10,504 and 8,734 18S rDNA reads and 17,378 and 14,201 ITS reads for Villefranche and the Ionian Sea, respectively). To allow direct comparison between the two fragments targeted by pyrosequencing (18S rDNA and ITS) within each sample, we limited the pyroread datasets to the same size by random selection of sequences within each set (i.e., 10,504 18S rDNA and ITS reads for VilleFr43, and 8,734 18S rDNA and ITS reads for Ioni7, see Table 1).

In the coastal sample VilleFr-43, the tintinnid community encompassed 21 morphospecies, largely dominated by the family Tintinnidae, with *Salpingella faurei* and *Steenstrupiella steenstrupii* representing 54% of the total number of cells (Figure 1). The corresponding 18S rDNA clone library was also dominated by Tintinnidae (37.3% of clones), followed by Undellidae (26%) and Codonellidae (18%). Similarly, pyroreads affiliating to Tintinnidae were the most abundant in both the 18S rDNA and ITS datasets, where they represented 64.6% and 43.2% of pyroreads, respectively. In the offshore sample Ioni-7, a more diverse community with 36 morphospecies was uncovered and the number of cells identified as belonging to the families Tintinnidae (23%), Undellidae (22%) and Eutintinnidae (18%) co-dominated the tintinnid community. By contrast, the proportions of tintinnid families in the clone library was different from the morphological observations, with a majority of 18S rDNA sequences attributed to Undellidae (47%), followed by Tintinnidae (22%) and Codonellidae and Codonellopsidae (12% each). Pyrosequencing of the 18S rDNA region revealed a majority of reads identified as Tintinnidae (51%) and Undellidae (32%) while the ITS region inversely revealed a majority of Undellidae (51%) and then Tintinnidae (16%) (Figure 1). Overall, the distribution of 18S rDNA and ITS pyroreads related well with that of clone sequences and with the diversity of tintinnids based on microscopic observations at the family level. However, 18S rDNA and ITS pyroread datasets differed significantly from each other in Ligurian coastal and Ionian off-shore samples (χ^2 test, p-value < 0.001), showing the difficulties in comparing diversity estimates when different markers are used to assess diversity. As the PCR-based methods can be biased, it is likely that both the 18S rDNA and ITS regions could recover somewhat different relative proportions.

The microscopy-based determination of species diversity showed that Ligurian and Ionian samples were inhabited by diverse tintinnid communities. The differences observed between the two samples were probably due to the different sampling locations and periods, separated by more than 1,000 kilometers and collected at one-year interval. In particular, despite the number of cells was twice lower in the Ioni-7 sample than in the VilleFr-43 sample, the Ionian sample harboured a more diverse community than that of the VilleFr-43 sample (Supplementary Table S1). Moreover, the coastal sample was dominated by a few forms, with >50% of the tintinnid community composed of two species while the Ioni-7 sample showed a more 'equitable' species distribution with five species which reached just 42% of the total community. These differences in tintinnid communities in terms of cell abundances and

species compositions conformed to expectations based on previous studies (e.g., Dolan et al 2002, Dolan et al 2006). Thus, the differences of tintinnid communities in term of richness and species dominance provided us with two different species distribution models to study the performance of the different molecular approaches to assess OTU composition.

Species richness estimation

To estimate species richness from sequence data, we defined tintinnid OTUs applying a sequence similarity threshold of 99%, which was based on the 18S rDNA nucleotide differences observed between distinct tintinnid morphospecies (Bachy et al 2012, and Supplementary Figure S1). A lower value of sequence similarity would entail the mixing not only of different species, but even of different genera, such as *Dictyocysta* and *Codonella*, *Amphorella* and *Amphorellopsis*, or *Tintinnopsis* and *Climacocylis*. Moreover, considering both full-length 18S rDNA sequences or exclusively the short ~340 bp fragment targeted by our pyrosequencing analysis, the use of $\geq 99\%$ sequence identity provided the best species-level resolution. In fact, applying this threshold on the Sanger sequences from the clone libraries, we retrieved comparable OTU numbers for the complete 18S rDNA sequences and for the fragments corresponding to the region chosen for pyrosequencing (25 and 22 OTUs in VilleFr-43 and 17 and 22 OTUs in Ioni-7, respectively). The rarefaction curves corresponding to these clone libraries were close to saturation (Figure S2). In the case of VilleFr-43, not only the number of OTUs but also the species abundance distributions (with 32-40% of rare OTUs) were quite close to the ones described by morphological identification. By contrast, the OTUs identified in Ioni-7 represented only ~50% of the 36 species identified by morphological examination. However, the percentage of rare OTUs ($\leq 1\%$ of the total number of sequences in a dataset) in this sample ranged from 35% to 52% and was similar to the 42% observed by morphological identification.

A variety of approaches have been developed to analyse pyroreads to assess species richness (see review in Zinger et al 2012). We applied several of these methods in order to compare the results among them and with the classical 18S rDNA clone libraries / Sanger sequencing and traditional morphological analysis (Table 2 and Figure S3). As explained above, we first defined an adequate sequence similarity threshold (99%) to identify OTUs in

the 18S rDNA pyroread dataset. According to pairwise sequence comparisons of reference sequences belonging to identified species and sequences from clone libraries, the 18S rDNA and ITS regions targeted by our pyrosequencing analysis diverged with similar rates, comparable to those of complete 18S rDNA sequences (Supplementary Figure S1 and S4). Therefore, we used the 99% sequence identity threshold to identify OTUs with both the 18S rDNA and ITS pyroreads. Nevertheless, we have shown in previous work that phylogenetically distinct lineages may contain identical ITS region sequences (Bachy et al 2012), so it should be noted that ITS pyroreads were not as accurate as the 18S rDNA pyroreads in discriminating different species (see Supplementary Figure S1). It may be that the ITS region we selected for pyrosequencing amplification is subjected to a greater selective constraint than the targeted 18S rDNA variable region V4, which is characterized by particularly rapid rates of evolution within eukaryotes. With the 99% sequence identity threshold established, we tested several sequence alignment, clustering and denoising methods on the 18S rDNA or ITS pyroread datasets (see Materials and Methods). Surprisingly, depending on the analytical method and molecular marker used, 18S rDNA or ITS, the number of OTUs varied from only 7 up to 1,848. Accordingly, the proportion of rare OTUs represented by few reads showed extreme variations, between 14% and 99%. Table 2 shows that the number of OTUs detected using the common method of multiple sequence alignment (MSA) or the more stringent pairwise alignment (PW) and average-linkage exceeded the number of expected species (based on unambiguous morphological taxon identification) by one or two orders of magnitude. The greedy algorithm for clustering of representative sequences implemented in Usearch (Edgar 2010) appeared to be more efficient, but the number of OTUs still exceeded the number of expected species by an order of magnitude. We could reject these diversity estimates with confidence, not only because they overestimated the number of species observed under the microscope but also because they exceeded, sometimes by two-fold or more, the total number of tintinnid cells observed in twin volumes of samples.

To avoid this kind of artificial inflation, we applied several approaches recently developed to remove PCR and pyrosequencing errors, including sequence preclustering and the more complex pipelines implemented in CD-HIT-OTU (Wu et al 2011) and AmpliconNoise (Quince et al 2011). The agglomerative approach applying single-linkage preclustering (SLP) before the classical clustering step also overestimated species richness, especially for the 18S

rDNA pyroreads. This could be due to the fact that the preclustering step at 98% sequence identity usually used for SLP (Huse et al 2010) was too large compared to the threshold that we empirically determined for tintinnids (Table 2). Nevertheless, applying a preclustering step with a 99.5% similarity threshold was not sufficient to denoise the data efficiently (between 166-1848 OTUs were defined depending on the marker, Table 2). Compared to the SLP approach, the denoising and clustering procedures implemented in CD-HIT-OTU and AmpliconNoise were the only ones able to produce estimates coherent with those obtained from classical clone library / Sanger sequencing and especially with morphological taxonomic identification. These methods inferred a number of OTUs of ~25 for the 18S rDNA region and ~20 for the ITS region in VilleFr-43 and 12-23 for the 18S rDNA region and 7-18 for the ITS region in Ioni-7. Since AmpliconNoise produced reasonable species estimates by taking into account all sequences, we kept the OTUs identified from 18S rDNA pyroreads by this method for subsequent analyses.

Phylogenetic analyses

We built a phylogenetic tree based on sequences representative of all OTUs identified in clone libraries and pyrosequencing data together with their closest described relatives available in public databases (Figure 2). The resulting tree was strongly supported at the genera-level, giving us the opportunity to compare at this fine taxonomic rank the abundance patterns generated by both molecular approaches and the morphological analysis. From a qualitative point of view, both the cloning and pyrosequencing approaches detected the majority of families and genera observed by morphology, although a slightly larger diversity was detected by cloning / Sanger sequencing than by pyrosequencing. Surprisingly, cloning and pyrosequencing approaches seemed thus to be almost equally efficient for detecting the species observed in our samples (*Salpingella*-, *Undella*-, *Codonella*- and *Dictyocysta*-related species), even the ones present in low abundance (e.g., those of families Xytonellidae or Codonellopsidae). Moreover, whereas no *Tintinnidium* spp. was observed in the coastal sample (VilleFr-43) under the microscope, 18S rDNA sequences affiliated to this genus were detected (4% of pyroreads and 0.5% of clones). By contrast, *Rhabdonella* sp. and *Steenstrupiella* sp. specimens observed in the sample Ioni-7 were not detected with any of the molecular approaches. The comparison between both molecular approaches revealed that

more than 50% of all OTUs were detected by both methods (16 of the 27 OTUs in the coastal sample VilleFr-43 and 11 of the 17 OTUs in the offshore sample Ioni-7, see Figure 2). The dissimilarities between the two methods concerned relatively rare OTUs, such as those affiliated to *Tintinnopsis* sp. (OTU 24, 5 clones in VilleFr-43) or *Codonella aspera* (OTU 36, 3 clones in VilleFr-43), undetected by pyrosequencing, or three OTUs within the *Eutintinnus* clade (OTU 14, 15 and 18, respectively 37, 22 and 17 18S rDNA pyrosequence reads) undetected in clone libraries. Part of the diversity found only by pyrosequencing corresponded to OTUs represented by singletons or doubletons (OTU 7, 19 and 21).

From a semi-quantitative point of view, the species found to be the most abundant by microscopic examination were also the ones preferentially detected by the molecular approaches. Thus, in the VilleFr-43 sample, OTUs 6, 8 and 27, which were dominant in the clone and pyroread libraries, were related to *Salpingella* sp., *Steenstrupiella steenstrupi* and *Undella* sp. respectively, and found to dominate the microscopically-determined tintinnid community (Figure 2). In the Ioni-7 sample, the *Undella*-related OTU 27 was the dominant in the clone and pyroread datasets and was also the dominant genus based on microscopic analysis (22.6% of cells). However, there were some discrepancies comparing datasets. *Eutintinnus* spp. specimens, which were abundant based on microscopic observations, especially in Ioni-7, were detected only in low abundances by the molecular approaches. In contrast, Codonellopsidae appeared overrepresented in the molecular results because only a few *Codonellopsis* sp. and *Stenosemella* sp. specimens were encountered under the microscope (Figure 1).

Discussion

Pyrosequencing versus classical clone libraries / Sanger sequencing

Pyrosequencing and the classical cloning method revealed a relatively similar phylogenetic structure of the tintinnid community at the species level. These observations resembled other molecular comparative studies of whole eukaryotic communities targeted with eukaryotic universal primers (Edgcomb et al 2011) or with particular groups, such as fungi, targeted with specific primers (Tedersoo et al 2010). In the case of our tintinnid communities, although the presence and proportion of the dominant lineages were similar for both molecular methods,

the OTUs unique to each method corresponded to rare lineages. These discrepancies in low abundant OTUs might be induced by the relative low quantity of DNA coming from rare species in the extracted DNA samples, and so, randomly amplified or not amplified during PCR steps. Some discrepancies could also be due to the very different length of clone sequences and pyroreads, and consequently to a different amount of phylogenetic information, leading to a different number of potentially detectable OTUs. The fact that differences in amplicon length could notably influence estimates of richness has been shown with amplicon libraries constructed for bacteria (Engelbrektson et al 2010, Huber et al 2009).

97.7% of clone sequences and 81.5% of pyroreads (18S rDNA and ITS together) could be reliably assigned to known tintinnid taxa at the genus/species level ($\geq 98\%$ sequence identity), and still a majority of them (85% of clones and 69.5% of pyroreads) could be related with strong confidence to the particular species ($\geq 99\%$ sequence identity). In fact, no novel distinct clades without described representatives were detected despite deep sequence coverage, especially by the pyrosequencing approach. The lack of novel clades could be explained by recent efforts of species genetic characterization in the Mediterranean area (Bachy et al 2012). However, for certain genera such as *Salpingella*, some lineages (OTUs 2, 5 and 6) were only represented by environmental sequences, which argues for the need for further taxonomic work to genotype more tintinnid species.

Molecular versus morphological identification

In contrast with previous studies addressing the accuracy of pyroreads and/or clone libraries to describe microbial diversity in natural samples, we targeted a particular protist group with conspicuous morphological traits that allowed us obtaining a precise description of the community using morphological inspection in parallel with molecular techniques. This provided us not only with a maximum diversity limit (no more OTUs than cells could exist in the samples), but also with a control reference for the expected number of OTUs and their relative abundances. In fact, the accurate characterization of microbial communities requires not only taxonomic identification but also measurement of relative abundances. OTU frequencies (cloning- or pyrosequencing-derived) are often used in molecular studies of microbial communities to infer the relative ecological relationships among taxa (e.g.,

dominance or evenness) and so, to extrapolate the total number of species in an environment (Hughes et al 2001).

In our study, the sequence frequency distribution for both clone and pyroread sequences did not fit perfectly the frequency distribution of tintinnid morphospecies encountered under the microscope. The most noticeable divergence in frequencies between morphology and sequencing data concerned *Eutintinnus* and *Tintinnidium* species and showed opposite trends. *Eutintinnus* spp. constituted the second and third more abundant morphotype in Ioni-7 and VilleFr-43, respectively, but was poorly detected by cloning and pyrosequencing (for both 18S rDNA and ITS reads, see Figure 2). In contrast, *Tintinnidium* was found in sequence libraries but not observed under the microscope. In both cases, there is no obvious explanation for the disagreement between the morphological and molecular estimations. In the case of *Tintinnidium*, with its amorphous lorica which can be mistaken for detrital material, it is possible that it was overlooked in microscopic examinations if it was present in low numbers. In addition, differences in rRNA gene copy number among taxa probably exist, which is often cited as the major reason of incongruence between environmental sequence data and morphology (Medinger et al 2010). Among tintinnids, the number of rRNA gene repeats probably varies, in particular because species may contain different number of nuclei (Agatha and Strüder-Kypke 2007). In addition, in the case of the less abundant species, our use of two different subsamples for each locality may explain some differences between morphological- and molecular-based estimations, as those very rare species could have been unequally distributed when the initial water samples were divided into two subsamples.

Influence of marker choice and computational analyses on diversity estimates

Identifying and counting each tintinnid cell in our samples allowed us to unambiguously compare their species richness and distribution between different analyses. As previously found by Doherty and colleagues (2007) for planktonic ciliates, we found that the number of recognizable tintinnid species was roughly similar to the number of OTUs detected in clone libraries. However, pyrosequencing data, and the multiple approaches existing to analyze them, yielded much more contradictory results. For our samples, the number of OTUs ($\geq 99\%$ identity) increased by a factor up to 100 depending on the *in silico* approach applied. The

multiple sequence alignment and average clustering grossly overestimated the number of OTUs. For the other methods tested (pairwise alignment, greedy algorithm for clustering of representative sequences and single-linkage preclustering), the number of inferred OTUs did not approach the levels estimated by morphology until we went down to a 95% identity threshold for OTU definition (Figure S3). A 95% threshold is clearly not realistic since we know from described tintinnid morphospecies that members of two different species may have 99% identical sequences (Figure S1 and Santoferrara et al 2012). Only the denoising algorithms based on sequence clustering approaches implemented in AmpliconNoise (Quince et al 2011) and CD-HIT-OTU (Wu et al 2011) were able to correctly estimate OTU diversities at 99% sequence identity to yield values comparable with those observed by morphological study and clone library analysis.

While the denoising methods have been demonstrated to minimize the number of spurious OTUs (Quince et al 2011), they tended to underestimate to some extent the species richness in the Ionian sample (Ioni-7). Compared with the 36 morphospecies found in this sample, the estimates obtained from the clone libraries (17-22 OTUs), 18S rDNA pyroreads (12-23 OTUs) and ITS pyroreads (7-18 OTUs) were low. The distinction as different species of what might be just morphological varieties could explain, at least partially, why we found a reduced diversity in molecular approaches. For instance, *Undella* spp. encompassed six morphotypes in the sample (Supplementary Table S1) while it is known that there is very low genetic divergence in this genus for the markers we studied (Bachy et al 2012). Moreover, the relatively small power of the ITS region selected for pyrosequencing to discriminate closely related species could explain the even lower diversity in found in the ITS pyroread dataset.

In addition to potential species identification problems, we realized in our study that, depending on the method applied to analyze pyrosequencing data, the percentage of rare OTUs ($\leq 1\%$ of the total number of sequences in a library) could lead to very different interpretations of species distributions, in particular the length of the tail of the rank abundance curve. As shown above, only the values estimated from clone libraries and from pyrosequencing data analyzed with AmpliconNoise or CD-HIT-OTU were reasonable according to the morphological observations. Rare species represented 42.9% and 41.7% of the cells inspected by morphological analysis in VilleFr-43 and Ioni-7, respectively. Whereas the part of rare OTUs detected with AmpliconNoise and CD-HIT-OTU accounted for 36.4-

55.6% and 33.3-56.5% of all OTUs in VilleFr-43 and Ioni-7, respectively, the other methods inferred a proportion ranging from 94% to 99.2%, grossly overestimated (Figure 1).

It is well-known that PCR and sequencing errors may inflate diversity estimates by creating false taxa (Kunin et al 2010, Quince et al 2009, Reeder and Knight 2009). It appears to be less well-appreciated that the methodological approach selected to analyze the pyrosequencing data is also crucial to avoid the artificial inflation of diversity estimates. The use of a morphological reference allowed us to show that certain methods led to massive overestimation of OTU numbers, in particular for rare OTUs, not only because of sequence errors but also by other methodological problems. Notably, misaligned sequences in multiple sequence alignments (MSA) can be considered as different OTUs. The size and diversity of the "rare biosphere" may be somewhat less than what has been so far inferred by the first pyrosequencing analyses of microbial diversity.

Acknowledgements

We are thankful to L. Gasperini and G. Bortoluzzi of the Istituto di Geologia Marina (ISMAR), CNR, Bologna (Italy) for allowing P.L.G. to participate in the Marmara2010 cruise. We kindly thank the officers and the crew as well as the scientific team on-board of the RV Urania for operating the CTD rosette and providing helpful assistance. We acknowledge financial support from the French CNRS and the ANR programs Biodiversity (ANR BDIV 07 004-02 'Aquaparadox') and Génomique microbienne à grande échelle (ANR-08-GENM-024-001, EVOLDEEP).

References

Agatha S, Strüder-Kypke MC (2007). Phylogeny of the order Choreotrichida (Ciliophora, Spirotricha, Oligotrichea) as inferred from morphology, ultrastructure, ontogenesis, and SSrRNA gene sequences. *Eur J Protistol* 43: 37-63.

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

Bachy C, López-García P, Vereshchaka A, Moreira D (2011). Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the North Pole at the end of the polar night. *Front Microbiol*.

Bachy C, Gómez F, López-García P, Dolan JR, Moreira D. Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora). *Protist*; e-pub ahead of print 8 February 2012, doi: 10.1016/j.protis.2012.01.001.

Bik HM, Sung WAY, De Ley P, Baldwin JG, Sharma J, Rocha-Olivares A *et al* (2012). Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Mol Ecol* 21: 1048-1059.

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

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD *et al* (2009). Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* 75: 5797-5808.

Doherty M, Costas BA, McManus GB, Katz LA (2007). Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. *Aquat Microb Ecol* 48: 141-154.

Dolan JR, Claustre H, Carlotti F, Plounevez S, Moutin T (2002). Microzooplankton diversity: relationships of tintinnid ciliates with resources, competitors and predators from the Atlantic Coast of Morocco to the Eastern Mediterranean. *Deep-Sea Research part I* 49: 1217-1232.

Dolan JR, Lemée R, Gasparini S, Mousseau L, Heyndrickx C (2006). Probing diversity in the plankton: using patterns in tintinnids (planktonic marine ciliates) to identify patterns. *Hydrobiologia* 555: 143-157.

Dolan JR, Ritchie ME, Tunin-Ley A, Pizay M-D (2009). Dynamics of core and occasional species in the marine plankton: tintinnid ciliates in the north-west Mediterranean Sea. *J Biogeogr* 36: 887-895.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*.

Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C *et al* (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J*.

Engelbrekton A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H *et al* (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* 4: 642-647.

Gihring TM, Green SJ, Schadt CW (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ Microbiol* 14: 285-290.

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

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G *et al* (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21: 494-504.

Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Mark Welch DB (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* 11: 1292-1302.

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001). Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. *Appl Environ Microbiol* 67: 4399-4406.

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

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

Jobb G, von Haeseler A, Strimmer K (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4: 18.

Katoh K, Misawa K, Kuma K, Miyata T (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059-3066.

Kofoed CA, Campbell AS (1929). *A conspectus of the marine and freshwater ciliata belonging to the suborder Tintinnoinea, with descriptions of new species principally from the Agassiz expedition to the Eastern Tropical Pacific 1904–1905*, vol. 34. Univ Calif Publ Zool.

Kofoed CA, Campbell AS (1939). *The Ciliata: the Tintinnoinea*, vol. 84. Bull Mus Comp Zool Harvard.

Kunin V, Engelbrektsen A, Ochman H, Hugenholtz P (2010). Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. *Environ Microbiol* 12: 118-123.

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

López-García P, Philippe H, Gail F, Moreira D (2003). Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci U S A* 100: 697-702.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.

Marshall SM (1969). Protozoa, order Tintinnia. *Conseil International pour l'exploration de la Mer, Fiches d'Identification de Zooplancton: fiches 117-127.*

Massana R, Pedrós-Alió C (2008). Unveiling new microbial eukaryotes in the surface ocean. *Curr Opin Microbiol* 11: 213-218.

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwalder B, Schlotterer C *et al* (2010). Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol Ecol* 19: 32-40.

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.

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

Not F, Valentin K, Romari K, Lovejoy C, Massana R, Tobe K *et al* (2007b). Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* 315: 253-255.

Pedrós-Alió C (2007). Dipping into the Rare Biosphere. *Science* 315: 192-193.

Philippe H (1993). MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* 21: 5264-5272.

Pierce RW, Turner JT (1993). Global biogeography of marine tintinnids. *Mar Ecol Prog Ser* 94: 11-26.

Price MN, Dehal PS, Arkin AP (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol Biol Evol* 26: 1641-1650.

Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM *et al* (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Meth* 6: 639-641.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 12: 38.

Reeder J, Knight R (2009). The 'rare biosphere': a reality check. *Nat Meth* 6: 636-637.

Saccà A, Strüder-Kypke MC, Lynn DH. Redescription of *Rhizodomus tagatzi* (Ciliophora: Spirotrichea: Tintinnida), Based on Morphology and Small Subunit Ribosomal RNA Gene Sequence. *J Eukaryot Microbiol*; e-pub ahead of print 27 March 2012, doi: 10.1111/j.1550-7408.2012.00615.x.

Santoferrara LF, McManus GB, Alder VA. Utility of Genetic Markers and Morphology for Species Discrimination within the Order Tintinnida (Ciliophora, Spirotrichea). *Protist*; e-pub ahead of print 20 January 2012, doi: 10.1016/j.protis.2011.12.002.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* 75: 7537-7541.

Schloss PD (2010). The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. *PLoS Comput Biol* 6: e1000844.

Snoeyenbos-West OLO, Salcedo T, McManus GB, Katz LA (2002). Insights into the diversity of choreotrich and oligotrich ciliates (Class: Spirotrichea) based on genealogical analyses of multiple loci. *Int J Syst Evol Microbiol* 52: 1901-1913.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* 103: 12115 - 12120.

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

Strüder-Kypke MC, Lynn DH (2008). Morphological versus molecular data - Phylogeny of tintinnid ciliates (Ciliophora, Choreotrichia) inferred from small subunit rRNA gene sequences. *Denisia* 23: 417-424.

Sun Y, Cai Y, Liu L, Yu F, Farrell ML, McKendree W *et al* (2009). ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res* 37: e76-e76.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I *et al* (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol* 188: 291-301.

Wu S, Zhu Z, Fu L, Niu B, Li W (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12: 444.

Zinger L, Gobet A, Pommier T (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Mol Ecol* 21: 1878-1896.

Table 1 Number of tintinnid cells, clone sequences and pyroreads retained for comparison from the VilleFr43 and Ioni7 samples

	VilleFr43	Ioni7
Cells	465	208
Clone sequences	200	200
18S rDNA pyroreads	10504	8734
ITS pyroreads	10504	8734

Table 2 Number of tintinnid species and total and rare OTUs identified using morphological and molecular methods

	VilleFr43		Ioni7	
	Total	Rare	Total	Rare
Cell identification				
Morphotypes	21	9 (42.9%)	36	15 (41.7%)
Clone sequences				
Complete sequences	25	10 (40.0%)	17	6 (35.3%)
Partial "pyroread" region	22	7 (31.8%)	22	11 (50.0%)
18S rDNA pyroreads				
MSA	1494	1474 (98.7%)	1089	1073 (98.5%)
PW	1159	1144 (98.7%)	854	840 (98.4%)
Usearch	518	498 (96.1%)	400	375 (93.8%)
SLP-PW	1848	1833 (99.2%)	1383	1372 (99.2%)
CD-HIT-OTU	27	15 (55.6%)	23	13 (56.5%)
AmpliconNoise	22	8 (36.4%)	12	4 (33.3%)
ITS pyroreads				
MSA	1036	1022 (98.6%)	663	653 (98.5%)
PW	832	818 (98.3%)	519	510 (98.3%)
Usearch	544	525 (96.5%)	259	251 (96.9%)
SLP-PW	166	156 (94.0%)	616	608 (98.7%)
CD-HIT-OTU	23	15 (65.2%)	18	12 (66.7%)
AmpliconNoise	19	11 (57.9%)	7	1 (14.3%)

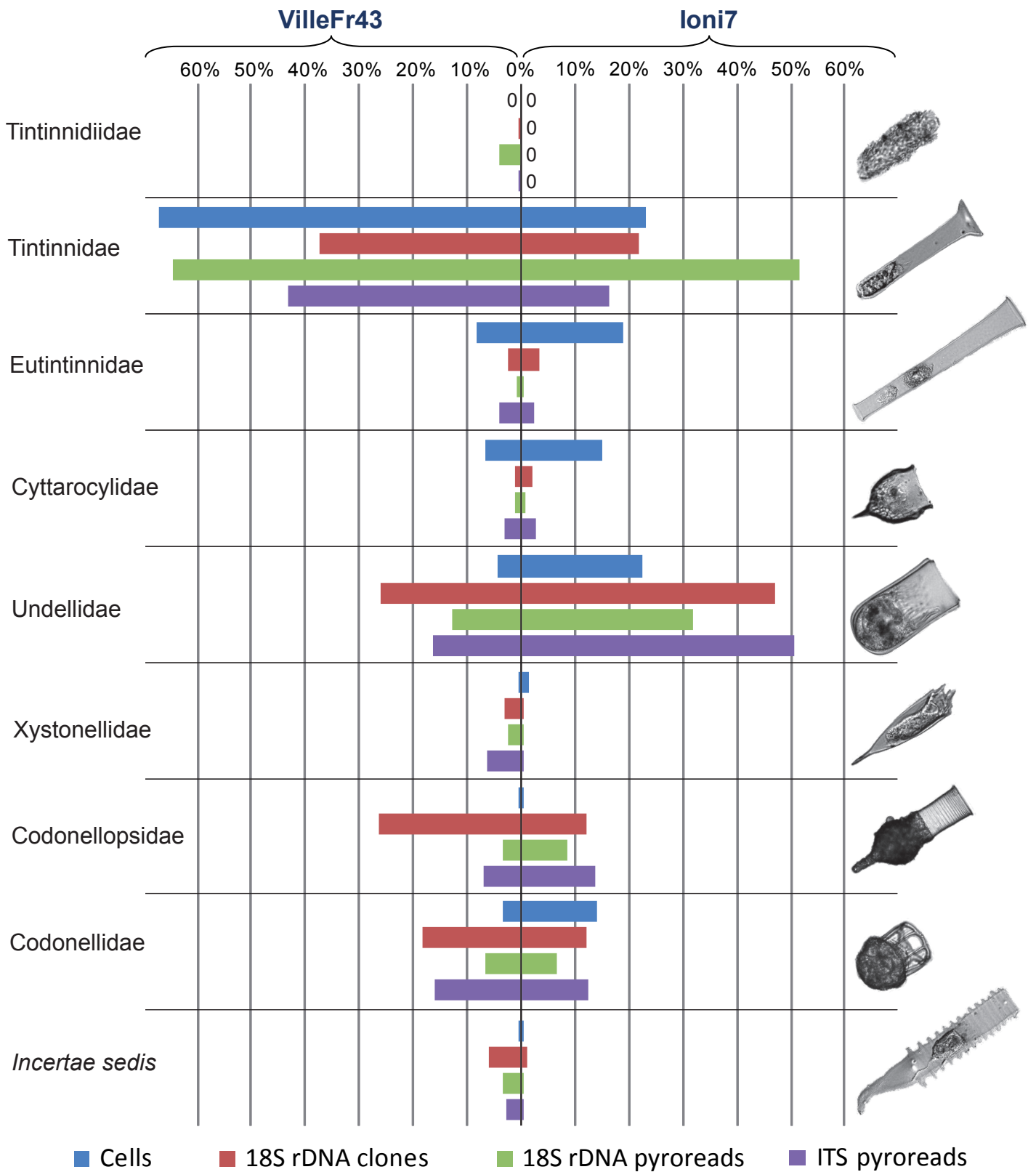
Abbreviations: MSA, multiple sequence alignment; PW, pairwise alignment; SLP-PW, single-linkage preclustering and pairwise alignment.

For rare OTUs, total numbers and relative proportions (%) are indicated.

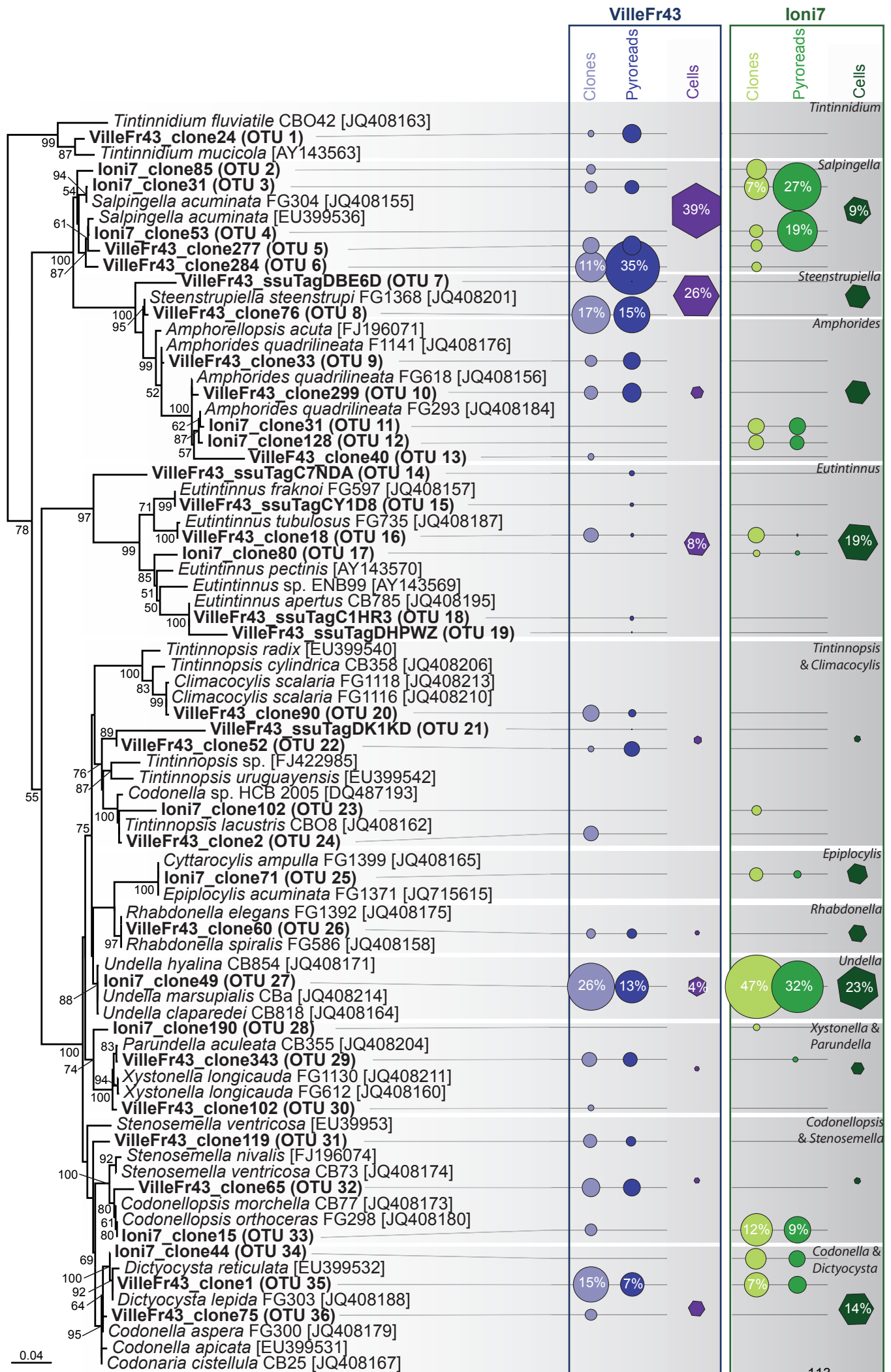
Figure legends

Figure 1- Frequencies of the different tintinnid families found in the VilleFr-43 and Ioni-7 samples by morphology (cells), 18S rDNA library sequencing (clones) and pyrosequencing of 18S rDNA and ITS regions. Sequences were assigned to families by BLAST searches against a curated tintinnid sequence database. Micrographs on the right show representative morphologies encountered in each family. Complete absence of a family in a sample is indicated by 0.

Figure 2- Maximum likelihood phylogenetic tree of 18S rDNA sequences of detected OTUs and their closest tintinnid relatives, based on 1,200 aligned positions. Sequences from this study are in bold. Relative proportions of the different OTUs in each sample estimated from clone libraries and 18S rDNA pyroreads are indicated by circles of proportional size on the right. Relative proportions of the different tintinnid species estimated from cell counts under the microscope are indicated by hexagons of proportional size on the right. The numbers at nodes are bootstrap values (values under 50% are omitted). Accession numbers are provided between brackets. The scale bar represents the number of substitutions for a unit branch length.



Bachy et al., Figure 1



Bachy et al., Figure 2

Table S1- Species list and abundances of tintinnid cells observed in VilleFr43 and Ioni7 samples. Taxonomic classification follows clades based on tintinnid molecular phylogeny (Bachy et al. 2012).

VilleFr-43 (Villefranche Oct 2009)	Classification (Family/Genus)	# specimens
<i>Salpingella faurei</i>	Tintinnidae/Salpingella sp.	131
<i>Steenstrupiella steenstrupi</i>	Tintinnidae/Steenstrupiella sp.	122
<i>Acanthostomella conicoides</i>	undetermined	37
<i>Salpingella curta</i>	Tintinnidae/Salpingella sp.	36
<i>Dadayiella ganymedes</i>	Cyttarocylididae/undetermined	30
<i>Eutintinnus elongatus</i>	Eutintinnidae/Eutintinnus sp.	27
<i>Undella clevei</i>	Undellidae/Undella sp.	20
<i>Salpingella acuminata</i>	Tintinnidae/Salpingella sp.	13
<i>Dictyocysta mitra</i>	Codonellidae/Codonella and Dictyocysta spp.	11
<i>Amphorides quadrilineata</i>	Tintinnidae/Amphorides sp.	10
<i>Craterella tortulata</i>	undetermined	6
<i>Eutintinnus apertus</i>	Eutintinnidae/Eutintinnus sp.	5
<i>Dictyocysta lepida</i>	Codonellidae/Codonella and Dictyocysta spp.	4
<i>Eutintinnus fraknoi</i>	Eutintinnidae/Eutintinnus sp.	3
<i>Eutintinnus lusundae</i>	Eutintinnidae/Eutintinnus sp.	3
<i>Codonellopsis orthocercas</i>	Codonellopsidae/Codonellopsis and Stenosemella spp.	2
<i>Protorhabdonella curta</i>	Cyttarocylididae/Rhabdonella sp.	1
<i>Tintinnopsis</i> sp.	<i>incertae sedis</i> /Tintinnopsis and Climacocyliis	1
<i>Dictyocysta entzi</i>	Codonellidae/Codonella and Dictyocysta spp.	1
<i>Xystonella clavata</i>	Xystonellidae/Xystonella and Parundella	1
<i>Climacocyliis scalaria</i>	<i>incertae sedis</i> /Tintinnopsis and Climacocyliis	1
Ioni-7 (Ionian Sea Oct 2010)	Classification (Family/Genus)	# specimens
<i>Eutintinnus macilintus</i>	Eutintinnidae/Eutintinnus sp.	31
<i>Undella clevei</i>	Undellidae/Undella sp.	18
<i>Dictyocysta elegans</i>	Codonellidae/Codonella and Dictyocysta spp.	16
<i>Amphorides quadrilineata</i>	Tintinnidae/Amphorides sp.	11
<i>Steenstrupiella steenstrupi</i>	Tintinnidae/Steenstrupiella sp.	11
<i>Dadayiella ganymedes</i>	Cyttarocylididae/undetermined	9
<i>Epiplocyliis undella</i>	Cyttarocylididae/Epiplocyliis sp.	9
<i>Protohabdonella curta</i>	Rhabdonella spp.	9
<i>Salpingella attenuata</i>	Tintinnidae/Salpingella sp.	9
<i>Dictyocysta mitra</i>	Codonellidae/Codonella and Dictyocysta spp.	8
<i>Proplectella</i> sp ØLOD37	Undellidae/Undella sp.	8
<i>Salpingacantha ampla</i>	Tintinnidae/Salpingella sp.	8
<i>Undellopsis marsupialis</i>	Undellidae/Undella sp.	8
<i>Proplectella fastigata</i>	Undellidae/Undella sp.	7
<i>Acanthostomella conicoides</i>	undetermined	6
<i>Eutintinnus fraknoi</i>	Eutintinnidae/Eutintinnus sp.	6
<i>Amphorides laackmanni</i>	Tintinnidae/Amphorides sp.	3

<i>Codonella aspera</i>	Codonellidae/Codonella and Dictyocysta spp.	3
<i>Proplectella cuspidata</i>	Undellidae/Undella sp.	3
<i>Steenstrupiella intum</i>	Tintinnidae/Steenstrupiella sp.	3
<i>Undella subacuta</i>	Undellidae/Undella sp.	3
<i>Dadayiella pachytoeus</i>	Cyttarocylididae/undetermined	2
<i>Dictyocysta lepida</i>	Codonellidae/Codonella and Dictyocysta spp.	2
<i>Epiplocylis blanda</i>	Cyttarocylididae/Epiplocylis sp.	2
<i>Eutintinnus tubulosus</i>	Eutintinnidae/Eutintinnus sp.	2
<i>Acanthostomella obtusa</i>	undetermined	1
<i>Canthariella pyrimidata</i>	undetermined	1
<i>Climacocylis scalaria</i>	<i>incertae sedis</i> /Tintinnopsis and Climacocylis	1
<i>Codonellopsis orthocercas</i>	Codonellopsidae/Codonellopsis and Stenosemella spp.	1
<i>Craterella tortulata</i>	undetermined	1
<i>Craterella urceolata</i>	undetermined	1
<i>Parundella aculeata</i>	Xystonellidae/Xystonella and Parundella	1
<i>Salpingella accuminata</i>	Tintinnidae/Salpingella sp.	1
<i>Salpingella decurtata</i>	Tintinnidae/Salpingella sp.	1
<i>Xystonella lohmannii</i>	Xystonellidae/Xystonella and Parundella	1
<i>Xystonellopsis paradoxa</i>	Xystonellidae/Xystonella and Parundella	1

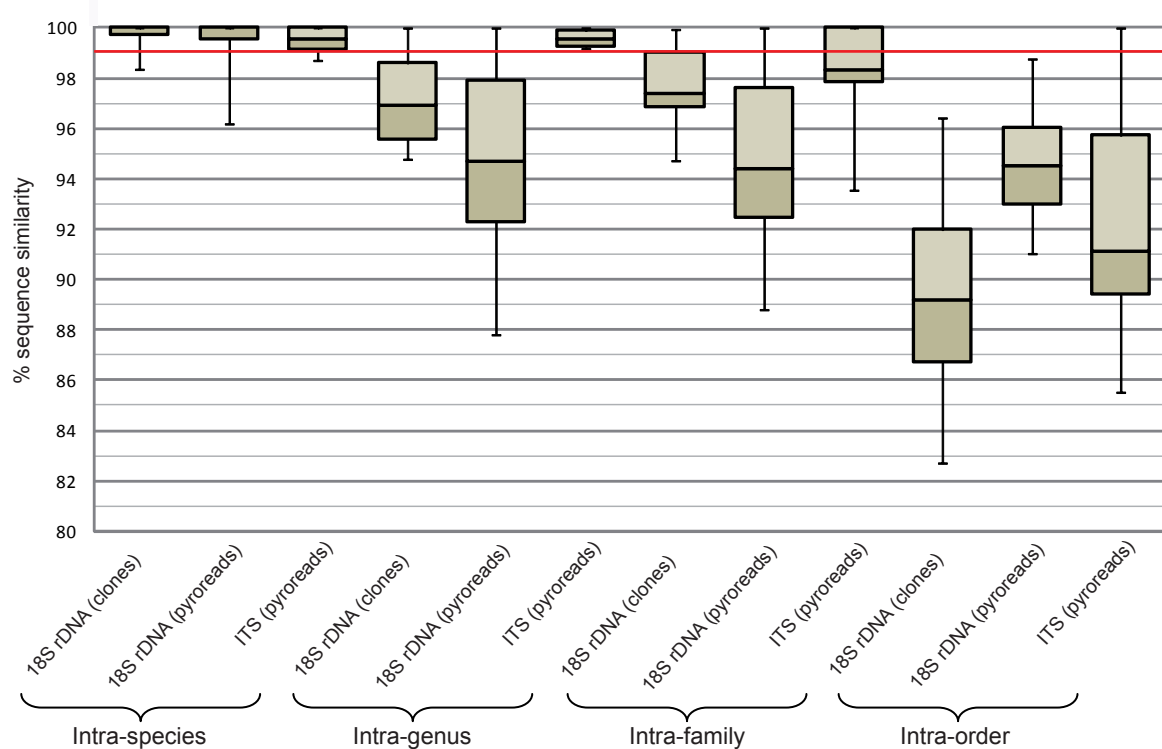


Figure S1 Box-Whisker plots (BWPs) depicting sequence similarity found in pairwise comparisons of 18S rDNA sequences of different length (indicated as Sanger and pyrotags) and ITS region sequences at different taxonomic ranks. Sequence similarities were calculated from 89 high-quality tintinnid 18S rRNA gene sequences publicly available in GenBank. Sequences were trimmed to include the region corresponding to nucleotide positions 384-1522 and 398-739 in *Amphorellopsis acuta* (accession number EU399530) to simulate “Sanger” and “pyrotag” sequences, respectively. An additional set of 23 tintinnid sequences covering the ITS regions was trimmed to include the nucleotide positions 66-298 in *Metacylis angulata* (accession number AF399077), corresponding to pairwise sequence comparisons of ITS pyrotags. The BWPs for intra-species comparisons summarizes 43 pairwise comparisons (5 for the ITS region), the intra-genus BWPs 204 comparisons (6 for ITS), the intra-family BWPs 160 comparisons (12 for ITS) and the intra-order BWPs 3236 comparisons (221 for ITS). Min-Max Whisker=observed sample minimum and maximum, Q1-Q3 box=lower and upper quartile, median=50th percentile (Q2). Sequences were classified following the molecular phylogeny-based taxonomy of Bachy et al. (2012).

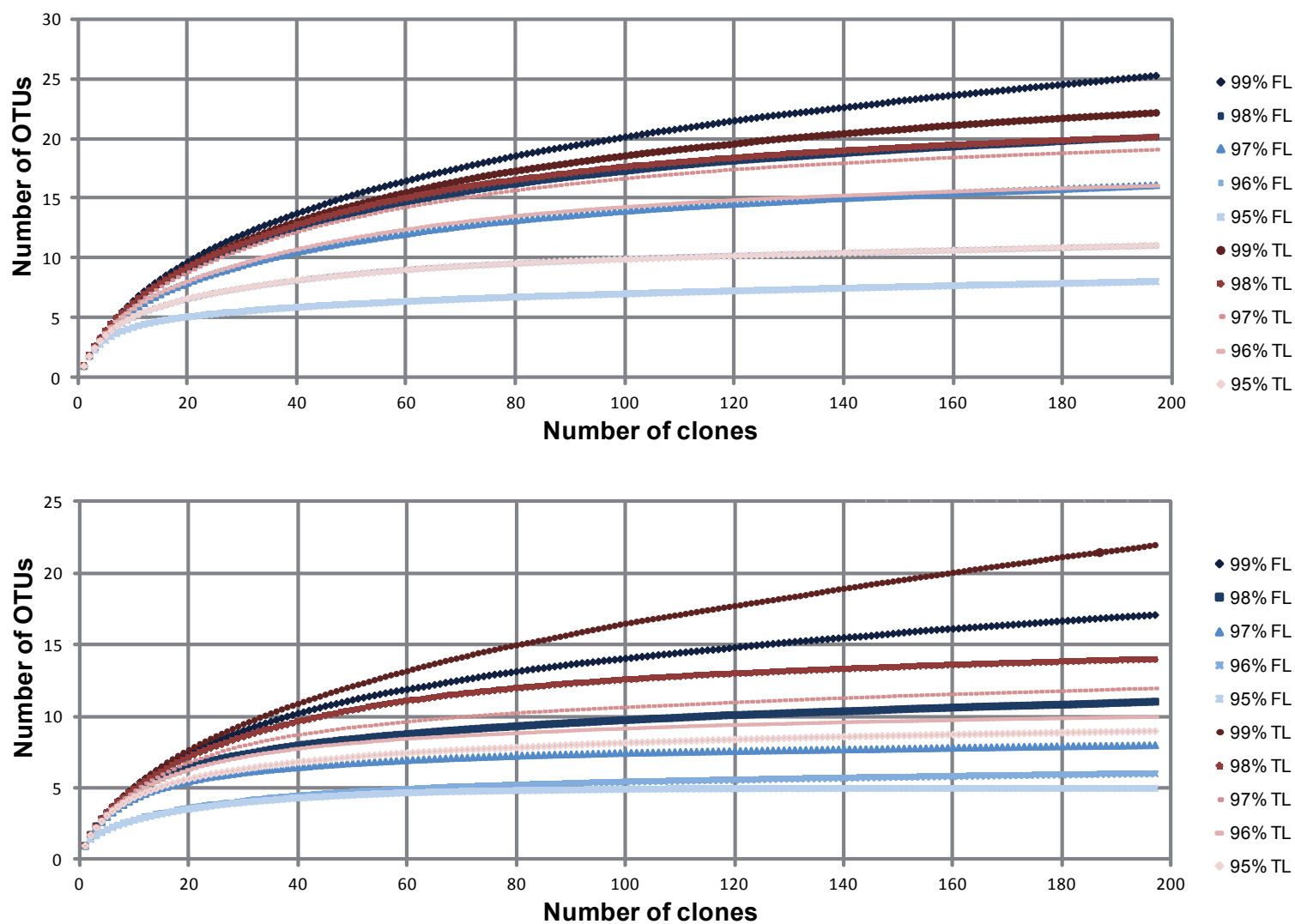


Figure S2 Rarefaction analyses of 18S rDNA clone library sequences at different levels of sequence similarity (from 95% to 99%). Analyses were carried out for the full length of the sequences and for a shorter fragment (~350 nt) corresponding to the region targeted for pyrosequencing. FL, full-length sequences; TL, pyroread-length sequences.

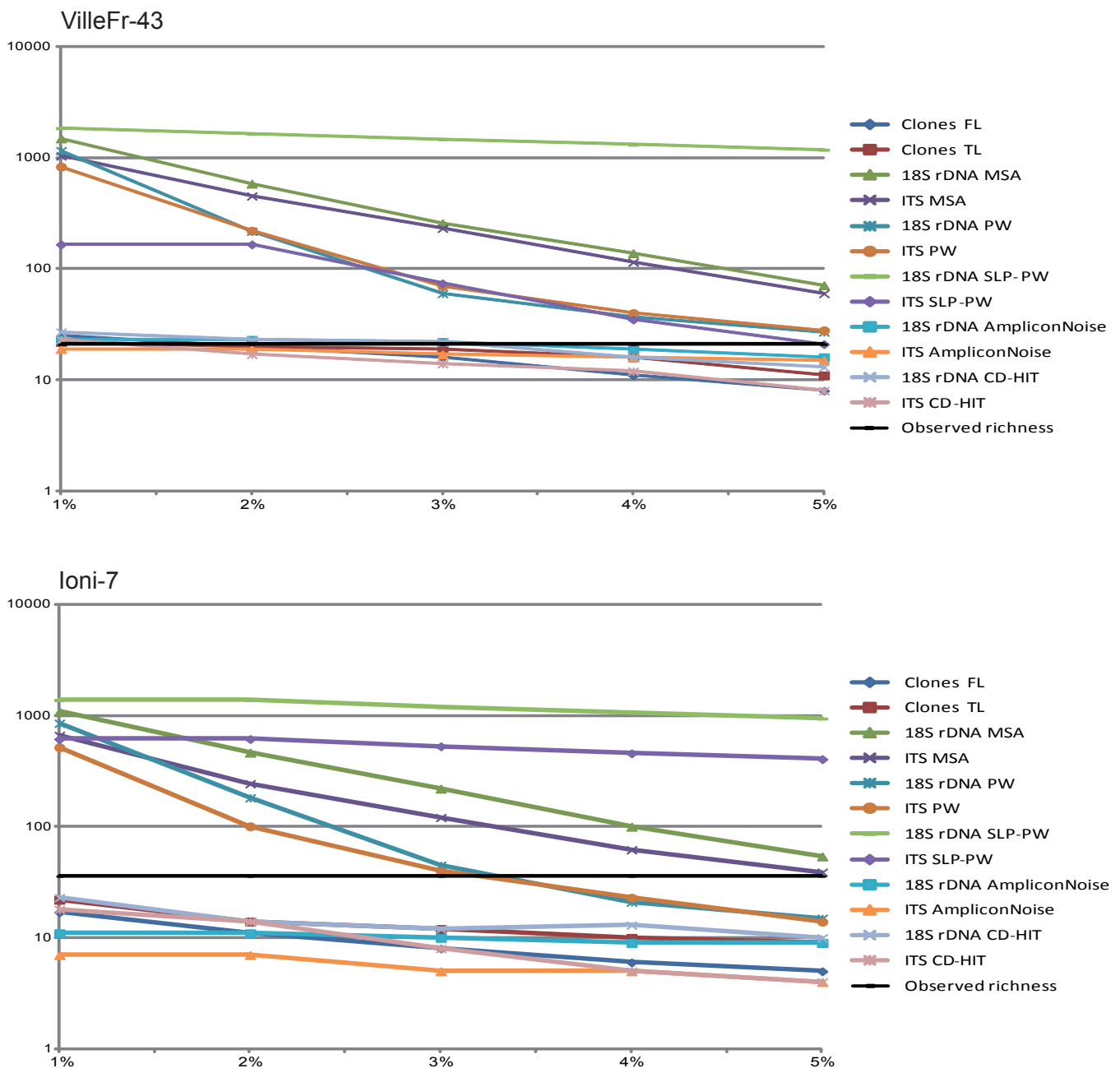


Figure S3 OTU numbers in the clone and pyroread libraries as a function of sequence divergence according to different analytical approaches. OTUs were counted at different percent of similarity (from 95% to 99.5%) after average linkage clustering. For clone library datasets, results are shown for both full-length 18S rDNA sequences (clones FL, blue diamonds) and only taking into account the region targeted for pyroread analysis (clones TL, red squares). For pyroread datasets, results are shown according to the different approaches used: multiple sequence alignment (MSA) for 18S rDNA (green triangles) and ITS (purple crosses) sequences, pairwise alignment (PW) for 18S rDNA (cyan crosses) and ITS (orange circles) sequences, pyrosequencing noise removal by SLP-PW for 18S rDNA (green sticks) and ITS (purple diamonds) sequences, pyrosequencing noise removal by CD-HIT for 18S rDNA (light blue crosses) and ITS (pink crosses) and finally, pyrosequencing noise and PCR errors removal by AmpliconNoise for 18S rDNA (cyan squares) and ITS (orange triangles). As a control reference, the number of species identified by morphology analysis is shown as a black line. The y-axis is scaled logarithmically.

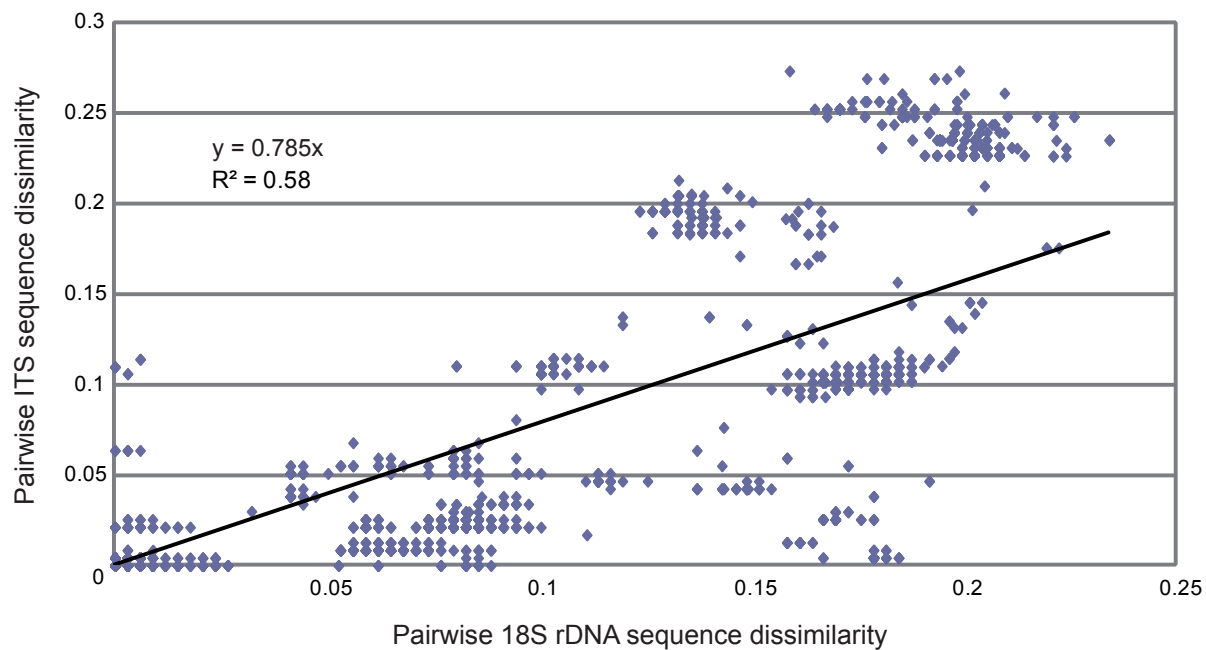


Figure S4 Comparison of sequence divergence between the 18S rDNA and ITS regions targeted for pyrosequencing. Data were obtained from comparisons of all sequences in the “ITS” dataset and all the clone sequences spanning the 18S rDNA, ITS+5.8S and 28S rDNA.

3. Synthèse et éléments de discussion

Le nombre d'UTOs de tintinnidés estimé à partir des banques de clones générées pour les échantillons de Villefranche en utilisant comme seuil 99% d'identité de séquence (entre 22-25 UTOs en fonction de la taille du marqueur), est très proche du nombre d'espèces défini sur des critères morphologiques dans le même échantillon (21 morpho-espèces). Nous avons toutefois relevé que le nombre d'UTOs estimé dans l'échantillon de la Mer Ionienne (entre 17-22 en fonction de la taille du marqueur) est inférieur au nombre d'espèces observées au microscope (36). Ces observations confirment le choix du seuil de 99% d'identité de séquence comme une approximation minimale pour corréler la diversité des séquences environnementales avec la diversité morphologique. Ce pourcentage a été également choisi par Santoferrara et collègues (2012), en se basant sur une phylogénie moléculaire construite avec les séquences disponibles publiquement d'ADNr 18S et de nouvelles séquences d'ADNr 18S et 28S obtenues à partir de nombreux spécimens de *Tintinnopsis*, *Tintinnidium*, *Favella*, *Eutintinnus*, *Helicostemella* et *Stenosemella*.

En revanche, le traitement des données de pyroséquençage s'est révélé plus discutable. En effet, en fonction du débruitage ou non des données de séquençage et des différentes méthodes employées pour les débruiter, pour aligner les séquences entre elles et les regrouper en UTOs, les estimations ont été extrêmement variables. Par exemple, pour l'échantillon de Villefranche-sur-Mer, quand les pyroséquences ont été traitées de la même manière que les séquences des banques de clones (sans débruitage, matrice de distance deux à deux, regroupement en groupes partageant 99% d'identité de séquence), les UTOs ont été près de 50 fois plus nombreuses (1159 vs 22 UTOs), soit deux fois plus que le nombre de cellules individuelles de tintinnidés présent dans l'échantillon. Un nombre d'UTOs se rapprochant du nombre de morpho-espèces a été atteint seulement quand les pyroséquences ont été rassemblées en groupe partageant plus de 95% d'identité. En termes de distribution d'espèces, pour cette même analyse, 98,7% des UTOs se trouvaient représentés par moins de 1% des pyroséquences. Cette valeur excède considérablement les 42,9% des espèces observées avec moins de 1% du nombre de tintinnidés dans la communauté sous le microscope et suggère une surestimation du nombre d'espèces à partir des pyroséquences regroupées par cette stratégie. Pour éviter ces surestimations artificielles de la richesse spécifique et de la représentation des espèces rares, il est apparu que les méthodes AmpliconNoise (Quince et al. 2011) ou CD-HIT-OTU (Wu et al. 2011) sont les plus efficaces pour décrire la composition des

communautés de tintinnidés en accord avec la réalité observable sous le microscope. Notamment, la méthode AmpliconNoise, qui permet de prendre en compte la totalité des pyroséquences générées, donne des résultats de diversité similaires à ceux estimés à partir des banques de clones, qui sont en assez bon accord avec les données morphologiques.

L'analyse phylogénétique a permis de faire la correspondance à l'échelle spécifique entre les séquences environnementales d'ADNr 18S générées par les deux approches moléculaires et les séquences d'ADNr 18S amplifiées par PCR directe à partir des cellules individuelles, pour lesquelles la diversité morphologique était connue. Ainsi, ces résultats ont montré que les morpho-espèces dominantes et rares sont représentées efficacement par les études de diversité moléculaire environnementales. De plus, il est intéressant de noter que les approches moléculaires plus conventionnelles (amplification, clonage, séquençage Sanger) et récentes (pyroséquençage des amplicons, si débruités correctement) donnent des résultats en composition phylogénétique extrêmement similaires. Cette correspondance entre la diversité moléculaire environnementale avec la diversité morphologique valide l'utilisation des approches de diversité moléculaires environnementales qui visent spécifiquement les tintinnidés pour décrire la composition des communautés.

Il s'avère que les nouvelles techniques de séquençage des amplicons d'ADNr 18S, bien que permettant d'obtenir un nombre de séquences considérable et de s'affranchir de l'étape de clonage, ne sont pas plus efficaces pour décrire la diversité de notre groupe d'étude que les méthodes standards. Il s'avère surtout qu'il faut traiter les données de pyroséquençage avec beaucoup d'attention et de précaution pour éviter de surestimer de manière absurde la diversité, bien que cela risque d'être bien plus difficile dans les cas où on ne dispose pas d'un contrôle morphologique de référence comme c'était le cas chez les tintinnidés. De manière paradoxale, il se trouverait alors que le meilleur contrôle pour limiter la surestimation de la diversité soit l'estimation de la diversité par amplification, clonage et séquençage Sanger.

C. Diversité spatio-temporelle des ciliés tintinnidés dans la rade de Villefranche-sur-Mer

1. Contexte et Objectifs

Après avoir construit une base phylogénétique de référence et avoir montré l'efficacité des études de diversité moléculaire environnementales pour décrire la diversité des morpho-espèces des tintinnidés, et ce, plus particulièrement dans la rade de Villefranche (voir parties IV.A et IV.B), nous avons voulu étudier la dynamique temporelle des communautés de tintinnidés à différentes profondeurs dans notre site d'étude. Chez les protistes, les études environnementales ont montré des communautés d'eucaryotes changeantes en fonction des conditions environnementales, particulièrement entre des échantillons de surface et de grande profondeur (2500 m) (Countway et al. 2007). Cependant, les assemblages naturels de protistes en fonction de la profondeur restent peu documentés. En outre, il existe encore moins de données moléculaires sur l'évolution des assemblages naturels de protistes au cours de l'année en relation avec les paramètres environnementaux. Les objectifs de cette partie sont de décrire la composition des communautés de tintinnidés en fonction de la profondeur (avec un gradient environnemental allant de la surface à 160 m de profondeur), et ensuite, pour une même profondeur, de savoir comment la communauté évolue sur une année entière. Pour répondre à ces questions, j'ai caractérisé la diversité des tintinnidés par une analyse *fingerprinting* par DGGE en utilisant la région de l'ITS. J'ai ensuite appliqué l'approche standard d'amplification, de clonage et de séquençage de l'ADNr 18S, 5.8S, une partie de l'ADNr 28S et les ITS 1 et 2 comme marqueurs génétiques. J'ai également cherché à mesurer la corrélation des différentes compositions des communautés avec les données environnementales disponibles. Le fait que l'échantillonnage à Villefranche ait été réalisé sur une longue période devrait également nous permettre d'évaluer l'influence de la période d'échantillonnage sur la composition des communautés de tintinnidés. Nous avons aussi l'idée d'avoir un aperçu de la diversité intra-spécifique temporelle de certaines UTOs en utilisant l'information phylogénétique complémentaire contenue dans la région des ITS et la partie de l'ADNr 28S amplifiées et séquencées. Ceci ne serait évidemment possible que si on identifiait des UTOs communes à plusieurs saisons étudiées.

2. Manuscrit de l'article3 : Seasonal dynamics of free-living tintinnid ciliate communities revealed by environmental sequences from the North-West Mediterranean Sea

*Article in preparation***Seasonal dynamics of free-living tintinnid ciliate communities revealed by environmental sequences from the North-West Mediterranean Sea****Charles Bachy¹, David Moreira¹, John R. Dolan², Purificación López-García^{1*}**

1 Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, Bâtiment 360, 91405 Orsay Cedex, France; charles.bachy@u-psud.fr, david.moreira@u-psud.fr, puri.lopez@u-psud.fr

2 Université Pierre et Marie Curie and Centre National de la Recherche Scientifique (CNRS), UMR 7093, Laboratoire d'Océanographie de Villefranche, Marine Microbial Ecology, Station Zoologique, B.P. 28, 06230 Villefranche-sur-Mer, France; dolan@obs-vlfr.fr

*Corresponding author; fax +33169154697

e-mail (puri.lopez@u-psud.fr)

Running title: Temporal dynamics of planktonic tintinnids

Summary

During the last decade, the use of culture-independent molecular approaches to describe the communities of microbial eukaryotes present in natural environments has led to the discovery of a huge diversity of these organisms. However, these analyses have rarely taken into account the temporal variations of diversity patterns. In this work, we studied temporal changes of genetic diversity of tintinnids (Ciliophora) over a two-year survey in a Mediterranean Sea location (Villefranche-sur-Mer, France). The species-rich order Tintinnida is a group of eukaryotes commonly encountered in the coastal marine waters and which is efficiently described using molecular estimates of diversity. We studied the genetic diversity of tintinnids over a one-year survey in a coastal Mediterranean Sea location (Villefranche-sur-Mer, France) at five different depths by combining community fingerprinting analyses using denaturing gel gradient electrophoresis (DGGE) with direct PCR amplification and sequencing of small subunit (SSU), 5.8S, and large subunit (LSU) rRNA genes together with the corresponding internal transcribed spacers (ITS). We amplified tintinnid sequences, detectable by DGGE analysis, in all the samples collected. The analysis of tintinnid community composition by direct PCR amplification and sequencing revealed marked seasonal changes. Despite a good phylogenetic reference sequence dataset for the tintinnids, we also retrieved two abundant phylotypes without any closely related known species that we named Environmental Tintinnid Clade 1 and 2 (ETC-1 & 2). Surprisingly, the tintinnid species composition based on the SSU rDNA alone did not show any clear relationship to measured environmental parameters (temperature, salinity, light, phytoplankton biomass). Nonetheless, a comparison of tintinnid community composition using Unifrac revealed three significant clusters of sequences grouping, respectively, samples collected in winter, autumn and summer, suggesting a seasonal effect on tintinnid community composition. In addition, phylogenetic trees based on the concatenated SSU + LSU rDNA and ITS sequences showed a better resolution than SSU rDNA alone to discriminate closely related species.

Introduction

The use of culture-independent molecular approaches to describe the communities of microbial eukaryotes, or protists, present in natural environments has led to the discovery of a huge diversity of these organisms (López-García et al., 2001; Moon-van der Staay et al., 2001; Epstein and López-García, 2008). Most of these studies are based on the PCR amplification of the nuclear encoded small-subunit rRNA gene (SSU rDNA). Additional genetic markers such as the internal transcribed spacers (ITS) or the large-subunit rRNA gene (LSU rDNA) have also, though more rarely, been used to improve the phylogenetic resolution of SSU rDNAs (e.g. Marande et al., 2009). Because many protist species cannot currently be identified with a microscope, microbial ecologists apply these culture-independent approaches to identify their distribution patterns as a required step to further understand the underlying mechanisms driving biodiversity and biogeography. Several studies have revealed that microbial eukaryotic community composition varies across environmental gradients related to depth (Countway et al., 2007; surface vs 2500m), salinity (Estrada et al., 2004; 4% vs 37%) or oxygen content (Orsi et al., 2012; anoxic vs oxic). However, the seasonality of aquatic protists has been insufficiently addressed. Despite so, temporal changes in community composition may be remarkable as revealed in marine surface waters (Massana et al., 2004; Romari and Vaultot, 2004), continental lakes (Lepère et al., 2006; Nolte et al., 2010) and oligotrophic aquatic environments such as peatlands (Lara et al., 2011).

If studies on spatial and temporal variation of protist communities using environmental surveys are limited, those of specific protist taxa are even rarer. Nonetheless, the study of temporal variation of particular protist lineages provides an increased level of resolution that may be useful to assess specific ecological questions. The control of dinoflagellate populations over time by specific parasites is such an example (Chambouvet et al., 2008). Environmental molecular surveys using specific primers to amplify the SSU rDNA of certain protist lineages allows to efficiently describe the diversity of clades which are often overlooked by using eukaryotic universal primers. For example, studies targeting specifically Cercozoa (Bass and Cavalier-Smith, 2004) or Haptophyta (Liu et al., 2009) showed an unsuspected genetic diversity within these groups. A few of these taxon-specific approaches aimed at revealing spatial patterns in marine ecosystems, showed, for instance, a depth-dependent community structure of diplomonads (Euglenozoa) in the water column (Lara et al., 2009) or a variation of kinetoplastid (Euglenozoa) composition with distance in the deep-sea

floor (Salani et al., 2012). A relative intense work has been done on marine ciliates (choreotrichs and oligotrichs), for which strong differences in community composition across environmental and depth gradients have been reported (Doherty et al., 2007; Doherty et al., 2010; Tamura et al., 2011). But, so far, temporal variation of marine ciliate taxa has not been explored.

Tintinnid ciliates are in some way an ideal model group to study how assemblages of microbial eukaryotes vary seasonally. Among ciliates, they are recognizable by the elaboration of a shell, or lorica. As part of the microzooplankton, tintinnids are a category of planktonic heterotrophs responsible for grazing most of the algal production in marine systems. Accordingly, there is a long tradition of studying their seasonal dynamics in marine coastal waters, notably in the Mediterranean Sea based on morphological surveys, which usually revealed a strong seasonal variation (e.g. Modigh and Castaldo, 2002; Dolan et al., 2006; Sitran et al., 2007). In addition, the determination of SSU rDNA sequences from single Mediterranean tintinnid cells has recently enriched the genetic database for these ciliates and provided a strong phylogenetic framework for further environmental molecular analyses in this marine region (Bachy et al., 2012). Furthermore, the distinctive morphology of tintinnids has made it possible to carry a recent comparative study of molecular (SSU rDNA sequences) and morphological estimates of species richness in Mediterranean plankton samples, which showed that the number and identity of species observed morphologically was roughly similar to the number and identity of operational taxonomic units (OTUs; defined as groups of sequences with $\geq 99\%$ identity) identified in SSU rDNA libraries from the same samples (Bachy et al., submitted, see article section IV.B). This dual correspondence between the diversity of sequences in gene libraries and morphology estimates allows making safe conclusions about the real significance of SSU rDNA-based observable changes across samples as well as to merge morphology and sequence-based observations across time and space.

In this study, we aimed at testing whether spatial and temporal changes in tintinnid community composition classically detected by microscopy observations are also detected by molecular approaches as well as determining at which level of taxonomic resolution those changes occur. To do so, we studied the genetic diversity of tintinnids over a one-year survey in a coastal Mediterranean Sea location (Villefranche-sur-Mer, France) at five different depths by combining community fingerprinting analyses using denaturing gel gradient

electrophoresis (DGGE) with the direct PCR amplification and sequencing of SSU, 5.8S, and LSU rRNA genes together with the corresponding internal transcribed spacers (ITS).

Results

Fingerprinting analysis of planktonic tintinnid communities in the water column along one year of survey

As a first approximation to the composition of the tintinnid communities in the Bay of Villefranche, we generated tintinnid-specific fingerprints of a relatively variable genomic area covering the ITS region using denaturing gel gradient electrophoresis (DGGE) for all the samples collected monthly from February 2009 to January 2010 at five different depths (5 m, 25 m, 50m, 100 m, 160 m) (see Supporting information, Table S1). We observed bands corresponding to tintinnid ITS regions in all the DGGE profiles (Fig. S1), ranging from 2 bands for the less diverse samples (VilleFr-12 and -19) to 12 for the richer ones (VilleFr-36 and -37). This indicates that planktonic ciliates are present throughout the year. A cluster analysis of DGGE profiles divided the samples in two major groups seemingly according to their depth in the water column. One cluster corresponded to surface and intermediate depth samples (5-50 m) obtained in different dates, including five 100 m samples and one deep sample (VilleFr-5, 160 m), and the second cluster included the other deep samples (100-160 m), the surface waters (5-25 m) of June and July 2009 (VilleFr-21, -22, -26 and -27) and the 25 m sample of September 2009 (VilleFr-32). There were also nine samples (VilleFr-2, -6, -7, -11, -12, -13, -17, -19 and -56) which did not group with the two major clades. All the bands identified in DGGE patterns were cut and their corresponding DNA fragments purified and sequenced, which showed that all of them did indeed affiliate with tintinnid sequences (Table S2).

Spatio-temporal patterns of tintinnid diversity based on SSU rDNA sequences

For further in-depth diversity analyses of tintinnids, we selected the five samples VilleFr-1 (5 m), VilleFr-2 (25 m), VilleFr-3 (50 m), VilleFr-4 (100 m), VilleFr-5 (160 m), corresponding to the different sampling depths of the Villefranche water column in February 2009 plus three additional samples at a same depth (50 m) collected at the other three seasons VilleFr-23, VilleFr-58 and VilleFr-43, which in general displayed rich band profiles in DGGE gels and

grouped in different subgroups within the first major cluster (Fig. S1). The sample VilleFr-43 was the subject of a recent study comparing the tintinnid diversity observed by morphology under the microscope with that observed by SSU rDNA and ITS amplification, cloning and Sanger sequencing and pyrosequencing, showing a good agreement between the morphospecies and the OTUs identified (Bachy et al., submitted, see article section IV.B).

For each selected sample, we characterized the tintinnid diversity by cloning and (Sanger) sequencing of SSU, 5.8S and partial LSU rRNA genes and internal transcribed spacers (ITS) using tintinnid-specific primers. The analysis of several markers from the ribosomal RNA cluster, including the ITS which is generally more variable than the coding regions, frequently allows describing diversity at a finer taxonomic level as compared with that derived from SSU rDNA sequences alone. However, because we have robust information on how the tintinnid genetic diversity relates to the morphological diversity at the species-level only for the SSU rRNA gene, we first measured the composition of tintinnid OTUs using a 99% sequence similarity threshold for this marker. A total of 19 OTUs were detected for all the 8 samples. Species richness varied between 3 in VilleFr-1 and 9 in VilleFr-23 and VilleFr-43 (Table 1). Despite a relative low number of clones analyzed per library, we obtained a good coverage for them, in most cases well above 90% , and in any case higher than 82%, which suggested that we had described most of the tintinnid specific diversity present in our samples (Table 1).

For subsequent phylogenetic analyses, we selected one representative of our SSU rDNA sequences per OTU and aligned these sequences along with a representative set of sequences of well-described Mediterranean tintinnids (Fig. 1). Even if the relationships between many tintinnid families and some lineages *incertae sedis* (*Tintinnopsis* spp and *Clymacocylis* sp.) were unresolved, as shown in a previous taxonomical study (Bachy et al. 2012), the sequences within families grouped together with strong support (ML bootstrap values -BV- $\geq 92\%$). The OTUs identified in this study were distributed across 9 families of the order Tintinnida, with the richest diversity found within the Tintinnidae with 5 OTUs related with *Amphorides quadrilineata*, *Steenstrupiella steenstrupi* and *Salpingella acuminata*. Some OTUs were detected in low amounts in only one sample, suggesting that they are less abundant, at least at the studied depths, along the year. This was the case of the OTU17 related with *Tintinnidium mucicola* only found in November 2009 (VilleFr-43) or the OTU10 related with *Parundella*

aculeata only found in June 2009 (VilleFr-23) (Figure 1). For more represented OTUs, we observed differences both in space and time.

Regarding the tintinnid community composition along the water column analyzed in February 2009, some OTUs appeared more abundant at higher depths (OTU13 related to *Cyrtarocylis* spp.), others at intermediate depths (OTU3 related to *Undella* spp.) and others in surface waters (the environmental OTU4). At the same time, other OTUs were detected in low amounts at various depths from surface to deeper waters, suggesting that they were present all throughout the water column (OTU5, related to *Dictyocysta* spp. or OTU1 related to *Codonelopsis* spp.) (Figure 1).

Regarding the tintinnid community composition along the seasons, we also observed marked patterns. One of the most notable differences observed was the absence of tintinnids belonging to the families Tintinniidae, Eutintinnidae, Tintinnidae and Xytonellidae in February 2009, members of which were detected in other seasons. Only members of the related families Codonellidae, Codonellopsidae, Undellidae and Cyrtarocylidae were detected in winter in diverse relative abundances (Figure 1). If we compared samples collected at the same depth, 50 m, but at different seasons, some OTUs were detected only in winter (the environmental OTU2, the OTU6 related to *Stenosemella*) and others only in other seasons (e.g. the OTU12 related to *Clymacocylis* and members of the Tintinniidae, Eutintinnidae, Tintinnidae and Xystonellidae), whereas some OTUs were detected throughout all the seasons. The latter case was that of the OTU3 – *Undella* spp.- which was remarkably present both along the year and across the water column (Figure 1).

Surprisingly, the tintinnid OTU4 and OTU6 were solely represented by environmental sequences and had no closely related sequences of described tintinnid species (their SSU rDNA similarity with that of the closest described species was 97.81% and 97.92%, respectively, far below the tintinnid morphospecies threshold that can be placed conservatively at 99% similarity). In addition, their phylogenetic position within tintinnids is not resolved; they branched in basal position of Codonellidae and Codonellopsidae representatives with extremely weak support (BV <50%). We named these two phylotypes Environmental Tintinnid Clade 1 and 2 (ETC-1 and -2). ETC-1 corresponded to OTU4, which was numerically dominant in all winter samples, being present at all depths in February 2009 and in the 50 m-deep February 2010 but not the rest of the year (Fig. 1). The prevalence of

this OTU in the clone libraries from February 2009 decreased with the depth, ranging from 86% of all tintinnid sequences at the surface sample (VilleFr-1, 5m depth) to 29% at the deepest sample (VilleFr-5, 160m depth).

Tintinnid community comparison and abiotic environmental factors

In order to see whether the tintinnid community composition was influenced by physicochemical environmental parameters we measured temperature, salinity, light and fluorescence at the time of sampling. Although temperature and light (photosynthetic active radiation, PAR) varied with depth and seasons, fluorescence and salinity remained relatively constant in the Villefranche water column (Table S3). Furthermore, temperature changes were very limited; temperature varied from 13.4 °C (February 2010, 50 m) to 18 °C (June 2009, 50 m) over all seasons, and was constant in February 2009 in all samples (13.7 °C) due to the mixed-layer structure of the water column. This remarkable constancy of parameters (with the exception of light) did not seem to correlate, at first sight, with the observed changes in the tintinnid specific richness or assemblages.

To test whether there was a correlation between physicochemical parameters and tintinnid community composition, we carried out a Monte Carlo permutation test, using unconstrained permutations in redundancy analyses (RDA). The results showed no significant effect of the measured physicochemical parameters on tintinnid assemblages (p-value=0.806). Nonetheless, the phylogenetic method Weighted UniFrac for comparing compositions of tintinnid communities showed grouping between samples depending on the period of sampling (Fig. 2), which suggests that the causes of such clustering may be not related directly to abiotic parameters. We applied principal coordinate analysis (PCoA) of the pairwise weighted UniFrac distances to the tintinnid communities detected in our 8 selected samples from Villefranche plus one additional sample that we have previously studied from the Ionian Sea (Ioni-7) as well as cluster analysis using UPGMA. They showed that the Villefranche and Ionian Sea samples fell into three well supported and distant groups (Fig. 2). The first two principal coordinates which explained 68.4% and 16.7% of the variation showed that the samples VilleFr-1 to -5 (Feb 2009) and VilleFr-58 (Feb 2010) collected in winter grouped together, samples VilleFr-43 (Nov 2009) and Ioni-7 (Oct 2010) collected in autumn grouped together, and the remaining sample VilleFr-23 (Jun 2009) was apart (Fig. 2A).

Jackknife bootstraps supported (>90%) these samples repartition (Fig. 2B). To see how the different OTUs influenced community changes we carried out an ordination analysis based on a linear ordination method (PCA, Fig. S2). It indicated that the main gradient of species composition change was related to the abundance of the OTU4 (ETC-1), and that the divergent position of the sample VilleFr-43 was explained by the more frequent occurrence of the OTU5 (*Dictyocysta*-related) and the OTU3 (*Undella*-related).

Intraspecific tintinnid diversity

To study the diversity of tintinnids at a finer taxonomic scale, we estimated the number of OTUs by applying a threshold of $\geq 99\%$ sequence identity at the full-length sequences amplified (SSU + ITS1 + 5.8S + ITS2 + partial LSU rDNA). For the eight samples studied, we detected a total of 26 OTUs according to this finer criterion, 7 more than when we considered the SSU rDNA alone. This increase in the number of OTUs was not surprising since the ITS region (ITS1+5.8S rDNA+ITS2) and the beginning of the LSU rDNA are more divergent than the SSU rDNA (Bachy et al., 2012; Santoferrara et al., 2012). For example, the sequences of OTU3, OTU4 and OTU6 related with, respectively, *Undella*, ETC-1 and ETC-2 (Fig. 1), were split in the new clustering analysis into four, three and three OTUs, respectively (data not shown). In order to know whether this richer diversity correlates with different morphological features, we built an ML tree with the full-length sequences that composed the perennial OTU3 (*Undella* sp.) based on the SSU rDNA, together with the two sequences of *Undella* present in GenBank, *Undella claparedei* isolate CB818 and *Undella hyalina* isolate CB854. The Fig. 3A shows different sub-clusters of sequences that are either closely related with known species or composed exclusively of environmental sequences. A pool of 11 sequences detected in VilleFr-2 to 5 and VilleFr-23 were phylogenetically associated with *Undella claparedei* (sub-clade 3-I). Two distant clades were exclusively composed of 4 and 2 sequences detected in VilleFr-43 (sub-clade 3-II and 3-III), among which the latter ones appeared to be associated with *Undella hyalina* with strong support (BV 100 %). The six remaining sequences (VilleFr-4_clone18 and VilleFr-58_clone19, VilleFr-3_clone18 and VilleFr-43_clone45, VilleFr-5_clone21 and VilleFr-2_clone4) were more distantly related, clustering in three clades (Sub-clade 3-IV, 3-V, 3-VI, respectively). This may suggest that part of the morphologic diversity existing within the genus *Undella* remains to be related with these SSU rDNA and ITS sequences.

The case of the OTU5 related with *Dictyocysta* illustrates the recurrent problem of lack of resolution of the SSU rDNA to discriminate closely related species in some tintinnid families. The sequences clustering as OTU5 based on SSU rDNA sequences alone were still part of the same OTU when considering also the ITS, 5.8S and partial LSU sequences. However, a ML tree inferred with the full-length sequences of this OTU, together with the sequences of *Codonella aspera* isolate CB235 and *Codonaria* spp. isolates CB25, CB82 and FG42, revealed that sequences attributed to *Dictyocysta* based on the SSU rDNA sequences are composed of 3 sub-clades moderately to strongly supported (BV 75 to 100%)(Fig. 3B). Surprisingly, these 3 sub-clades were related with distinct tintinnid species. Sequences from VilleFr-4, -5, -23, -43 and -58 clustered with *Codonella* and *Codonaria* (sub-clade 5-I). Whereas, sequences unambiguously identified as *Dictyocysta* (based on the similarity between clone sequences and *Dictyocysta lepida* isolate FG303 (JQ408188) on the SSU rDNA $\geq 99.8\%$) from VilleFr-1, -3, -23, -43 and -58 clone libraries clustered together in two closely related sub-clades (sub-clade 5-II and 5-III) with no described relatives having the ITS region sequenced.

Concatenated SSU rDNA+ITS+5.8S rDNA+LSU rDNA phylogeny

To improve the resolution of the current tintinnid phylogeny, we carried out a molecular phylogenetic analysis of concatenated gene markers. A total of 293 long environmental tintinnid sequences (~1900 bp) encompassing the SSU rDNA and up to the beginning of the LSU rDNA were retrieved from the eight studied samples and 17 other samples from Villefranche analyzed in this work (Table S1) and the Ionian Sea. Sequences were classified based on the SSU rDNA for which there is a more extensive phylogenetic data set of sequences from described tintinnids. We aligned these environmental sequences to 19 taxonomically-defined sequences publicly available covering the same genomic region (Fig. 4). After molecular phylogenetic analyses (Fig. 4), the environmental sequences could be classified into clusters of tintinnids corresponding to family or environmental clade levels: Tintinnidiidae (1 sequence), Tintinnidae (41 sequences), Eutintinnidae (9 sequences), Cyttarocylididae (12 sequences), Xystonellidae (12 sequences), Undellidae (46 sequences), ETC-1 (86 sequences), ETC-2 (11 sequences), Codonellopsidae (20 sequences), Codonellidae (33 sequences), *incertae sedis* (corresponding to *Tintinnopsis* spp. and *Climacocylis* sp., 6 and 14 sequences, respectively). Two of the OTUs showed inconsistencies between the phylogenetic position based on SSU rDNA sequences alone and that of the concatenated SSU

rDNA+ITS 1+5.8S rDNA+ITS 2+LSU rDNA. Considering the SSU rDNA alone, VilleFr-5_clone 22 and VilleFr-4_clone 23 branched with strong supports (-BV- 100%) within *Stenosemella* spp. and ETC-1, respectively (data not shown). However, considering the concatenated sequences, the first sequence branched in a basal position of ETC-2 with strong support (BV 100%) and the second one branched with moderate support (BV 68%) in a basal position of an *incertae sedis* closely related with Codonellopsidae and Codonellidae (Figure 4). These sequences which presented incongruent phylogenetic positions might be chimerae undetected by our several filtering methods (see Materials and methods), showing the difficulty to detect chimeric sequences when dealing with closely related species. But they potentially are true sequences, thus we actually kept them in our analyses.

In addition to the 8 Villefranche samples explored in greater details, we also constructed gene libraries and sequenced several (but few) clones from other Villefranche samples corresponding to other depths and/or seasons. We obtained in this way a total of 65 additional sequences (Table S1), whose analysis revealed some interesting features. For example, sequences grouping with the abundant environmental clade ETC-1 were also retrieved in March 2009 at 50 m depth, in June 2009 at 100 m and 160 m depths and in November 2009 at 100 m depth. So, this clade was not exclusively present in the samples from February 2009 and 2010, but also in samples from June and November 2009, but at depths deeper than 50 m (Fig. 4). This was in agreement with the results reported from the DGGE analysis, where the bands we cut, purified and sequenced showed that ETC-1 representatives were present during the whole temporal survey, and at all the sampling depths (Fig. S1 and Table S2).

Discussion

Our molecular survey of tintinnid community composition revealed a relatively wide and rich phylogenetic diversity. A total of 22 distinct DGGE bands and 19 OTUs based on SSU rDNA sequence similarity were detected (Fig. 1 and Fig. S1). The OTUs were distributed in 9 different families of Tintinnida, from the most basal family Tintinnidiidae to the most distal families Codonellopsidae and Codonellidae. Currently, the numerous tintinnid morphospecies with published SSU rDNA sequences make it reliable the phylogenetic identification at the genus/species level within this order (Bachy et al., 2012). Indeed, all the SSU rDNA sequences retrieved from the 8 samples studied in more detail from the Villefranche Bay were

≥98% identical to tintinnid sequences publicly available. By contrast, in other ciliate groups such as the *Oligotrichia*, a sister clade of the *Tintinnida*, the morphospecies sampling by molecular methods is sparse and a large phylogenetic diversity observed in environmental molecular studies lacks described relatives (Doherty et al., 2007; Tamura et al., 2011). Establishing the link between morphology and SSU rDNA sequences has been facilitated by the distinctive morphology of tintinnids and their easy identification in planktonic samples using an inverted microscope as compared with other ciliate species for which identification needs a more complex process (Montagnes and Lynn, 1987). However, despite an extensive work of single-cell isolation, identification, and subsequent SSU-rRNA and ITS sequencing in general and at our sampling site of Villefranche (Bachy et al., 2012), the analysis of tintinnid communities by direct PCR amplification and sequencing of rRNA genes from plankton samples revealed two phylotypes without any phylogenetically closely related described species.

The phylotype ETC-1 was particularly interesting in this sense. Sequences related to ETC-1 were the most abundant in the clone libraries of February 2009 at all depths (VilleFr-1 to -5) and February 2010 at 50 m (VilleFr-58), suggesting that these organisms may be an important ecological component of planktonic communities in this coastal Mediterranean site. Furthermore, some sequences related to this clade were also retrieved in some other seasons but in samples deeper than 50 m. Similarly, a second group also composed exclusively of environmental sequences (ETC-2) was found in lower abundance in three clone libraries from 50 m to 160 m depth in February 2009 (VilleFr-3, -4 and -5), and detected in the clone library from 50 m depth in December 2010 (VilleFr-73). The lack of closely related described species to these environmental clades could be explained by several hypotheses. First, these species could be rare in plankton (and hence little observed) but possess a high number of SSU rDNA copies in the genome, which would revert in their higher relative abundance in clone libraries, as already shown for *Codonellopsis* spp. (Bachy et al., submitted, see article section IV.B). Second, ETC-1 and ETC-2 sequences may represent non-conventional tintinnids (naked or minute specimens) for which no morphological description has been made and that have been overlooked during our microscopy observations under the microscope and not isolated for SSU rDNA amplification and sequencing. And third, as the ETC-1 and 2 sequences branched at the base of the Codonellidae and Codonellopsidae, the corresponding specimens might present a very similar morphology to those of these family representatives that we had already

genotyped and, hence, for which we did not continue to isolate for SSU rDNA analysis, thus preventing their identification by sequence divergence.

The analysis of SSU rDNA libraries revealed distinct spatial and seasonal patterns in the occurrences of tintinnid OTUs detected. A limited number of OTUs were found in two or more seasons. These temporal shifts in communities are characteristic of tintinnid assemblages described using microscopic analysis in most coastal ecosystems, and specifically in our well-studied site of Villefranche Bay (Balech, 1959; Rassoulzadegan, 1979; Dolan et al., 2006). We did not observe a clear relationship between the OTU diversity within our samples and several environmental variables (depth, salinity, temperature, PAR and fluorescence). The lack of correlation between abiotic parameters and tintinnid community composition may seem surprising. Therefore, we hypothesize that biotic factors may be a more important factor shaping the tintinnid community in coastal waters. In fact, several ecological studies based on morphological specificities of tintinnids, mainly the oral diameter which has been related to the optimum prey size, showed that communities are determined by resource partitioning among different species in coastal waters (Dolan et al., 2012). In addition, tintinnid communities could also be influenced by competition with other microplankton grazers and by mortality due to metazoan grazers (Pierce and Turner, 1992) or parasites (Bachvaroff et al., 2012). Consequently, in order to understand the driving forces shaping tintinnid communities, the study of their diversity by molecular approaches, which efficiently describes their morphological diversity (Bachy et al., submitted, see article section IV.B), should be combined with both the study of abiotic and, most importantly, biotic (resources, predation, competition) environmental factors.

As we have mentioned, although the causes are not well established, we observed a strong seasonal effect on tintinnid community composition. Although the tintinnid composition varied across samples, some samples resembled between them more than others, as showed by PCoA analyses (Figure 2). A cluster grouped the five February 2009 samples from all the depths (5 m to 160 m). At this season, the water column experiences mixing from the surface to the floor, which would explain a homogeneous tintinnid distribution through the water column, even if subtle differences exist between samples. These variations of the tintinnid composition with depth could be explained by local scale variability as evidenced in the Villefranche Bay at very small geographic scale based on tintinnid morphology (Dolan and Stoeck, 2011). They could also be explained by a stochastic sampling effect in the case of

OTUs with relative low abundances. However, stochastic effects are unlikely to explain the decrease of the dominant ETC-1 with depth (Figure 1). In this case, we cannot discard that the apparent indirect effect of biotic factors related, for instance, with the availability of preys whose distribution is stratified, for instance photosynthetic organisms adapted to particular light intensities. This cluster also contained the other winter sample collected one year later in February at 50 m depth. So, the tintinnid communities are reproducible in winter from one year to the next, mainly because of dominance of the ETC-1 phylotype. Similarly, a second cluster contained the two samples collected during autumn 2009 and 2010 from distant locations (coastal Ligurian Sea vs offshore Ionian Sea). A distinctive cluster represented the only sample collected in late spring, which may be potentially characterized by tintinnid community living in a stratified environment.

The study of tintinnid community using SSU rDNA+ 5.8S rDNA+ LSU rDNA and ITS revealed a greater specific diversity than that detected from the clustering of SSU rDNA sequences alone. For example, the perennial *Undella*-related OTU4 defined on the strict criterion of $\geq 99\%$ SSU rDNA sequence identity, revealed to be more diverse when considering ITS and LSU rDNA positions. In this case, the detailed phylogenetic approach showed that, at least, a part of this diversity is linked with morphological variations within the genus. The case of the *Dictyocysta*-related OTU5 is more difficult to interpret. Considering SSU rDNA alone or SSU rDNA+ITS+LSU rDNA positions, the clustering based on sequence identity defined only one OTU. However, the ML tree revealed that this OTU was composed of different well-supported phylogenetic groups related with particular morphological traits. In environmental molecular surveys, to discriminate species by choosing appropriate sequence similarity threshold is a crucial prerequisite for a fair assessment of diversity. For practical reasons, the cut-off level for species is based on SSU rDNA sequence similarity (frequently established at 99% for protists). However, this criterion does not accommodate the whole complexity of eukaryotic, ciliate or tintinnid diversity (Caron et al., 2009; Nebel et al., 2010; Santoferrara et al., 2012) and needs to be adapted to each eukaryotic taxon whenever molecular and morphological data are available. We previously showed that even partial sequences of SSU rDNA are reasonably accurate to discriminate tintinnid species diversity (Bachy et al., submitted, see article section IV.B). However, the sequencing of ITS region and partial LSU rDNA together with the SSU rDNA has revealed more extensive genetic diversity. Whether this genetic diversity detected within and among OTUs correlates with

inter- or intra-specific diversity, or even with intra-individual variability is unclear. To unravel these different low taxonomic levels of diversity within tintinnids, finer population genetic studies based on clearly defined species and targeting molecular markers other than the SSU rDNA will be required.

Experimental procedures

Sampling site and environmental measures

We studied samples from a station named 'Point B+' (43°41'00"N, 7°18'44"E) at the Bay of Villefranche which allows to sample until relatively deep waters for a coastal site (180 m depth). Sampling was conducted monthly from February 2009 to February 2010 at five different depths in the water column: 5, 25, 50, 100, and 160 m depth (Table S1). In addition, three samples from 50 m depth were collected in May, September and December 2010, representing different seasons of a second year. Plankton samples were collected using 12 l Niskin bottles, and immediately pre-filtered through 200 µm Nitex mesh and then filtered through 5 µm pore diameter TMTP (Millipore) filters to concentrate biomass in the 5-200 µm fraction. Filters were conserved in absolute ethanol at -20°C until further processing. Systematically, the day before sampling, temperature, salinity, chlorophyll a and photosynthetically available radiation (PAR) were measured by the SOMLIT (Service d'Observation en Milieu Littoral, <http://somlit.epoc.u-bordeaux1.fr/fr/>). Temperature-induced stratification occurred in the water column from April to October in 2009 and 2010.

DNA purification and Denaturing Gradient Gel Electrophoresis

Genomic DNA was purified from the biomass retained in filters using the Ultraclean MoBio Soil DNA kit (MoBio, Solana Beach, CA, USA). Nucleic acids were resuspended in 100 µl 10 mM Tris-HCl, pH 8. From the February 2009 to January 2010 samples, ITS region fragments of approximately 250-260 bp were amplified with the specific tintinnid forward primer Tin454-ITSFw-GCclamp (5'-CGCGCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCGCCCCGCAATYTGTTCAGRCGMAAGC-3') and the reverse tintinnid primer Tin454-ITSRev (5'-AGCAATAGAAGGGCATCTA-3'). These group-specific primers were especially designed to amplify the 5.8S rDNA and partial adjacent ITS 1 and 2, which present variable GC

content (Bachy et al., submitted, see article section IV.B). PCR reactions were carried out in 25 μ l of reaction buffer containing 2 μ l of the total DNA, 1.5 mM MgCl₂, dNTPs (10 nmol each), 20 pmol of each primer, and 0.2 U Taq Platinum DNA Polymerase (Invitrogen). PCR amplifications were performed with the following conditions: an initial denaturation step at 94°C for 3 min, 30 cycles consisting of a denaturation step at 94°C for 15 s, an annealing step of 30 s (a touch down procedure with a decreasing annealing temperature from 60°C to 50°C for the 10 first cycles was applied followed by an annealing temperature of 50°C for the following 20 cycles) and a polymerization step at 72°C for 1.5 min, and a final step of 1 h extension at 72°C. Migration of PCR products was done in a denaturing gradient gel using the CBS (California, USA) electrophoresis system. A solution of urea (7 M) and 40% formamide was used as denaturing agent with a concentration gradient from 20% to 40% in a polyacrylamide gel (8%). 25 μ l of PCR product were loaded for each sample in parallel with 50 bp ladder DNA markers (Promega, Lyon, France) every four lanes for next-step computational normalization. The gel was run at 150 V for 6 hours at 60°C in 0.5 \times Tris-acetate-EDTA (1X: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8,0) buffer and then stained with SYBR Gold (Invitrogen, Carlsbad, CA, USA) for 1 hour and photographed under UV light. Characteristic bands of various migration levels were cut and purified for sequencing. Photographed gels were normalized using the Bionumerics® 5 software (AppliedMaths, Saint-Martens-Latem, Belgium) based on the position of ladder standards. Band positions were assigned manually. Clustering analysis of all the DGGE patterns was carried out using UPGMA (Unweighted Pair Group Method with Arithmetic mean) clustering applying a Dice coefficient.

Small-subunit rRNA gene, ITS region and partial large-subunit rRNA gene library construction

We amplified and cloned a region encompassing the SSU rDNA, the ITS region and partial LSU rDNA (amplicon size ~1900 bp) using specific primers for tintinnids from seven selected samples: five samples from five different depths of the Villefranche water column collected in February 2009, VilleFr-1 (5 m), VilleFr-2 (25 m), VilleFr-3 (50 m), VilleFr-4 (100 m), VilleFr-5 (160 m), and two samples collected at 50 m depth in June 2009 (VilleFr-23) and February 2010 (VilleFr-58). To amplify the SSU rDNA-ITS-partial LSU rDNA fragment, the set of tintinnid-specific primers 18S-Tin3F (5'-

GCGGTATTTATTAGATAWCAGCC-3') and 28S-TinR1 (5'-TGGTGCACTAGTATCAAAGT-3') was used (Bachy et al., 2012). PCR was carried out under the following conditions: 35 cycles (denaturation at 94 °C for 15 s, annealing at 52 °C for 30 s, extension at 72 °C for 2 min) preceded by 3 min denaturation at 94 °C, and followed by 15 min extension at 72 °C. The corresponding amplicon libraries were constructed using the Topo TA cloning system (Invitrogen) following the instructions provided by the manufacturer. Positive inserts of expected size were selected from each library and completely sequenced with the forward and reverse primers, and also the internal eukaryotic-specific primer 1498R (5'CACCTACGCAAACCTTGTTA; Lopez-Garcia et al. 2003). We determined between 20 (VilleFr-5) and 36 sequences (VilleFr-23) for each library. Long amplicon libraries were already constructed for the VilleFr-43 (Nov 2009, Villefranche Bay, 50 m depth) and Ioni-7 (Oct 2010, Ionian Sea, 77 m depth) samples (Bachy et al., submitted, see article section IV.B). We searched for recombination events to detect potential chimeric artefacts within our clone sequences, which turned out to be a complicated matter when working on phylogenetically closely related sequences. The Recombination Detection Program, RDP version 3.42 (Martin et al., 2005), was used with settings (a Bonferroni corrected P value cut-off of 0.01) for the different detection methods, including GENECONV (Padidam et al., 1999), Chimaera (Posada and Crandall, 2001) and MaxChi (Maynard Smith, 1992) implemented in RDP3.42. With a reference dataset of 23 tintinnid sequences covering the same region (Bachy et al., 2012), a total of 6 potential PCR-derived recombined sequences (3 in VilleFr-3, 3 in VilleFr-4) were removed from further investigations. A total of 197 complete high-quality sequences generated from the Villefranche Bay samples remained for clustering and phylogenetic analyses. Sequences newly reported in this paper have been deposited in the GenBank database under accession numbers AA000000-AA000000.

Clustering of tintinnid sequences and phylogenetic analyses

Sequences were aligned using MAFFT (Katoh et al., 2002). Then, two distance matrices were generated in Phylip format based either on the SSU rDNA alone, or on the complete sequences. The resulting matrix was used as input for MOTHUR (Schloss et al., 2009) in order to group sequences in Operational Taxonomic Units (OTUs) by average linkage clustering. In this study, an OTU is defined as a cluster of SSU rDNA sequences sharing $\geq 99\%$ of similarity and has to be considered as a good proxy for protist and, most particularly,

tintinnid species identification (Santoferrara et al., 2012; Bachy et al., submitted, see article section IV.B). Coverage values were calculated using the Good estimator (Good, 1953) following the equation $C = (1 - n/N) \times 100$, where C is the percentage of coverage of the library, n the number of singletons and N the total number of clones examined. A subsequent clustering step of the complete sequences sharing $\geq 99\%$ of similarity was then realized. For in-depth phylogenetic analyses, we selected a representative sequence for each OTU. Together with their most similar sequences (known/described species) identified by BLAST (Altschul et al., 1997) in GenBank and some typical Mediterranean tintinnid sequences (Bachy et al., 2012), they were aligned using MAFFT. A phylogenetic tree was constructed using maximum likelihood (ML) with the software TREEFINDER (Jobb et al., 2004) under the general time reversible (GTR) model of sequence evolution with four substitution rate categories and a Gamma law to accommodate for among-site variation. Non-parametric bootstrap analyses were inferred using 1000 replicates. To test if the general topology of the tree remained congruent, we performed a Bayesian analysis on the same dataset with the GTR + Γ + I model of sequence evolution, using the software MrBayes (Huelsenbeck and Ronquist, 2001). Bayesian inference was also conducted using MrBayes, starting with a random tree, ran for one million generations in four chains and excluding the first 35000 generations as a burnin. The ML and Bayesian analyses produced congruent topologies. We then conducted a phylogenetic analyses with all the sequences and positions available from the fully sequenced SSU rDNA+ITS+LSU rDNA regions, to determine whether working with more additional molecular markers (implying more positions) could improve the phylogenetic resolution within the tintinnid order. We obtained a dataset containing 293 sequences of ~1900 bp from the 8 clone libraries (see above), from 16 other clone libraries generated from other Villefranche samples (VilleFr-8, -18, -22, -24, -25, -28, -31, -32, -33, -34, -44, -48, -53, -63, -68, -73) but for which we generated few sequences (between 1 and 11), and from the Ioni-7 clone library (Bachy et al., submitted, see article section IV.B). Individual clone sequences were identified to the species-level based on the SSU rDNA similarity, since for this marker we have a more extensive phylogenetic data set of sequences from described single-cells. We constructed a ML tree with the complete clone sequences and 19 reference tintinnid sequences covering the full-length SSU+5.8S+partial LSU and ITS.

Multivariate analyses

In order to compare tintinnid diversity between the 8 studied samples of our seasonal sampling and the distant Ioni-7 sample, we performed a hierarchical clustering in UniFrac (Lozupone et al., 2006) with an ML tree built as mentioned above with representative sequences of each OTU labeled according to the sample it comes from. Using the weighted and unweighted UniFrac, we determined whether our tintinnid communities were related to one another based on the different depth and seasonal sampling. The principal coordinates describe how much the variation of our data set is explained by the first two axes. The weighted and unweighted analyses gave very similar results. We also assessed the ordination of tintinnid communities with the associated environmental variables by studying the same data set that included the OTUs from the coastal site of Villefranche. For this, we performed a constrained ordination (redundancy analysis, RDA) on this community data (OTU absolute frequency data) with all the environmental variables available. To test whether there was an effect of environmental variables on the tintinnid community compositions, we applied a Monte Carlo permutation test, using unconstrained permutations (null hypothesis: there is no effect of the environmental variables on species/OTU representation, i.e. the effect of variables is zero). We also performed a principal component analysis (PCA) with the same dataset including the community data (OTU absolute frequency data) to visualize how OTU occurrences correlated with the ordination of samples. The RDA, Monte Carlo permutation test and PCA were done using the software Canoco version 3.1 (Braak, 1988).

Acknowledgements

We thank the vessel crew from the Villefranche-sur-Mer oceanological observatory, in particular Jean-Yves Carval. We also thank to the SOMLIT group responsible for the environmental data collection in the Villefranche Bay. This work was funded by the ANR programs Biodiversity (ANR BDIV 07 004-02 'Aquaparadox') and Génomique microbienne à grande échelle (ANR-08-GENM-024-001, EVOLDEEP).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.

Bachvaroff, T.R., Kim, S., Guillou, L., Delwiche, C.F., and Coats, D.W. (2012) Molecular diversity of the syndinean genus *Euduboscquella* Based on Single Cell PCR. *Appl Environ Microbiol*.

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D (submitted). Accuracy of protist diversity assessments: morphology compared to cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *ISME J*.

Bachy, C., Gómez, F., López-García, P., Dolan, J.R., and Moreira, D. (2012) Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora). *Protist*.

Balech, E. (1959) Tintinnoinea del Mediterraneo. *Trab Inst Esp Oceanogr* 28: 1-88.

Bass, D., and Cavalier-Smith, T. (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Microbiol* 54: 2393-2404.

Braak, C.J.F.t. (1988) *CANOCO - a FORTRAN program for canonical community ordination by [partial] [etrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1)*. Wageningen: MLV.

Caron, D.A., Countway, P.D., Savai, P., Gast, R.J., Schnetzer, A., Moorthi, S.D. et al. (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* 75: 5797-5808.

Countway, P.D., Gast, R.J., Dennett, M.R., Savai, P., Rose, J.M., and Caron, D.A. (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). *Environ Microbiol* 9: 1219-1232.

Doherty, M., Costas, B.A., McManus, G.B., and Katz, L.A. (2007) Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. *Aquat Microb Ecol* 48: 141-154.

Doherty, M., Tamura, M., Costas, B.A., Ritchie, M.E., McManus, G.B., and Katz, L.A. (2010) Ciliate diversity and distribution across an environmental and depth gradient in Long Island Sound, USA. *Environ Microbiol* 12: 886-898.

Dolan, J.R., and Stoeck, T. (2011) Repeated sampling reveals differential variability in measures of species richness and community composition in planktonic protists. *Environmental Microbiology Reports* 3: 661-666.

Dolan, J.R., Lemée, R., Gasparini, S., Mousseau, L., and Heyndrickx, C. (2006) Probing diversity in the plankton: using patterns in tintinnids (planktonic marine ciliates) to identify patterns. *Hydrobiologia* 555: 143-157.

Dolan, J.R., Montagnes, D.J.S., Agatha, S., Coats, D.W., and Stoecker, D.K. (2012) *The Biology and Ecology of Tintinnid Ciliates*: John Wiley and Sons Ltd Wiley-Blackwell

Epstein, S., and López-García, P. (2008) "Missing" protists: a molecular prospective. *Biodivers Conserv* 17: 261-276.

Estrada, M., Henriksen, P., Gasol, J.M., Casamayor, E.O., and Pedrós-Alió, C. (2004) Diversity of planktonic photoautotrophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods. *FEMS Microbiol Ecol* 49: 281-293.

Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* 40: 237-264.

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

Jobb, G., von Haeseler, A., and Strimmer, K. (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4: 18.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059-3066.

Lara, E., Moreira, D., Vereshchaka, A., and Lopez-Garcia, P. (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonads. *Environ Microbiol* 11: 47-55.

Lara, E., Mitchell, E.A.D., Moreira, D., and López-García, P. (2011) Highly Diverse and Seasonally Dynamic Protist Community in a Pristine Peat Bog. *Protist* 162: 14-32.

Lepère, C., Boucher, D., Jardillier, L., Domaizon, I., and Debroas, D. (2006) Succession and Regulation Factors of Small Eukaryote Community Composition in a Lacustrine Ecosystem (Lake Pavin). *Appl Environ Microbiol* 72: 2971-2981.

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

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

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

Marande, W., Lopez-Garcia, P., and Moreira, D. (2009) Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. *Environ Microbiol* 11: 3179-3188.

Martin, D.P., Williamson, C., and Posada, D. (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260-262.

Massana, R., Balagué, V., Guillou, L., and Pedro's-Alió, C. (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol Ecol* 50: 231-243.

Maynard Smith, J. (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126-129.

Modigh, M., and Castaldo, S. (2002) Variability and persistence in tintinnid assemblages at a Mediterranean coastal site. *Aquat Microb Ecol* 28: 299-311.

Montagnes, D.J.S., and Lynn, D.H. (1987) A quantitative Protargol stain (QPS) for ciliates: Method description and test of its quantitative nature. *Mar Microb Food Webs* 2: 83-93.

Moon-van der Staay, S.Y., De Wachter, R., and Vaultot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409: 607 - 610.

Nebel, M., Pfabel, C., Stock, A., Dunthorn, M., and Stoeck, T. (2010) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. *Environmental Microbiology Reports*: no-no.

Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwalder, B., Boenigk, J., and Schlotterer, C. (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol Ecol* 19: 2908-2915.

Orsi, W., Song, Y.C., Hallam, S., and Edgcomb, V. (2012) Effect of oxygen minimum zone formation on communities of marine protists. *ISME J*.

Padidam, M., Sawyer, S., and Fauquet, C.M. (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218-225.

Pierce, R.W., and Turner, E. (1992) Ecology of Planktonic Ciliates in Marine Food Webs. *Rev Aquat Sci* 6: 139-181.

Posada, D., and Crandall, K.A. (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences* 98: 13757-13762.

Rassoulzadegan, F. (1979) Évolution annuelle des ciliés pélagiques en Méditerranée nord-occidentale. II. Ciliés oligotriches. Tintinnides (Tintinnina). *Invest Pesq* 43: 417-448.

Romari, K., and Vaultot, D. (2004) Composition and Temporal Variability of Picoeukaryote Communities at a Coastal Site of the English Channel from 18S rDNA Sequences. *Limnol Oceanogr* 49: 784-798.

Salani, F.S., Arndt, H., Hausmann, K., Nitsche, F., and Scheckenbach, F. (2012) Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. *ISME J* 6: 713-723.

Santoferrara, L.F., McManus, G.B., and Alder, V.A. (2012) Utility of Genetic Markers and Morphology for Species Discrimination within the Order Tintinnida (Ciliophora, Spirotrichea). *Protist*.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* 75: 7537-7541.

Sitran, R., Bergamasco, A., Decembrini, F., and Guglielmo, L. (2007) Temporal succession of tintinnids in the northern Ionian Sea, Central Mediterranean. *J Plankton Res* 29: 495–508.

Tamura, M., Katz, L.A., and McManus, G.B. (2011) Distribution and diversity of oligotrich and choreotrich ciliates across an environmental gradient in a large temperate estuary. *Aquat Microb Ecol* 64: 51-67.

Table 1- Summary of SSU rRNA gene sequences analyzed from the 8 tintinnid gene libraries and the associated number of clones, OTUs and coverage indices.

Sample	Depth (m)	Date	No. of clones	No. of OTUs	Coverage (%)
VilleFr-1	5	Feb 09	21	3	95
VilleFr-2	25	Feb 09	22	4	95
VilleFr-3	50	Feb 09	23	6	91
VilleFr-4	100	Feb 09	22	7	82
VilleFr-5	160	Feb 09	20	6	100
VilleFr-23	50	June 09	36	9	94
VilleFr-43	50	nov-09	30	9	83
VilleFr-58	50	Feb 10	24	5	92

Figure legends

Fig. 1. Maximum likelihood (ML) phylogenetic tree of SSU rDNA of tintinnids identified at different depths and seasons in the Villefranche Bay, based on 1,149 aligned positions. Numbers at nodes indicate bootstrap values. Sequences determined in this study are indicated in grey. Relative proportions of the different OTUs in each sample are indicated by circles of proportional size on the right. The percentage of sequences for the most abundant OTUs is indicated within the circles. The scale bar indicates the number of substitutions per site for a unit branch length.

Fig. 2. (A) Hierarchical clustering of samples using principal coordinate analysis (PCoA) (B) Hierarchical clustering of samples using UPGMA. Numbers at nodes indicate jackknife support values. The scale bar shows the distance between clusters in UniFrac units: a distance of 0 means that two environments are identical and a distance of 1 means that two environments contain mutually exclusive lineages.

Fig. 3. Maximum likelihood (ML) phylogenetic trees of the SSU+5.8S+LSU rDNA and ITS region showing the position of the tintinnid OTUs 3 and 5, based on 1,681 and 1,790 aligned positions, respectively. Numbers at nodes indicate bootstrap values. Sequences from this study are colored depending on their origin and date of sampling. Numbers of sub-clades cited in the text are indicated on the molecular phylogenies.

Fig. 4. SSU+5.8S+LSU rDNA + ITS based ML phylogenetic tree, based on 1,853 aligned positions, including all tintinnid environmental sequences retrieved in this study and a wide taxonomic representation of known tintinnid species. The colour code used in the tree topology highlights the date and geographical origin of the sequence. The taxonomic statuses are based on the SSU rDNA sequence identity and the sequences originating from isolated tintinnid specimens (Bachy et al., 2012). The two sequences showing incongruent positions are indicated with arrows.

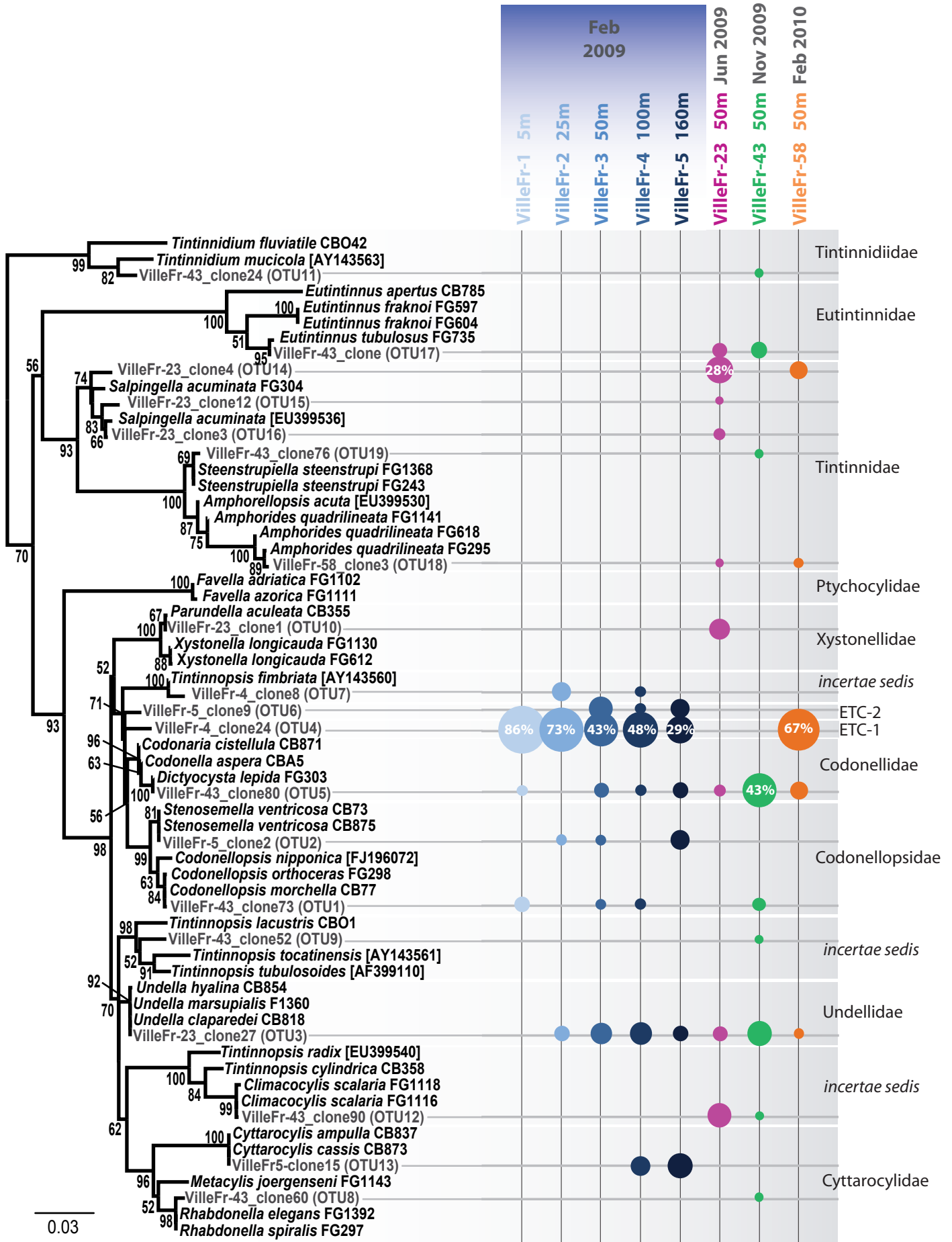
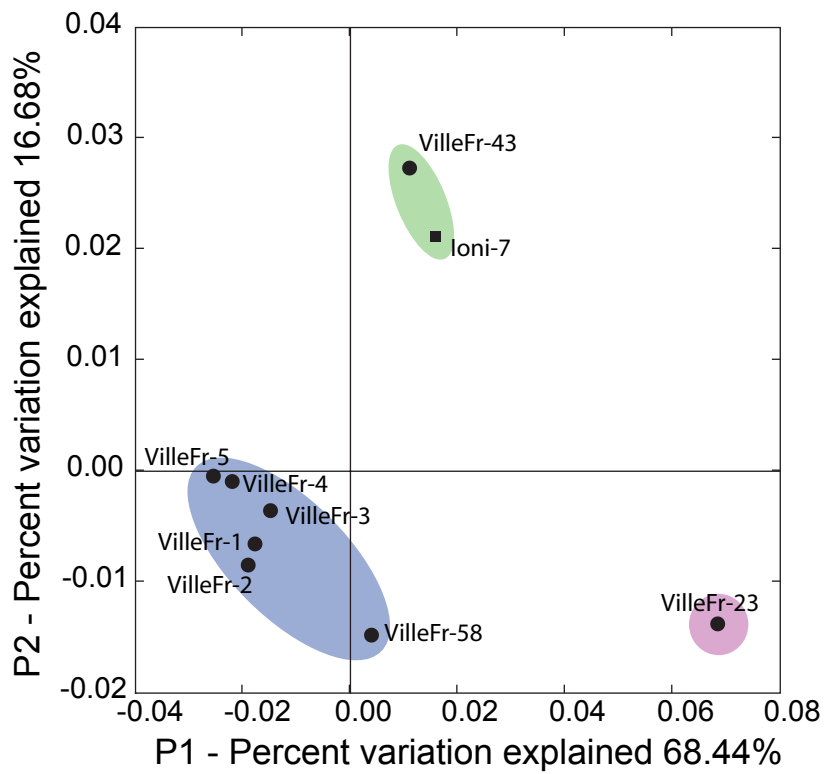


Figure 1, Bachy et al.

(A)



(B)

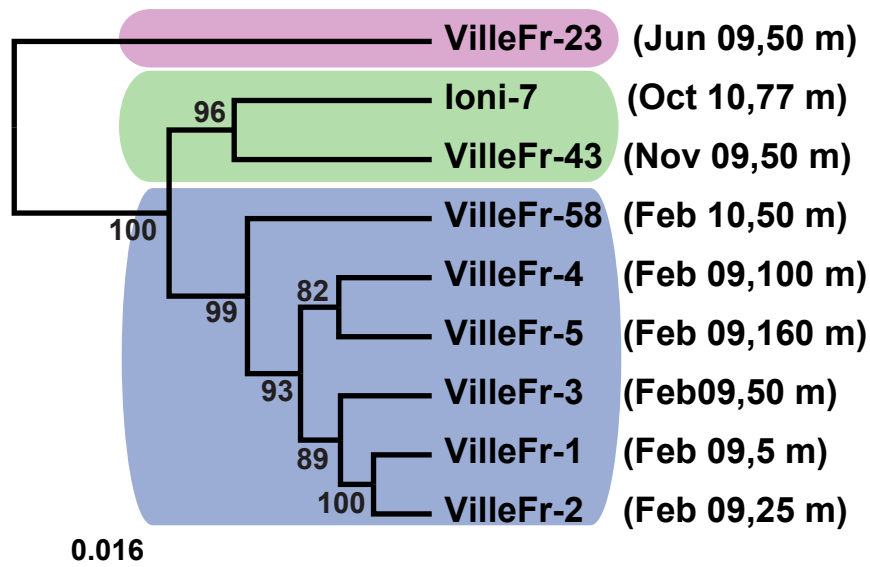


Figure 2, Bachy et al.

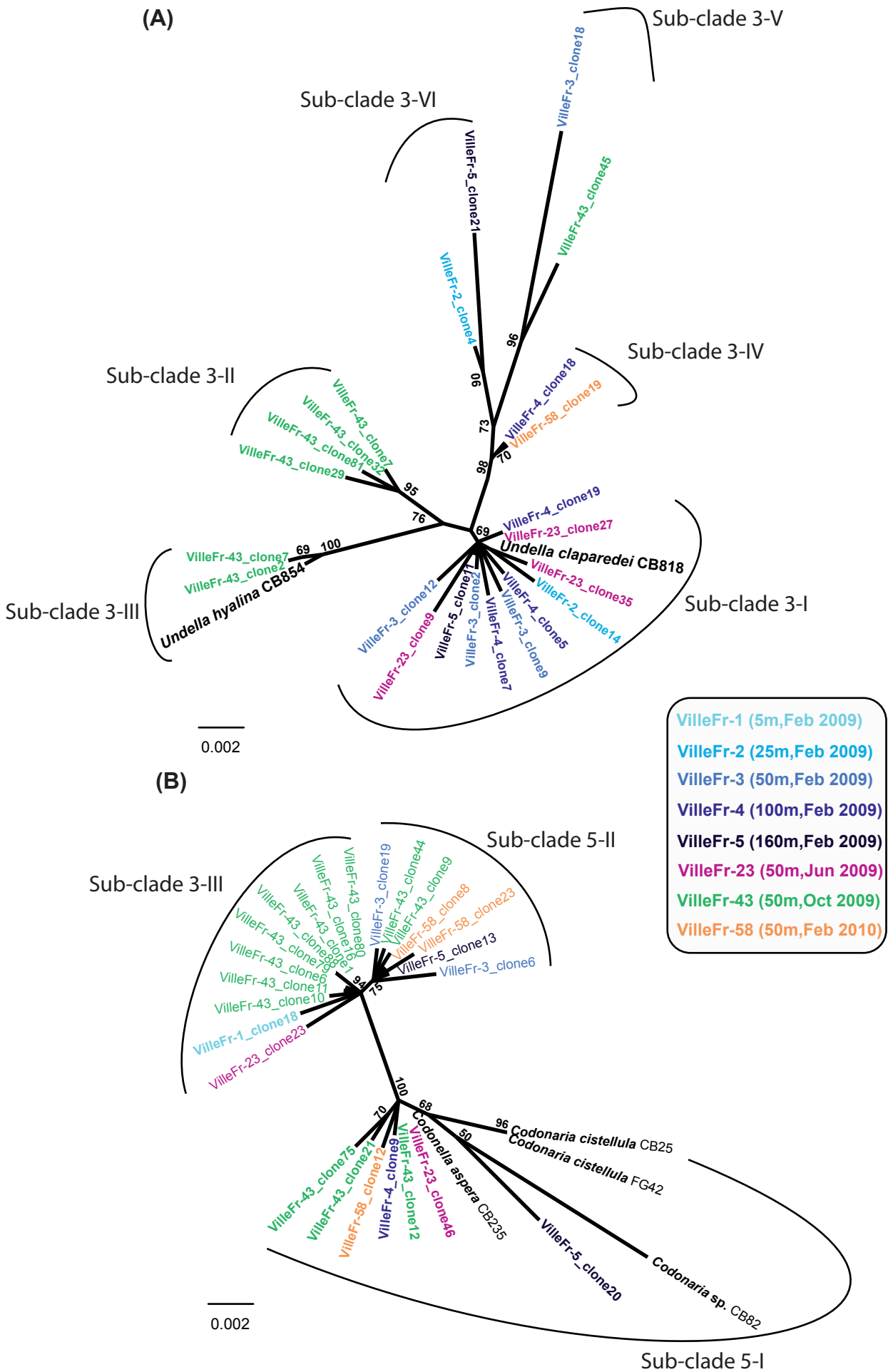


Figure 3, Bachy et al.

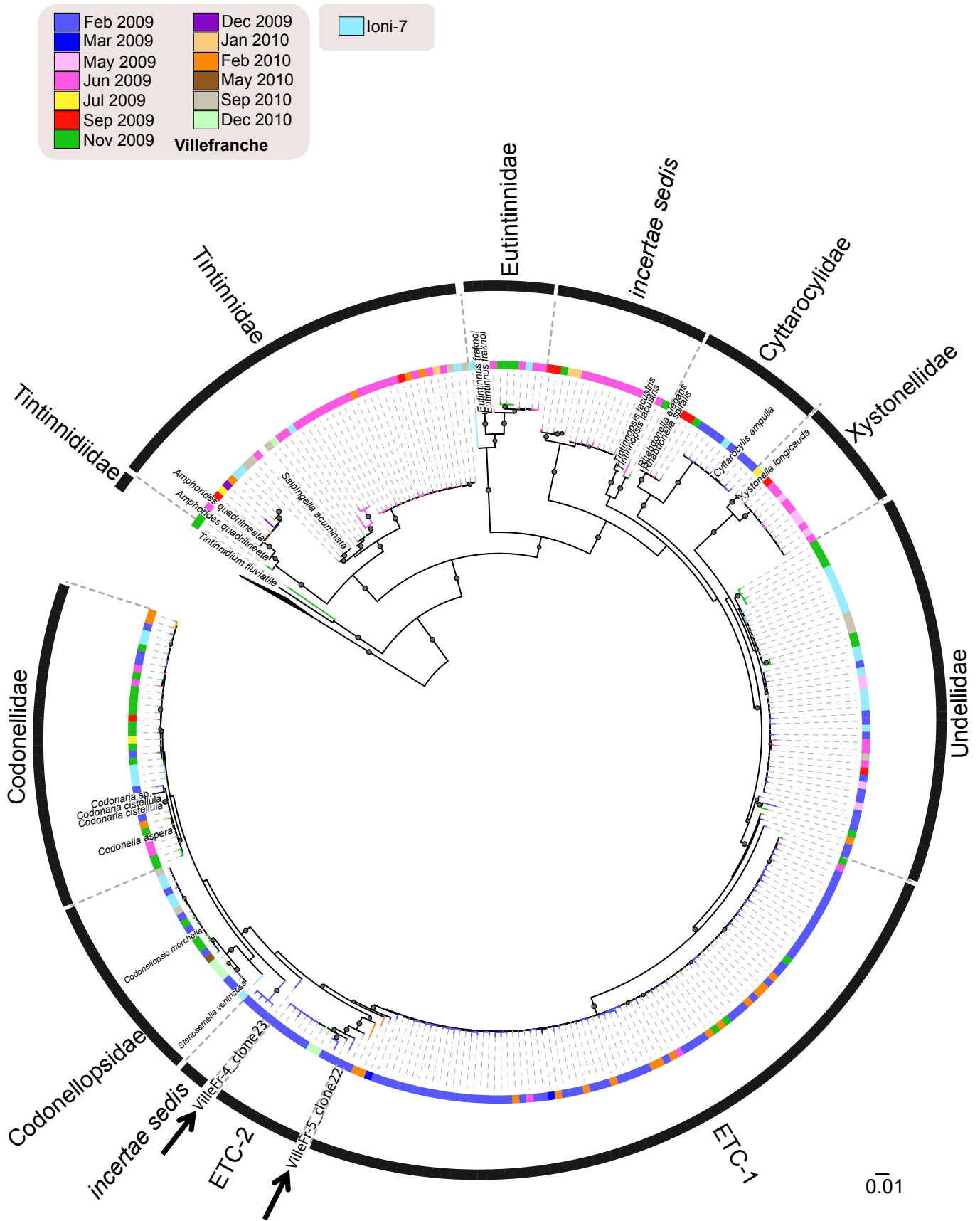


Figure 4, Bachy et al.

Table S1. List of samples collected in the Villefranche Bay during the temporal monitoring

Sample names	Depth (meters)	Date of sampling	Amplification and cloning
VilleFr-1	5	18 February 2009	x
VilleFr-2	25	18 February 2009	x
VilleFr-3	50	18 February 2009	x
VilleFr-4	100	18 February 2009	x
VilleFr-5	160	18 February 2009	x
VilleFr-6	5	18 March 2009	
VilleFr-7	25	18 March 2009	
VilleFr-8	50	18 March 2009	o (2)
VilleFr-9	100	18 March 2009	
VilleFr-10	160	18 March 2009	
VilleFr-11	5	17 April 2009	
VilleFr-12	25	17 April 2009	
VilleFr-13	50	17 April 2009	
VilleFr-14	100	17 April 2009	
VilleFr-15	160	17 April 2009	
VilleFr-16	5	27 May 2009	
VilleFr-17	25	27 May 2009	
VilleFr-18	50	27 May 2009	o (8)
VilleFr-19	100	27 May 2009	
VilleFr-20	160	27 May 2009	
VilleFr-21	5	25 June 2009	
VilleFr-22	25	25 June 2009	o (5)
VilleFr-23	50	25 June 2009	x
VilleFr-24	100	25 June 2009	o (6)
VilleFr-25	160	25 June 2009	o (5)
VilleFr-26	5	24 July 2009	
VilleFr-27	25	24 July 2009	
VilleFr-28	50	24 July 2009	o (3)
VilleFr-29	100	24 July 2009	
VilleFr-30	160	24 July 2009	
VilleFr-31	5	24 September 2009	o (1)
VilleFr-32	25	24 September 2009	o (2)
VilleFr-33	50	24 September 2009	o (5)
VilleFr-34	100	24 September 2009	o (1)
VilleFr-35	160	24 September 2009	
VilleFr-36	5	22 October 2009	
VilleFr-37	25	22 October 2009	
VilleFr-38	50	22 October 2009	
VilleFr-39	100	22 October 2009	
VilleFr-40	160	22 October 2009	
VilleFr-41	5	18 November 2009	
VilleFr-42	25	18 November 2009	
VilleFr-43	50	18 November 2009	x
VilleFr-44	100	18 November 2009	o (5)
VilleFr-45	160	18 November 2009	
VilleFr-46	5	16 December 2009	
VilleFr-47	25	16 December 2009	
VilleFr-48	50	16 December 2009	o (1)
VilleFr-49	100	16 December 2009	
VilleFr-50	160	16 December 2009	

VilleFr-51	5	27 January 2010	
VilleFr-52	25	27 January 2010	
VilleFr-53	50	27 January 2010	o (3)
VilleFr-54	100	27 January 2010	
VilleFr-55	160	27 January 2010	
VilleFr-56	5	24 February 2010	
VilleFr-57	25	24 February 2010	
VilleFr-58	50	24 February 2010	x
VilleFr-59	100	24 February 2010	
VilleFr-60	160	24 February 2010	
VilleFr-63	50	26 May 2010	o (1)
VilleFr-68	50	26 September 2010	o (11)
VilleFr-73	50	1 December 2010	o (6)

The crosses (x) indicated the 8 selected samples for which we generated well-covered clone libraries (see Table 1). The rounds (o) indicated the 16 samples for which we generated clone libraries with few sequences (number indicated in brackets).

Table S2. Closest tintinnid OTUs to sequences of ITS region fragments amplified from DGGE bands.

Band	1st hit	identity (%)	Taxonomy	Corresponding OTU
A	VilleFr-5_clone6	99.02	ETC-1	VilleFr-4_clone24 (OTU4)
B	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
C	VilleFr-5_clone6	99.02	ETC-1	VilleFr-4_clone24 (OTU4)
D	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
E	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
F	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
G	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
H	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
I	VilleFr-5_clone6	99.51	ETC-1	VilleFr-4_clone24 (OTU4)
J	VilleFr-5_clone11	99.51	<i>Undella claparedei</i>	Villefr-23_clone 27 (OTU3)
K	VilleFr-43_clone60	99.5	<i>Rhabdonella</i> sp.	VilleFr-43-clone60 (OTU8)
L	VilleFr-23_clone3	99.49	<i>Salpingella</i> sp.	VilleFr-23_clone3 (OTU16)
M	VilleFr-23_clone3	99.49	<i>Salpingella</i> sp.	VilleFr-23_clone3 (OTU16)
N	VilleFr-23_clone3	98.46	<i>Salpingella</i> sp.	VilleFr-23_clone3 (OTU16)
O	VilleFr-58_clone3	100	<i>Amphorides quadrilineata</i>	VilleFr-58_clone3 (OTU18)
P	VilleFr-23_clone8	99.5	<i>Parundella aculeata</i>	VilleFr-23_clone1 (OTU10)
Q	n/a	n/a	n/a	n/a
R	n/a	n/a	n/a	n/a
S	VilleFr-23_clone12	100	<i>Salpingella</i> sp.	VilleFr-23_clone12 (OTU15)
T	VilleFr-23_clone12	100	<i>Salpingella</i> sp.	VilleFr-23_clone12 (OTU15)
U	VilleFr-23_clone12	100	<i>Salpingella</i> sp.	VilleFr-23_clone12 (OTU15)

n/a : band which did not be sequenced

Table S3. Environmental parameters measured on the eight studied samples collected in Villefranche-sur-mer at point B+.

Sample	Depth (m)	Temperature (°C)	Fluorescence	PAR	Salinity (PSU)	Date
VilleFr-1	5	13.64	0.13	274.7	37.96	02/18/2009
VilleFr-2	25	13.64	0.18	30.86	37.96	02/18/2009
VilleFr-3	50	13.64	0.18	2.95	37.96	02/18/2009
VilleFr-4	100	13.72	0.16	1.20	37.99	02/18/2009
VilleFr-5	160	13.73	0.15	1.22	37.99	02/18/2009
VilleFr-23	50	14.71	0.19	23.9	37.94	06/25/2009
VilleFr-43	50	18.04	0.14*	8.38	37.98	11/18/2009
VilleFr-58	50	13.14	0.18	1.02	37.89	02/24/2010

*n/a, value estimated by intrapolation

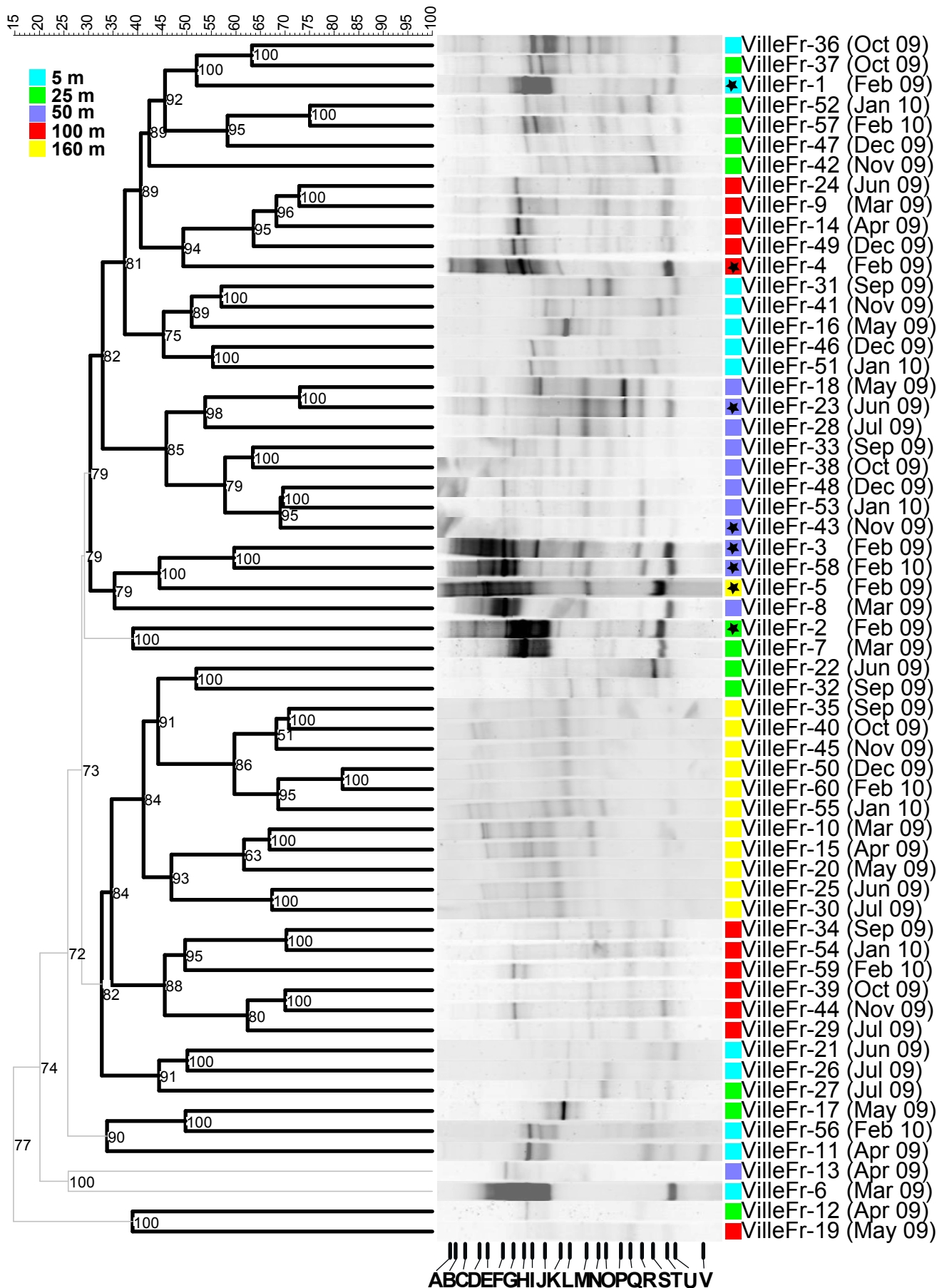


Figure S1. Clustering of DGGE profiles of tintinnid assemblages obtained at the Villefranche Bay at monthly intervals (February 2009 to February 2010) and at different depths (5 to 160 m). The name, month and depth of each sample are given on the right. The scale bar above the dendrogram shows distances (%) between samples based on presence/absence of bands. The numbers at the nodes correspond to cophenetic correlations. The most relevant clusters in the dendrogram as measured by the cluster cutoff method implemented in Bionumerics are highlighted with bold branches. Bands considered in the DGGE clustering analysis are indicated on the bottom side of the gel. These bands were cut off and sequenced from different samples (see the results in Table S3). The eight samples chosen for detailed molecular diversity analyses are labeled with a star.

PCA - P1 vs P2

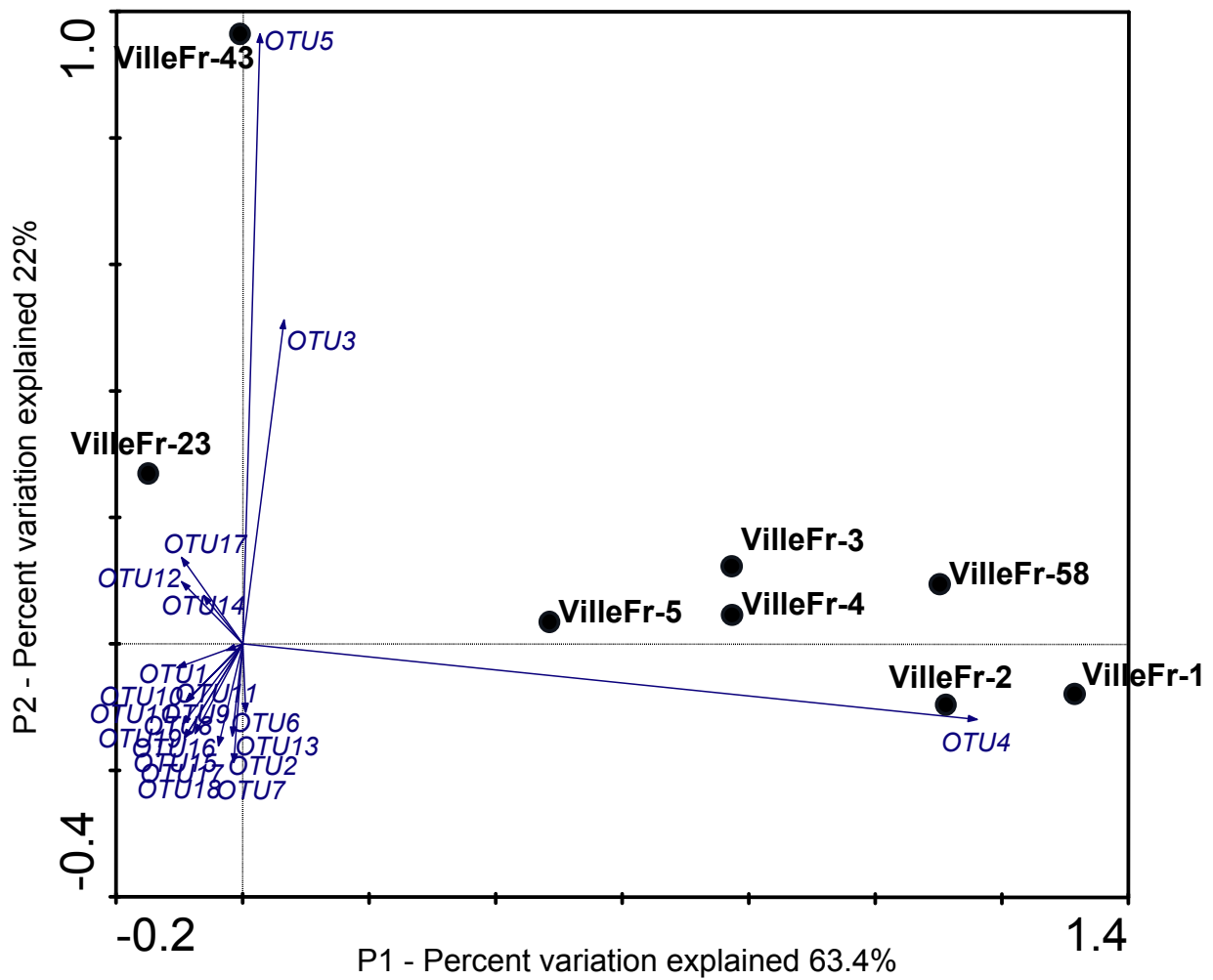


Figure S2. Principal component analysis (PCA) of tintinnid OTUs and the 8 selected samples from DGGE analysis for which the diversity was studied in greater depth. PC1 and PC2 are shown. The percentage of the variance in relative abundances of OTUs explained by each PC is shown on the axis. Each sample is represented by one black circle. Species are represented by arrows, and showed their relative contribution to the ordination subspace definition.

3. Synthèse et éléments de discussion

Les analyses de la dynamique temporelle des communautés tintinnidiennes de la rade de Villefranche par construction de banques de clones ont montré des changements dans leur composition entre les différentes saisons, avec par exemple des familles détectées uniquement en juin 2009, novembre 2009 et février 2010 mais pas en février 2009 (Tintinnidiidae, Eutintinnidae, Tintinnidae, Xystonellidae). On a également détecté des variations en termes d'abondance et de présence de certaines UTOs entre les différentes profondeurs en février 2009. De façon similaire à l'étude réalisée dans la partie IV.B, une grande diversité de familles de tintinnidés a été détectée par cette étude (9 familles). Cependant, contrairement à l'étude de la partie IV.B dans laquelle la majorité des séquences avaient été attribuée par comparaison à des séquences de tintinnidés de référence, ici, nous avons identifié deux UTOs qui n'ont pas pu être attribuées à des espèces morphologiquement décrites et qui en plus étaient très abondantes, regroupant la majorité des séquences des mois de février 2009 et 2010. Ces clades environnementaux (ETC-1 et ETC-2) branchent à la base des Codonellidae et des Codonellopsidae avec des supports faibles, mais ne sont groupées à aucune séquence de référence. Dans les études qui se concentrent sur des taxons spécifiques de protistes, notamment pour ceux qui n'ont pas de caractères morphologiques clairement distinguables, des séquences qui forment des groupes environnementaux sans représentants connus proches sont courantes. Par exemple, dans une étude de diversité moléculaire standard (amplification de l'ADNr 18S, clonage et séquençage Sanger) axée sur les cercozoaires, de nombreuses séquences n'ont pu être identifiées qu'à l'échelle du genre voire de la famille (Bass et Cavalier-Smith 2004). Toutefois, ce constat est plus étonnant dans notre cas des tintinnidés de la baie de Villefranche, du fait que ces ciliés ont une morphologie très distinctive, qu'ils ont été étudiés pendant des décennies sur ce site et que nous avons établi une phylogénie de référence pour les espèces connues sur place. Malgré notre échantillonnage extensif et varié de cellules individuelles (voir partie IV.A), il n'est donc pas possible à l'heure actuelle de rapprocher ces séquences avec des séquences de référence, et donc d'identifier clairement l'organisme d'origine. Cette absence d'identification est d'autant plus étonnante que ces séquences environnementales sont très abondantes à différentes profondeurs et de manière récurrente en hiver. À l'inverse, certaines espèces communément rencontrées dans notre étude sur cellules isolées (comme les *Favella* spp. par exemple) n'ont pas été détectées dans cette étude environnementale.

La comparaison de la composition des communautés de tintinnidés des différents échantillons semblerait révéler une structuration des communautés en fonction de la profondeur dans la colonne d'eau (différemment en fonction des UTOs), mais aussi et surtout de la saisonnalité. En effet, même si notre étude s'est limitée à deux ans, la composition des communautés des tintinnidés a tendance à être plus similaire dans les échantillons prélevés durant la même saison quelque soit l'année et quelque soit le site d'échantillonnage. Dans notre analyse comparative, la communauté de l'échantillon VilleFr-43 est plus proche de celle de l'échantillon Ioni-7, collecté à plus d'un millier de kilomètres du site de Villefranche et un an plus tard, mais toujours en automne (voir partie IV.B), que des autres communautés de Villefranche du reste du suivi. Je n'ai pas pu observer de corrélation entre les variables environnementales abiotiques disponibles et les communautés de tintinnidés, ce qui semble indiquer que d'autres paramètres différents de ceux mesurés déterminent la structure de ces communautés de ciliés. Toutefois, savoir quels sont les facteurs environnementaux qui conditionnent la coexistence de différentes espèces nécessiterait de collecter un plus grand nombre de paramètres environnementaux, notamment des paramètres biologiques.

Nous avons profité de la présence lors de différentes saisons et à différentes profondeurs de quelques UTOs (UTO3 et UTO5) pour nous intéresser de manière plus spécifique à la diversité au sein de ces UTOs en prenant en compte non seulement l'ADNr 18S mais également la région des ITS. Les séquences de ces UTOs groupent phylogénétiquement en différents clades, indépendamment de la saison ou de la profondeur d'échantillonnage. Et une partie de ces clades est identifiée à l'aide des séquences de référence comme des espèces morphologiquement différentes. Une autre partie de la diversité ne groupe avec aucune séquence de référence. Si l'ADNr 18S est relativement un bon marqueur pour distinguer le genre et souvent les espèces, dans certains cas, des marqueurs moins conservés semblent mieux indiqués pour travailler à fine échelle taxonomique dans les études environnementales.

Enfin, savoir si cette diversité détectée au sein des UTOs est corrélée à de la diversité inter-, intra- spécifique ou encore intra-individuelle n'est pas clair. Cela ouvre des perspectives du point de vue de la diversité à l'échelle intra-spécifique : il faut approfondir l'étude de la diversité génétique au sein d'une même espèce de tintinnidé, et définir quelle part est liée à la diversité entre espèces du même genre, entre individus de la même espèce et entre marqueurs du même individu.

D. Diversité temporelle intra-spécifique chez l'espèce *Undella claparedei*

1. Contexte et Objectifs

Les études de diversité intra-spécifique peuvent fournir des informations importantes sur la diversité génétique au sein des populations si des marqueurs moléculaires adéquats sont utilisés. Cependant, dans l'objectif d'interpréter une telle information, il convient de distinguer la diversité taxonomique cryptique de la structuration des populations d'une même espèce, puisque les deux peuvent se confondre.

Traditionnellement, la diversité et la distribution des protistes ont été étudiées par observation morphologique. Cependant, nous avons vu qu'il existe de nombreuses espèces cryptiques qui partagent une morphologie similaire mais sont extrêmement divergents génétiquement. Ceci constitue une limitation non seulement dans la description de la composition en espèces des communautés, mais également dans la détermination de la diversité intra-spécifique. Les études de diversité intra-spécifique réalisées qui comportent une dimension spatiale et/ou temporelle peuvent informer sur la diversité génétique au sein et entre les populations et éventuellement, relier les changements populationnels observés à la présence de divers écotypes par exemple. Pour prospecter dans ce sens, je me suis intéressé à l'espèce de tintinnidé *Undella claparedei*, aisément identifiable morphologiquement, présente dans les échantillons de plancton de Villefranche de façon récurrente tout au long du suivi temporel effectué et à des profondeurs différentes. Savoir identifier les bases génétiques et les mécanismes évolutifs permettant à des populations de se maintenir à l'échelle locale est une question importante, surtout dans les environnements marins aux conditions environnementales changeantes. Le premier objectif a donc été de caractériser la diversité génétique et son évolution au cours du temps dans plusieurs populations naturelles d'*U. claparedei*. Pour cela, nous avons disposé d'un lot d'individus échantillonnés pendant deux ans dans différentes populations séparées dans le temps d'un mois ou plus. À l'aide d'un jeu de deux marqueurs (ADNr 18S et ITS), nous avons cherché à caractériser la structure génétique des populations de cette espèce au cours du temps et à détecter d'éventuelles traces de recombinaison. Ces analyses initiales de génétique des populations et l'observation minutieuse de la diversité sur les différents marqueurs permettront de faire des inférences sur la dynamique évolutive de ces populations.

2. Manuscrit de l'article 4 : Temporal patterns of polymorphism at 18S rRNA genes and internal transcribed spacers in populations of *Undella claparedei* (Tintinnida) from Villefranche Bay, France

Article in preparation

Temporal patterns of polymorphism at 18S rRNA genes and internal transcribed spacers in populations of *Undella claparedei* (Tintinnida) from Villefranche Bay, France

Charles Bachy¹, David Moreira¹, John R. Dolan², Purificación López-García^{1*}

1 Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, Bâtiment 360, 91405 Orsay Cedex, France; charles.bachy@u-psud.fr, david.moreira@u-psud.fr, puri.lopez@u-psud.fr

2 Université Pierre et Marie Curie and Centre National de la Recherche Scientifique (CNRS), UMR 7093, Laboratoire d'Océanographie de Villefranche, Marine Microbial Ecology, Station Zoologique, B.P. 28, 06230 Villefranche-sur-Mer, France; dolan@obs-vlfr.fr

*Corresponding author; fax +33169154697

e-mail (puri.lopez@u-psud.fr)

Summary

The existence of high levels of genetic diversity behind apparent similar morphologies has been shown to be relatively common among microbial eukaryotes, with cases of cryptic species complexes. In certain cases, studies of the geographical component of genetic variation at a population level have demonstrated that the genetic diversity may be linked to geographical patterns. By contrast, the temporal structuring of genetic diversity below the species level remains poorly documented. To address this question, we have carried out genetic diversity analyses of populations of the tintinnid ciliate *Undella claparedei*, which is easily distinguishable in marine plankton samples by its conspicuous morphology. Over a period of almost two years, we regularly sampled these tintinnid populations in the Bay of Villefranche-sur-Mer (Mediterranean Sea). We amplified by PCR and sequenced from single cells two nuclear loci, the small subunit rRNA gene -18S rDNA- and the adjacent internal transcribed spacer (ITS). We identified three and two different sequence types for the 18S rDNA and the ITS, respectively, which did not appear to correlate with any observable morphological variation. Some individuals contained a mix of 18S rDNA or ITS sequences of several types, suggesting the occurrence of sexual reproduction and hybridization, with several cases of recombination detected at both the 18S rDNA and ITS sequences. The 18S rDNA appeared to be more resolute than the ITS, allowing to detect genetic differentiation between *U. claparedei* populations during the 2-year sampling. We tentatively explained this 18S rDNA genetic differentiation over time by the immigration in the local population of the Villefranche Bay of exogenous individuals of novel sequence types, which might have spread in the population by sexual reproduction.

Introduction

Traditionally, the diversity and distribution of protists (or unicellular eukaryotes in general) have been studied using classical morphological examination. However, the existence of cryptic species, i.e. species sharing similar morphologies but being genetically divergent has been shown to constitute a major limitation of this approach, not only to estimate the community species composition but also the intraspecies diversity. In fact, the advent of molecular techniques into this field during the last decade has revealed that behind identical or slightly different morphologies, several morphospecies are actually composed of several genetically distinct clades (e.g. Catania et al 2009, Lowe et al 2005, Montresor et al 2003, Slapeta et al 2006). Whereas some of those clades appear to be cosmopolitan, others show ecologically and/or geographically restricted distributions (Darling and Wade 2008, de Vargas et al 1999, Rodriguez et al 2005). Moreover, part of the detected cryptic diversity most likely represents distinct biological species, as attested by genetic evidence for reproductive isolation (Amato et al 2007). For instance, among photosynthetic picoeukaryotes, some cryptic species of prasinophyte algae have diverged for several millions of years (Slapeta et al 2006) and others are highly differentiated in their protein coding sequences (Piganeau et al 2011).

Yet, this field remains poorly explored and information about levels of genetic diversity within populations of members of the same morphospecies is limited. Published studies have most often been conducted to study biogeographical questions at a very large geographical scale but, by definition, biogeography is the study of biodiversity not only over space but also through time (Martiny et al 2006). However, with the exception of a few studies focusing on dinoflagellates (Shankle et al 2004) or diatoms (Casteleyn et al 2009b, Evans et al 2005), the temporal component of genetic variation at a population level has rarely been taken into account in the study of free-living protists. In addition, an ecological bias affected previous protist biogeography studies since most of them concerned autotrophic species rather than heterotrophic ones, although the latter are much more diverse than the autotrophic species (Vaulot et al 2002). This bias is partially due to the fact that many of those heterotrophs are difficult to culture in the laboratory for subsequent molecular analyses. Nevertheless, cultivation biases can be overcome at present by the technique of single-cell isolation from

natural samples and direct PCR amplification and sequencing of suitable molecular markers. Recently, this approach has been applied to ciliate protists, which are major consumers of phytoplanktonic biomass in coastal waters, to address phylogenetic and diversity assessment questions (e.g. Bachy et al 2012, Duff et al 2008, Snoeyenbos-West et al 2002). As ciliates possess many copies of their genome (Prescott 1994), the amplification of molecular markers is facilitated and does not need whole genome amplification or cultivation steps. This provides a rapid way to collect molecular data from large numbers of individuals, opening the possibility of carrying out population genetics studies.

The species-rich order Tintinnida groups free-living ciliate species which are easily identifiable by the presence of an external shell, named lorica, which allows species identification under the microscope (e.g. Dolan and Stoeck 2011). In fact, even if the morphology of the lorica can be highly variable in some species (Laval-Peuto and Brownlee 1986), it remains a discriminative criterion to identify many tintinnid species (Bachy et al 2012). Among them, *Undella claparedei* (Entz) Kofoid and Campbell is a tintinnid species restricted to warm-temperate marine waters, commonly found in the Mediterranean Sea at different depths and different periods of the year with fluctuating abundances (Balech 1975, Gomez and Gorsky 2003, Kršinić and Grbec 2006); Bachy et al., in prep, see section IV.C). We have used this species as a model to study genetic variation in a protist species over time at a small spatial scale in a coastal marine environment. Thus, we carried out a two-year monitoring of the genetic diversity of *U. claparedei* in the bay of Villefranche-sur-Mer (France) using genetic markers (18S RNA gene and ITS regions). Its abundance at this site makes this easy-to-recognize species an appropriate model for population genetic studies. In addition to estimating the degree of intraspecific genetic diversity in *U. claparedei*, our survey allowed us to address the question of whether temporally separated populations from the same location were genetically different, namely if the apparent continuous presence of this morphospecies actually reflects a succession of different genotypes.

Materials and Methods

Sampling and single-cell isolation

Specimens of the planktonic tintinnid *Undella claparedei* were collected over a period of two years in the northwestern Mediterranean Sea (Ligurian Sea) at the Bay of Villefranche-sur-

Mer. Sampling was performed at the site 'Point B+' (43°41'00"N, 7°18'44"E; water column depth ~200 m), located at the entrance of the bay, which is open towards the sea and is almost not protected from the wind (Nival and Corre 1976). During our sampling period, temperature-driven stratification occurred in the water column from April to October in both 2009 and 2010. Sampling was carried out monthly with a custom-made conical phytoplankton net (53 µm mesh size, 54 cm diameter and 280 cm length) in double oblique angle across the 0-50 meters layer from February 2009 to February 2010. We also obtained additional samples in May 2010, September 2010 and December 2010 (see Table 1). Planktonic cells were concentrated in Utermöhl chambers and individual *U. claparedei* specimens were isolated and photographed using an inverted microscope (Olympus IX51) equipped with an Olympus DP71 digital camera. *U. claparedei* was easily identified based on the morphological characteristics of its bowl-shape hyaline lorica with a suboral thickening. We measured for each specimen typical morphometric data (oral aperture diameter and the lorica length) with the ImageJ software, and noticed when some loricae presented subtle modifications such as a slightly pointed aboral end, a longer shape or an apparent aberrant morphology (Fig.1). The isolation of all specimens was done by the same researcher (CB) over the complete sampling season to avoid researcher-induced taxonomical identification bias. When *U. claparedei* was highly concentrated in the planktonic material, between 50-60 specimens were usually isolated for sequencing. When the species was more scarcely observed under the microscope, reflecting its lower abundance in the natural field, we isolated all the specimens encountered and rarely reached the number of 10 individuals per sampling, even after several hours of observation. After isolation, each cell was individually rinsed twice in two droplets of 0.2 µm-filtered and sterilized seawater. The isolated washed cells were placed in 0.2 ml Eppendorf tubes filled with several drops of absolute ethanol. The tubes were stored at 4 °C or at room temperature and in darkness until the molecular analysis could be performed.

PCR amplification of small subunit rRNA genes and internal transcribed spacer (ITS) region

The specimens fixed in ethanol were centrifuged gently for 5 min at 3,000 rpm. Ethanol was then evaporated in a vacuum desiccator and single cells were resuspended directly in 25 µl of Ex TaKaRa buffer (TaKaRa, distributed by Lonza Cia., Levallois-Perret, France). PCR reactions were done in a volume of 30 µl reaction mix containing 10-20 pmol of the forward

eukaryotic-specific 18S rDNA primer EK-42F (5'-CTCAARGAYTAAGCCATGCA-3') and the reverse eukaryotic large subunit rDNA (28S rDNA) primer EK-28S564R (5'-TGGTCCGTGTTTCRAGACG-3') (López-García et al 2001). The PCR reactions were performed under the following conditions: 2 min denaturation at 94 °C; 10 cycles of 'touch-down' PCR (denaturation at 94 °C for 15 s; a 30 s annealing step at decreasing temperature from 65 down to 55 °C employing a 1 °C decrease with each cycle, extension at 72 °C for 2 min); 20 additional cycles at 55 °C annealing temperature; and a final elongation step of 7 min at 72 °C. Two sets of internal primers were used separately to specifically target the tintinnid 18S rDNA and ITS region (ITS-1+5.8S rDNA+ITS-2) sequences. A first nested PCR reaction was carried out using 3 µl of the first PCR reaction diluted to 1/10 in a Gotaq (Promega, Lyon, France) polymerase reaction mix containing the tintinnid-specific primer 18S-Tin3F (5'-GCGGTATTTATTAGATAWCAGCC-3') and the eukaryotic-specific SSU-rDNA primer EK-1379R (5'-TGTGTACAAAGGGCAGGGAC-3'). A second PCR reaction was carried out using 3 µl of the first PCR reaction diluted to 1/10 in a Gotaq (Promega, Lyon, France) polymerase reaction mix containing the eukaryotic-specific 18S rDNA primer EK-1520F (5'-GTAGGTGAACCTGCRG-3') and the tintinnid-specific 28S rDNA primer 28S-TinR1 (5'-TGGTGCACTAGTATCAAAGT-3') (Bachy et al 2012). Fifty PCR cycles (denaturation at 94 °C for 15 s; annealing at 55 °C for 30 s and extension at 72 °C for 1 min) were performed, with an initial denaturation step at 94 °C for 3 min and a final elongation step at 72 °C for 7 min for the PCRs with the two different sets of primers. Amplicons of the expected size (1150 bp for the 18S rDNA fragments and 750 bp for the ITS region fragments) were then entirely sequenced using the forward and reverse amplification primers with an automated 96-capillary sequencer ABI PRISM 3730xl (Beckman Coulter Genomics, Takeley, U.K). Sequences were assembled using the CodonCode Aligner 3.7 program (CodonCode Corporation, USA) and all the chromatograms were manually checked to detect possible sequencing errors. The non-redundant sequences reported in this study have been deposited in the GenBank database (accession numbers AA00000-00).

Detection of sequence polymorphisms

Some polymorphic sites can be detected by examining the high-quality chromatograms searching for double overlapping peaks observed in the both sequencing directions. As we

believed that these double peaks indicated the presence of true intra-individual sequence diversity we kept them in the analyses as degenerated base pairs. All the high-quality 18S rDNA and ITS region sequences were aligned using MAFFT (Kato et al 2002), respectively generating alignments of 662 bp for the 18S rDNA sequences and of 437 bp for the ITS region sequences (which covered the complete 5.8S rDNA, the complete ITS-2 and a fragment of the 28S rDNA). Single nucleotide polymorphisms (SNPs) were identified by comparison of the aligned sequences. All polymorphic sites observed were re-checked on the chromatograms of both the reverse and forward strands. Unique mutations in the 18S rDNA or ITS region sequences (i.e., polymorphisms found only once in the total database) were removed from subsequent analyses to avoid any possible PCR and/or sequencing errors. This was a very conservative approach and we admitted that a possible loss of diversity was introduced in this way. Finally, no indels were found but only single base substitutions.

Statistical analyses

We carried out an Analysis of Molecular Variance (AMOVA; Excoffier et al 1992) using the 18S rDNA and ITS region sequences (for which polymorphic sequence types could be resolved into two sequence types per individual). True heterozygotes could not be distinguished from intragenomic variation, so, we calculated the pairwise fixation index (Φ_{ST}) and population pairwise F_{ST} values, and their significance in order to estimate population differentiation. Within- and between-population variation in sequence types was calculated with Arlequin 3.0 (Excoffier et al 2005) for both the 18S rDNA and the ITS region datasets using AMOVA with 1000 permutations. To study the effect of the sampling date on the genetic diversity of *U. claparedei* we firstly considered the specimens collected at each sampling date as a population (only populations with at least 10 individuals were taken into account). In a second time, to study the possible effect of environmental variables on population structure and also to take into account all the specimens that we collected, we subdivided our total set of individuals into two populations. The first one grouped all the individuals collected during months characterized by a well-mixed water column and the second one grouped all the individuals collected during months characterized by a stratifying or fully stratified water column.

Sequence secondary structure prediction

Predictions of the ITS-2 secondary structures were obtained using MFOLD with default parameters (Zuker 2003). The exact borders of the ITS-2 region were identified based on the base pairing between the end of the 5.8S and the start of the 28S rDNA (Keller et al 2009). To have a reference structure, we searched for existing secondary structures already defined for related species within ciliates in the ITS-2 database using a hidden Markov model, but it did not return any match (Koetschan et al 2012). Therefore, we manually checked the results of folding programs to select a common structure among our ITS-2 types which presented homology with the universal structure typical of eukaryotes (Schultz et al 2005). Then, the base changes due to sequence polymorphisms were mapped on the predicted structures.

Phylogenetic analyses

For all the specimens for which we successfully sequenced both the 18S rDNA and ITS region, we concatenated the sequences of both markers in order to increase the number of polymorphic sites and the phylogenetic signal. This approach gave us the opportunity to test for recombination events and assess the phylogenetic relationships. We used the SplitsTree4 package (Huson and Bryant 2006, available at <http://www.splitstree.org>) with the neighbour-net algorithm (Bryant and Moulton 2004) with default parameters, to infer the phylogenetic network of representative sequences of our concatenated dataset, together with a reference sequence of *Undella claparedei* specimen CB818 (#JQ408164) and a set of 18 environmental sequences clearly identified as *Undella* species based on their 18S rDNA sequence identity, and also covering the ITS region. The 18S rDNA+ITS phylogeny was inferred under maximum likelihood (ML) from that sequence dataset, to which we added previously published sequences of other tintinnid taxa (Bachy et al 2012). The ML analysis was performed using TREEFINDER (Jobb et al 2004). The J1 model of nucleotide substitution was selected as the best-fit evolutionary model applying the Akaike Information Criterion with correction for small sample size (AICc, Posada and Buckley 2004) using TREEFINDER. Non-parametric bootstrapping (1,000 replicates) was employed to evaluate nodal support among clades. The 50% consensus tree was visualized and edited with FigTree (Rambaut 2009).

We used several recombination detection methods to investigate recombination events from the concatenated 18S rDNA+ITS sequence dataset as it has been shown that conclusions about the presence of recombination should not be based on a single method (Posada and Crandall 2001, Posada 2002). We thus applied the maxchi and geneconv tests (Maynard Smith 1992, Sawyer 1989) using RecombiTEST (Piganeau et al 2004). We also applied the LDr^2 and the $LD|D'|$ tests which assessed the significance of the relationship between the measure of linkage disequilibrium, respectively r^2 (Hill and Robertson 1968) and $|D'|$ (Lewontin 1964), and distance between sites. 1000 randomizations were performed to assess significance for each test.

Results

During almost two years (from February 2009 to December 2010), we regularly collected plankton samples in the Bay of Villefranche-sur-Mer at twelve dates (Table 1). From these samples, we isolated a total of 417 specimens of the tintinnid ciliate *U. claparedei* (Fig. 1). Depending on the abundance of this species in the different samples, we collected between 7 and 70 individuals (Table 1). Among the 417 isolated specimens, we obtained 18S rRNA gene sequences for 156 (37.4%) of them, and of the ITS region for 154 (36.9%) of them (Table 1). Both markers were successfully sequenced for 79 (18.9%) of the collected individuals. We first examined the genetic diversity in the *U. claparedei* populations for the two molecular markers taken separately and then for the fraction of cells for which both markers were sequenced.

18S rDNA and ITS region sequence diversity

Three different 18S rDNA sequence types were detected from direct sequencing of the 662 bp PCR products obtained from isolated *U. claparedei* cells: type A (114 individuals), type B (18 individuals), and type C (1 individual) (Fig. 2). Some sequences had SNPs that showed double peaks in the chromatograms, which we interpreted as the simultaneous presence of several sequence types in the same individual. Thus, three different mixings of 18S rDNA sequence types were detected: type A/B (16 individuals), type A/C (1 individual) and type

A/B/C (6 individuals). For example, the mixed type A/B corresponded to individuals that had sequences which showed overlapping patterns of both 18S rDNA types A and B. In the case of the 18S rDNA mixed type A/B/C, our data did not allow establishing if this type resulted either from the simultaneous presence of the three different types A, B and C in the same individual or of the type A with a fourth type that would combine the patterns B and C but for which we did not have any reference sequence in our dataset. The 18S rDNA types showed a sequence divergence on the 662-bp 18S rDNA fragment studied ranging between 0.3% (A versus B) and 0.9% (B versus C) (Figure 2). In the case of the ITS region, two different sequence types were obtained: type A (132 individuals) and type B (9 individuals). Similarly, we also detected 13 individuals which presented an ITS mixed type A/B, namely showing overlapping patterns of both ITS types A and B. The ITS types A and B showed a sequence divergence on the 437-bp ITS region fragment studied of 0.9% (Figure 2), but a 2.1% sequence divergence considering the ITS-2 alone (189 bp length).

The specific nucleotide differences between the different sequence types are illustrated in Figure 2. 18S rDNA type A differed from type B in SNPs at positions 331 (A versus G) and 545 (C versus T) and from type C in SNPs at positions 296 (C versus G), 304 (G versus T), 306 (T versus C) and 353 (A versus G). An overlapping pattern between 18S rDNA types A and B was evident in the two SNPs at positions 331 and 545 in a number of individuals and was referred as to 18S rDNA mixed type A/B. The sequences corresponding to the 18S rDNA mixed type A/C showed overlapping patterns between the 18S rDNA types A and C in the SNPs at positions 296, 304, 306 and 353, whereas the sequences of the 18S rDNA mixed type A/B/C showed overlapping patterns between the three 18S rDNA types A, B and C in the SNPs at positions 296, 304, 306, 331, 353 and 545. Regarding the ITS types, the type A differed from the B in SNPs in the ITS-2 region at positions 190 (C versus T), 215 and 217 (G versus A), and 228 (T versus C). We did not identify any SNP in the 5.8S rRNA or 28S rRNA regions that we sequenced together with the ITS-2. The overlapping pattern between ITS types A and B to generate the mixed type A/B was clearly observed in the four SNPs at positions 190, 215, 217 and 228.

ITS-2 secondary structure

The ITS-2 secondary structures of the types A and B revealed that the 4 SNPs that differentiate the two sequence types were involved in non-compensatory changes (i.e., base pair ↔ non-pair), which resulted in internal bulges on two different helices and introduced a spacer of one nucleotide between the helices II and III. The SNPs at positions 190, 215 and 217 were located on the stem of helix II while the SNP at position 228 was located at the basis of the stem of helix III (Fig. 3).

Morphological characteristics of different genotypes

All our specimens exhibited the very typical round-shaped lorica specific of *U. claparedei* (Fig. 1A) but with minute variations in shape. The length of the lorica from the aboral end to the oral collar end ranged from 76.6 to 84.5 μm and that of the oral aperture diameter from 37.7 to 41.5 μm. The results of morphometric analysis of the oral aperture obtained from specimens belonging to all the 18S rDNA and ITS types did not reveal any significant differences. Moreover, some relatively divergent morphotypes that displayed a minute pointed aboral end (e.g. specimens CB536 and CB679, Fig. 1B), a slightly longer lorica (specimen CB760, Fig. 1C) or a peculiar lorica with irregular form (specimen CB695, Fig. 1D) were all them of type A for both the 18S rDNA and the ITS. Consequently, we did not observe any consistent morphological difference that distinguished the different 18S rDNA or ITS types, even those that occurred at low frequencies in our samples or that have the most atypical morphologies.

Analysis of Molecular Variance (AMOVA) of sequence types distribution

A simple projection of the absolute frequencies of sequence types (Fig. 4) showed that the 18S rDNA types A, B and A/B co-occurred all along the sampling period from February 2009 to December 2010. Among them, the 18S rDNA type A was the most abundant, averaging 72.5% of the total 18S rDNA sequences per sampling. We observed the occurrence of 18S rDNA type C only since December 2009 and then the co-occurrence of types A/C and A/B/C until December 2010. The distribution of ITS sequences showed a co-occurrence of A, B and A/B types all along the survey with, once again, the dominance of one type (A) for the whole

duration of the temporal monitoring (85.4% on average). Despite the clear dominance of certain types, it is important to note that we generally detected more than one type for both markers (18S rDNA and ITS). This is also true for the ITS even in the populations for which we obtained the smallest number of individuals (less than 18 individuals in March 2009, May 2009, November 2009, and May 2010).

Analyses of Molecular Variance (AMOVAs) were carried out for each molecular marker independently with two types of data subdivision: subset 1 contained a restricted set of the populations sampled along our temporal survey including only those populations with 10 individuals or more, whereas subset 2 contained all individuals arbitrarily classified into two clusters depending on the water column conditions at the moment of the sampling (i.e., mixed vs. stratified). For subset 1, the 18S rDNA genetic variation was hierarchically structured, with about 12.74% of the variation distributed among populations, and 87.26% within populations. Fixation indices showed moderate but significant genetic structuring at the between-populations level ($\Phi_{ST} = 0.127$, P-value < 0.001) (Table 2). The analysis of pairwise distances (F_{ST}), however, showed that this differentiation was mainly due the divergence of the December 2010 population, which was exceptionally rich in individuals of the mixed type A/B/C. Indeed, F_{ST} measures showed that there were significantly high differences between this population and the populations from June 2009 and January and September 2010. Moreover, we noted significantly low levels of difference between the population from February 2010 with those from September and December 2010 (Table 3). For subset 2, less than 3% of the 18S rDNA genetic variation was distributed between the two populations, with significant low differentiation. AMOVA showed that the ITS genetic variance in the samples was only attributable to within-population variation (100% of the variance) whatever the subdivision considered, and revealed no significant population genetic structure (Table 2).

Detection of recombination events

The analysis of the individuals for which we obtained the sequences of the two markers 18S rDNA and ITS showed that a majority of 18S rDNA type A sequences were linked to ITS type A sequences (Fig. 5). By contrast, only one specimen (CB655) had both 18S rDNA and ITS sequences of type B, and only another one (CB797) had both 18S rDNA and ITS sequences of the mixed type A/B. Interestingly, one specimen (CB676) with 18S rDNA of

type B, four cells (CB647, CB658, CB694 and CB802) with 18S rDNA of mixed type A/B and one specimen (CB692) of 18S rDNA of mixed type A/B/C were linked to ITS sequences of type A. These combinations of different 18S rDNA and ITS sequence types observed in single individuals were most likely indicative of recombination events linked to sexual reproduction. Indeed, we investigated this possibility using recombination tests and observed significant signals of recombination among those sequences with the LD_r^2 and $LD|D'$ tests (p -value < 0.001), although our data were not sufficient to carry out tests with *maxchi* and *geneconv*.

Phylogenetic analyses

The maximum likelihood (ML) phylogenetic analysis of the concatenated 18S rDNA and ITS sequences supported that all the *U. claparedei* sequences branched together with strong support (ML bootstrap value -BV- 85%) within the tintinnid ciliates (Fig. 6A). In order to see how the *Undella* sequences were related among them, we carried out a phylogenetic network analysis (Fig. 6B) including all the sequences ascribed to Undellidae. The results revealed that sequences could be divided into two subclusters. The first one clustered the environmental sequences of *Undella* previously assigned to sub-clade 3-I (Bachy et al., in prep, see section IV.C), the reference sequence of *Undella claparedei* CB818, the sequences with 18S rDNA and ITS of type A ($n=71$), the sequences with 18S rDNA of type B, those of mixed type A/B or A/B/C and ITS of type A ($n=6$), and the sequence of 18S rDNA and ITS of mixed type A/B ($n=1$). We noted that the reference sequence available in GenBank (specimen CB818, #JQ408164) was identical to the *U. claparedei* sequences with 18S rDNA and ITS of type A, which were the most frequently retrieved in our samples. Moreover, one environmental sequence (VilleFr-23_clone9) was identical to the *U. claparedei* sequence with 18S rDNA of type B and ITS of type A. The second subcluster grouped the *U. claparedei* specimen CB655, which had 18S rDNA and ITS sequences of type B, with the species *Undella hyalina* (specimen CB854, #JQ408171) and with several environmental sequences of *Undella* previously assigned to sub-clades 3-II and 3-III (Bachy et al., in prep, see section IV.C). This *U. hyalina* clade was also supported in the ML tree (BV 80%).

Discussion

The specimens with *Undella claparedei* morphology isolated from the Bay of Villefranche-sur-Mer in diverse months and seasons over our two-year survey were not genetically identical but showed different degrees of inter- and intra-individual diversity in the 18S rDNA and ITS loci. Nevertheless, all those specimens most likely belonged to the same species since the percentage of sequence identity was higher than the typical threshold found for 18S rDNA sequences as discriminant between species in tintinnid ciliates (99% ; Bachy et al 2012, Santoferrara et al 2012) or in ciliate species in general (98% ; Nebel et al 2011). Concerning the ITS region, even if recent work has given first insights about the evolutionary rate of this marker in tintinnids (Bachy et al 2012), there is still not enough information for this marker to establish a divergence threshold for species as in the case of the 18S rDNA.

We were unable to find a relationship between the sequence differences and any morphological variation in the lorica of the studied individuals. In fact, the overall morphology of all our *U. claparedei* cells remained globally similar during the temporal survey except for the very sporadic appearance of some cells with tenuous morphological differences. However, the sequences of these cells were identical to those of individuals with the typical morphology. Moreover, our phylogenetic analyses clustered the *Undella hyalina* specimen CB854 (#JQ408171), characterized by a very long tube-shaped lorica (~250 µm length) slightly angled in the aboral end, with the *Undella claparedei* specimen CB655 (with 18S rDNA and ITS sequences of type B). Thus, the different loricae found in the *Undella* morphological variants may just represent different life cycle stages or ecotypes, and do not preclude the possibility of interbreeding, as attested by the recombination events between different sequence types that we detected.

Intra-individual polymorphism

Intra-individual polymorphism was detected in numerous specimens which showed multiple 18S rDNA or ITS variants (A, B and C for the 18S rDNA; A and B for the ITS), detected as overlapping peaks in their sequences. This agreed with intra-individual polymorphism of ribosomal markers already noticed in a wide variety of protists, including foraminifers (Pawlowski et al 2007), diatoms (D'Alelio et al 2009, Orsini et al 2004), dinoflagellates (Wayne Litaker et al 2007), phaeophytes (Coyer et al 2001) and raphidophytes (Kooistra et al

2001). Sometimes, intra-individual variation was high enough to suppose that it arises from hybridization between morphologically different populations (Pillet et al 2012).

Unequal crossing over and gene conversion are generally thought to be responsible for multi-copy gene (such as the rRNA cistron) sequence homogenization within individuals and also among individuals in recombining populations (Arnheim 1983, Hillis and Dixon 1991). However, the multiple 18S rDNA and ITS copies diverging by a number of SNPs in some individuals might suggest a slow pace of concerted evolution occurring in our *Undella* populations.

Several main sources of intra-individual polymorphism in rDNA sequences are known, including hybridization (Casteleyn et al 2009a, D'Alelio et al 2009), presence of pseudogenes (Santos et al 2003, Xu et al 2009), functionally different ribosome populations (Rooney 2004, Thompson et al 1999) and presence of multiple nuclei (Kuhn et al 2001, O'Mahony et al 2007). In our case the presence of pseudogenes or different functional types of rDNA appeared unlikely, as our sequences did not show long indels or mutations in the 5.8S rDNA, and the ITS-2 secondary structures folded normally. As most ciliates, *Undella* possesses two types of nuclei: small diploid micronuclei containing standard chromosomes and large macronuclei, which are transcriptionally active. Moreover, tintinnid cells are multinucleated and have at least two micro- and two macro-nuclei (Agatha and Strüder-Kypke 2007). However, since in all ciliates macronuclei develop from the micronuclei just after conjugation, if different rDNA sequence types co-occur within the same individual, they have to be encoded in all the different nuclei. Finally, sexual reproduction is the most plausible scenario to explain intra-individual polymorphism within our populations. Although little is known at present about the life cycle of tintinnids in nature, the alternation of sexual and asexual phases is likely the rule (Montagnes in press). The point mutations detected in the ITS-2 sequences imply secondary structures with non-compensatory base changes. ITS-2 secondary structures have been proposed as a marker to predict the mating ability among closely related organisms under the CBC (Compensating Base Change) hypothesis. In fact, the presence of at least one CBC in the helices II or III of ITS-2 appears to be correlated with mating inability (Coleman 2009). Even if this concept has recently been seriously debated with the study of a group of green algae, the Ulvales (Caisova et al 2011), the fact that we detected no CBC in our *Undella* population sequence data suggests that individuals have the potential to reproduce between them. Actually, the 4 non-compensating substitutions found in

our ITS types A and B, and above all, their co-occurrence in the mixed type A/B, are a strong signal of sexual reproduction in our populations. In fact, if we consider each type of 18S rDNA and ITS sequence that has been found alone from the direct sequencing of single cell genes as a parental type, the easiest way to explain the mixed types is for them to be descendants from sexual reproduction of both different parental types. Moreover, the observation of those mixed types and the significant positive results from recombination tests applied on the 18S rDNA + ITS concatenated data provide additional evidence for sexual exchange of genetic information in *U. claparedei* populations in nature. Intraspecific hybridization in the field between different varieties of the marine free-living diatom *Pseudo-nitzschia pungens* has been detected in the same way using ITS sequences (Casteleyn et al 2009a, D'Alelio et al 2009).

Genetic diversity between and within populations

Our analyses using a population genetics approach has revealed that the 18S rDNA sequence diversity of *U. claparedei* presented a moderate genetic structure related with time, meaning that for this marker the populations that we studied appeared to be differentiated depending on the month of sampling. We observed that these differences were essentially due to the genetic compositions of the populations collected in February and December of the last year of sampling, 2010, with significant F_{st} values that ranged from 0.055 to 0.224. These values indicated moderate (0.05–0.15) to strong (0.15–0.25) genetic structure (Hartl and Clark 1997, Wright 1978). By contrast, the population subdivision by physical conditions at the moment of sampling (stratified vs. mixed water column) was less effective to explain the population genetic structure. These results were similar to those obtained with the dinoflagellate *Prorocentrum micans* over a period of two years, which also revealed a moderate differentiation between populations over time (Shankle et al 2004). However, a previous study conducted on the diatom *Pseudo-nitzschia pungens* over a period of 18 months revealed a weak genetic differentiation with time (Evans et al 2005). This was the case also for the ITS sequence diversity of our populations of *U. claparedei*, for which the total genetic diversity did not change drastically over time and the genetic variance was exclusively explained by the within population diversity. In consequence, there is no general trend common to all temporal surveys of genetic differentiation in marine protists. Moreover, our contrasting results

between the 18S rDNA and ITS markers even suggested that, depending on the molecular markers targeted, different conclusions are possible, which complicates the task and casts doubts about studies carried out with single markers.

Examination of the 18S rDNA and ITS sequence type frequencies showed that the populations of *U. claparedei* are generally dominated by one sequence type (A), although along time different types may appear and increase their proportion. As the 18S rDNA and ITS mixed types are as frequent as the minority sequence types, they likely result from relatively frequent hybridization events between the abundant and minority types. Our long-term sampling showed that the Bay of Villefranche is characterized by the dominance of *U. claparedei* of type A. Furthermore, we observed at the end of our sampling period an increase of the proportions of other 18S rDNA types, including the type B and the mixed types A/B and A/B/C, which represented up to 50% of the *Undella* population individuals in December 2010. Further work on these *Undella* local populations should help to show whether the distribution of 18S rDNA types remains as observed in our monitoring, or whether over longer periods (several years) local populations tend to shift towards other dominant types.

From a biogeographical point of view, to determine whether the degree of temporal genetic differentiation was similar or not to the degree of spatial genetic differentiation at a larger geographical scale would require the collection of *U. claparedei* specimens from other locations.

The study of populations in other Mediterranean regions could also help to test if the intraspecific and intra-individual 18S rDNA and ITS sequence variation observed in our populations result from a gene inflow from outside the Villefranche area. For instance, currents in the Ligurian Sea could bring immigrants from distant populations with different genotypes, which could explain the appearance of the 18S rDNA type C only at the beginning of 2010, after almost one year of sampling. Those aliens bearing the 18S rDNA type C may have hybridized with the local individuals and the new type spread within the population. However, because so little is known about the timing and prevalence of sexual reproduction in field populations of marine free-living protists, it is very difficult to quantify their role in contributing to the observed levels of genetic variation. If sexual reproduction is common, it could account for the moderate change in genetic diversity over time, especially when considering the ITS region, by continually replenishing the genetic diversity of the population

through recombination. Actually, the apparent perennial presence of *U. claparedei* in the Bay of Villefranche over the course of 2 years reflects the occurrence and the succession of different genotypes, with some that likely represent recombinant types. Tracking genetic diversity in microbial populations and recombination events over time appear to be essential to understand protist dynamics in complex environments, such as marine ecosystems.

Acknowledgements

We thank the vessel crew from the Villefranche-sur-Mer oceanological observatory, in particular Jean-Yves Carval, for help during sampling. This work was funded by the ANR programs Biodiversity (ANR BDIV 07 004-02 'Aquaparadox') and Génomique microbienne à grande échelle (ANR-08-GENM-024-001, EVOLDEEP).

References

- Agatha S, Strüder-Kypke MC (2007). Phylogeny of the order Choreotrichida (Ciliophora, Spirotricha, Oligotrichea) as inferred from morphology, ultrastructure, ontogenesis, and SSrRNA gene sequences. *Eur J Protistol* 43: 37-63.
- Amato A, Kooistra WH, Ghiron JH, Mann DG, Proschold T, Montresor M (2007). Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158: 193-207.
- Arnheim N (1983). Concerted evolution of multigene families. *Evolution of Genes and Proteins*.
- Bachy C, Gómez F, López-García P, Dolan JR, Moreira D (2012). Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora). *Protist*.
- Bachy C, Moreira D, Dolan JR, López-García P (in prep.). Seasonal dynamics of free-living tintinnid ciliate communities revealed by environmental sequences from the North-West Mediterranean Sea.
- Balech E (1975). La familia Undellidae (Protozoa, Ciliophora, Tintinnina). *Physis, Sección A, Buenos Aires* 34: 377-398.
- Bryant D, Moulton V (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Mol Biol Evol* 21: 255-265.
- Caisova L, Marin B, Melkonian M (2011). A close-up view on ITS2 evolution and speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). *BMC Evol Biol* 11: 262.

Casteleyn G, Adams NG, Vanormelingen P, Debeer A-E, Sabbe K, Vyverman W (2009a). Natural Hybrids in the Marine Diatom *Pseudo-nitzschia pungens* (Bacillariophyceae): Genetic and Morphological Evidence. *Protist* 160: 343-354.

Casteleyn G, Evans K, Backeljau T, D'hondt S, Chepurinov V, Sabbe K *et al* (2009b). Lack of population genetic structuring in the marine planktonic diatom *Pseudo-nitzschia pungens* (Bacillariophyceae) in a heterogeneous area in the Southern Bight of the North Sea. *Mar Biol* 156: 1149-1158.

Catania F, Wurmser F, Potekhin AA, Przyboś E, Lynch M (2009). Genetic Diversity in the *Paramecium aurelia* Species Complex. *Mol Biol Evol* 26: 421-431.

Coleman AW (2009). Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Mol Phylogenet Evol* 50: 197-203.

Coyer JA, Smith GJ, Andersen RA (2001). Evolution of *Macrocystis* spp. (Phaeophyceae) as determined by ITS1 and ITS2. *J Phycol* 37: 574-585.

D'Alelio D, Amato A, Kooistra WH, Procaccini G, Casotti R, Montresor M (2009). Internal Transcribed Spacer Polymorphism in *Pseudo-nitzschia multistriata* (Bacillariophyceae) in the Gulf of Naples: Recent Divergence or Intraspecific Hybridization? *Protist* 160: 9-20.

Darling KF, Wade CM (2008). The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Mar Micropaleontol* 67: 216-238.

de Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J (1999). Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences* 96: 2864-2868.

Dolan JR, Stoeck T (2011). Repeated sampling reveals differential variability in measures of species richness and community composition in planktonic protists. *Environmental Microbiology Reports* 3: 661-666.

Duff RJ, Ball H, Lavrentyev PJ (2008). Application of combined morphological-molecular approaches to the identification of planktonic protists from environmental samples. *J Eukaryot Microbiol* 55: 306-312.

Evans KM, Kühn SF, Hayes PK (2005). High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia pungens* (Bacillariophyceae) populations. *J Phycol* 41: 506-514.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 131: 479-491.

Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Libertas Academica*.

Gomez F, Gorsky G (2003). Annual microplankton cycles in Villefranche Bay, Ligurian Sea, NW Mediterranean. *J Plankton Res* 25: 323-339.

Hartl DL, Clark AG (1997). *Principles of Population Genetics*, 3rd edn. Sinauer Associates Inc.: Sunderland, MA.

Hill WG, Robertson A (1968). Linkage disequilibrium in finite populations. *Theor Appl Genet* 38: 226-231.

Hillis DM, Dixon MT (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly review of biology* 66: 411-453.

Huson DH, Bryant D (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Mol Biol Evol* 23: 254-267.

Jobb G, von Haeseler A, Strimmer K (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4: 18.

Katoh K, Misawa K, Kuma K, Miyata T (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059-3066.

Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M (2009). 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. *Gene* 430: 50-57.

Koetschan C, Hackl T, Müller T, Wolf M, Förster F, Schultz J (2012). ITS2 Database IV: Interactive taxon sampling for internal transcribed spacer 2 based phylogenies. *Mol Phylogenet Evol* 63: 585-588.

Kooistra WHCF, de Boer MK, Vrieling EG, Connell LB, Gieskes WWC (2001). Variation along ITS markers across strains of *Fibrocapsa japonica* (Raphidophyceae) suggests hybridisation events and recent range expansion. *J Sea Res* 46: 213-222.

Kršinić F, Grbec B (2006). Horizontal distribution of tintinnids in the open waters of the South Adriatic (Eastern Mediterranean). *Sci Mar* 70: 77-88.

Kuhn G, Hijri M, Sanders IR (2001). Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* 414: 745-748.

Laval-Peuto M, Brownlee DC (1986). *Identification and systematics of the Tintinnina (Ciliophora): evaluation and suggestions for improvement*, vol. 62. Institut océanographique: Paris, FRANCE.

Lewontin RC (1964). The interaction of selection and linkage. I. Genetic considerations; heterotic models. *Genetics* 49: 49-67.

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

Lowe CD, Day A, Kemp SJ, Montagnes DJS (2005). There are High Levels of Functional and Genetic Diversity in *Oxyrrhis marina*. *J Eukaryot Microbiol* 52: 250-257.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Micro* 4: 102-112.

Maynard Smith J (1992). Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126-129.

Montagnes DJS (in press). Ecophysiology and behaviour: a review of tintinnid biology and its application. In: Dolan JR, Agatha S, Coats DW, Montagnes DJS, Stoecker DK (eds). *Biology and Ecology of Tintinnid Ciliates: Models for Marine Plankton*. Wiley-Blackwell: Oxford.

Montresor M, Sgrosso S, Procaccini G, Kooistra WHCF (2003). Intraspecific diversity in *Scrippsiella trochoidea* (Dinophyceae): evidence for cryptic species. *Phycologia* 42: 56-70.

Nebel M, Pfabel C, Stock A, Dunthorn M, Stoeck T (2011). Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. *Environmental Microbiology Reports* 3: 154-158.

Nival P, Corre MC (1976). Variation annuelle des caractéristiques hydrologiques de surface dans la baie de Villefranche-sur-Mer. *Ann Inst Océanogr Paris* 52: 57-78.

O'Mahony EM, Tay WT, Paxton RJ (2007). Multiple rRNA Variants in a Single Spore of the Microsporidian *Nosema bombi*. *J Eukaryot Microbiol* 54: 103-109.

Orsini L, Procaccini G, Sarno D, Montresor M (2004). Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Mar Ecol Prog Ser* 271: 87-98.

Pawlowski J, Fahrni J, Lecroq B, Longet D, Cornelius N, Excoffier L *et al* (2007). Bipolar gene flow in deep-sea benthic foraminifera. *Mol Ecol* 16: 4089-4096.

Piganeau G, Eyre-Walker A, Grimsley N, Moreau H (2011). How and Why DNA Barcodes Underestimate the Diversity of Microbial Eukaryotes. *PLoS ONE* 6: e16342.

Piganeau GI, Gardner M, Eyre-Walker A (2004). A Broad Survey of Recombination in Animal Mitochondria. *Mol Biol Evol* 21: 2319-2325.

Pillet L, Fontaine D, Pawlowski J (2012). Intra-Genomic Ribosomal RNA Polymorphism and Morphological Variation in *Elphidium macellum* Suggests Inter-Specific Hybridization in Foraminifera. *PLoS ONE* 7: e32373.

Posada D, Crandall KA (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences* 98: 13757-13762.

Posada D (2002). Evaluation of Methods for Detecting Recombination from DNA Sequences: Empirical Data. *Mol Biol Evol* 19: 708-717.

Posada D, Buckley TR (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Syst Biol* 53: 793-808.

Prescott DM (1994). The DNA of ciliated protozoa. *Microbiol Mol Biol Rev* 58: 233-267.

Rambaut A (2009). FIGTREE <http://tree.bio.ed.ac.uk/>.

Rodriguez F, Derelle E, Guillou L, Le Gall F, Vaultot D, Moreau H (2005). Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol* 7: 853-859.

Rooney AP (2004). Mechanisms Underlying the Evolution and Maintenance of Functionally Heterogeneous 18S rRNA Genes in Apicomplexans. *Mol Biol Evol* 21: 1704-1711.

Santoferrara LF, McManus GB, Alder VA (2012). Utility of Genetic Markers and Morphology for Species Discrimination within the Order Tintinnida (Ciliophora, Spirotrichea). *Protist*.

Santos SR, Kinzie RA, Sakai K, Coffroth MA (2003). Molecular Characterization of Nuclear Small Subunit (ISS)-rDNA Pseudogenes in a Symbiotic Dinoflagellate (Symbiodinium, Dinophyta). *J Eukaryot Microbiol* 50: 417-421.

Sawyer S (1989). Statistical tests for detecting gene conversion. *Mol Biol Evol* 6: 526-538.

Schultz J, Maisel S, Gerlach D, Müller T, Wolf M (2005). A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11: 361-364.

Shankle AM, Mayali X, Franks PJS (2004). Temporal patterns in population genetic diversity of *Prorocentrum micans* (Dinophyceae) *J Phycol* 40: 239-247.

Slapeta J, López-García P, Moreira D (2006). Global Dispersal and Ancient Cryptic Species in the Smallest Marine Eukaryotes. *Mol Biol Evol* 23: 23-29.

Snoeyenbos-West OLO, Salcedo T, McManus GB, Katz LA (2002). Insights into the diversity of choreotrich and oligotrich ciliates (Class: Spirotrichea) based on genealogical analyses of multiple loci. *Int J Syst Evol Microbiol* 52: 1901-1913.

Thompson J, Van Spaendonk RML, Choudhuri R, Sinden RE, Janse CJ, Waters AP (1999). Heterogeneous ribosome populations are present in *Plasmodium berghei* during development in its vector. *Mol Microbiol* 31: 253-260.

Vaultot D, Romari K, Not F (2002). Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends Microbiol* 10: 266-267.

Wayne Litaker R, Vandersea MW, Kibler SR, Reece KS, Stokes NA, Lutzoni FM *et al* (2007). Recognizing dinoflagellate species using ITS rDNA sequences. *J Phycol* 43: 344-355.

Wright S (1978). *Evolution and the genetics of populations*, vol. 4. University of Chicago press: Chicago.

Xu J, Zhang Q, Xu X, Wang Z, Qi J (2009). Intragenomic variability and pseudogenes of ribosomal DNA in Stone flounder *Kareius bicoloratus*. *Mol Phylogenet Evol* 52: 157-166.

Zuker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415.

Table 1. Isolation date, number of isolated cells, number of 18S rDNA, ITS and both marker sequences generated by sampling date.

Sampling date	No of <i>Undella claparedei</i> specimens	No of 18S rDNA sequences (%)	No of ITS region sequences (%)	Both markers (%)
February the 18th, 2009*	41	15 (36.6%)	18 (43.9%)	13 (31.7%)
March the 18th, 2009	8	1 (12.5%)	4 (50.0%)	1 (12.5%)
April the 16th, 2009*	29	3 (10.3%)	5 (17.2%)	2 (6.9%)
May the 27th, 2009	7	1 (14.3%)	3 (42.9%)	0 (0%)
June the 25th, 2009*	30	11 (36.7%)	18 (60.0%)	11 (36.7%)
November the 18th, 2009	12	1 (8.3%)	9 (75.0%)	1 (8.3%)
December the 16th, 2009*	63	8 (12.7%)	13 (20.6%)	5 (7.9%)
January the 27th, 2010*	42	16 (38.1%)	29 (69.0%)	8 (19.0%)
February the 24th, 2010*	67	34 (50.7%)	32 (47.8%)	22 (32.8%)
May the 26th, 2010	18	2 (11.1%)	2 (11.1%)	0 (0%)
September the 29th, 2010*	30	20 (66.7%)	17 (56.7%)	13 (4.3%)
December the 1st, 2010*	70	39 (55.7%)	4 (5.7%)	3 (4.3%)
Total	417	156 (37.4%)	154 (36.9%)	79 (18.9%)

* date of high *Undella claparedei* concentrations in the material observed under the microscope.

Table 2. Analyses of molecular variance between and within populations for the 18S rDNA and ITS markers considering two subdivision types of populations: by sampling date (subset 1) and by physical conditions of the water column (subset 2).

18S rDNA Subset 1: 6 populations (June 2009, January, February, September and December 2010)

Source of variation	Degrees of freedom	Variance components	% of variation
Among populations	5	0.05192	12.74
Within populations	262	0.35554	87.26

Fixation index Φ_{ST} : 0.1274**

ITS region Subset 1: 6 populations (February, June and December 2009, January, February and September 2010)

Source of variation	Degrees of freedom	Variance components	% of variation
Among populations	5	0.979	0
Within populations	246	0.23966	100

Fixation index Φ_{ST} : 0.0045

18S rDNA Subset 2: 2 populations (mixed vs stratified)

Source of variation	Degrees of freedom	Variance components	% of variation
Among populations	1	0.01263	2.88
Within populations	298	0.42530	97.12

Fixation index Φ_{ST} : 0.0288*

ITS region Subset 2: 2 populations (mixed vs stratified)

Source of variation	Degrees of freedom	Variance components	% of variation
Among populations	1	0.00240	0
Within populations	304	0.36630	100

Fixation index Φ_{ST} : 0.0066

Statistical significance is based on 10,000 permutations in Arlequin: * $P < 0.05$; ** $P < 0.001$

Table 3. Pairwise genetic distances (Fst) between the studied populations

	June 2009	January 2010	February 2010	September 2010
June 2009				
January 2010	0.01557			
February 2010	0.09036	0.04007		
September 2010	0.00423	-0.02741	0.05818*	
December 2010	0.22414**	0.17744**	0.05546*	0.20292**

Statistical significance is based on 10,000 permutations in Arlequin: * P < 0.05; ** P < 0.001.

Figure legends

Figure 1- Light micrographs showing morphological variability of *Undella claparedei* specimens. **A.** Typical round-shape lorica morphology (CB502). **B.** specimen with with a slightly pointed lorica (CB536). **C.** specimen with lengthened lorica (CB760). **D.** specimen with an anomalous lorica (CB695). Scale bar 50 μm .

Figure 2- Variations in 18S rDNA and ITS sequence types in *U. claparedei* individuals. Positions indicated in the text refer to those in the alignments datasets. Dots indicate base pairs that are identical to type A. On the right the % of sequence dissimilarity between pure sequence types are indicated.

Figure 3- Predicted secondary structures of *U. claparedei* ITS-2 helices II and III of sequence type A. Circles indicate the base-pair changes between ITS-2 sequences of type A and B.

Figure 4- Frequencies of 18S rDNA and ITS types for the 12 sampling dates. The size of each pie chart is proportional to the number of sequences. Asterisks (*) indicated the sampling dates when the *U. claparedei* appeared to be abundant under the microscope. Physical conditions of the water column are indicated in parallel to the timeline.

Figure 5- Different combinations of 18S rDNA and ITS sequence types found in single *U. claparedei* individuals classified by abundance. The histogram on the right shows the relative abundance of each combination, based on the analysis of 79 individuals.

Figure 6- Phylogenetic analysis of the representative sequences of the 6 combinations of 18S rDNA and ITS sequence types recovered in the *U. claparedei* individuals. **A.** Consensus ML tree based on reference and environmental *Undella* sequences, and a set of 8 reference tintinnid sequences publicly available. Bootstrap supports >50% are indicated at nodes. Representative concatenated sequences of this study are in bold; the corresponding sequence types are indicated between parentheses with the same colour code as in Fig.5. The number of individuals with each combination is indicated on the right. **B.** Neighbor-Net network of an *Undella* sequence dataset comprising environmental and specimen sequences. The edge lengths illustrate the different levels of relatedness between the genetic types. Representative concatenated sequences from this study are in bold.

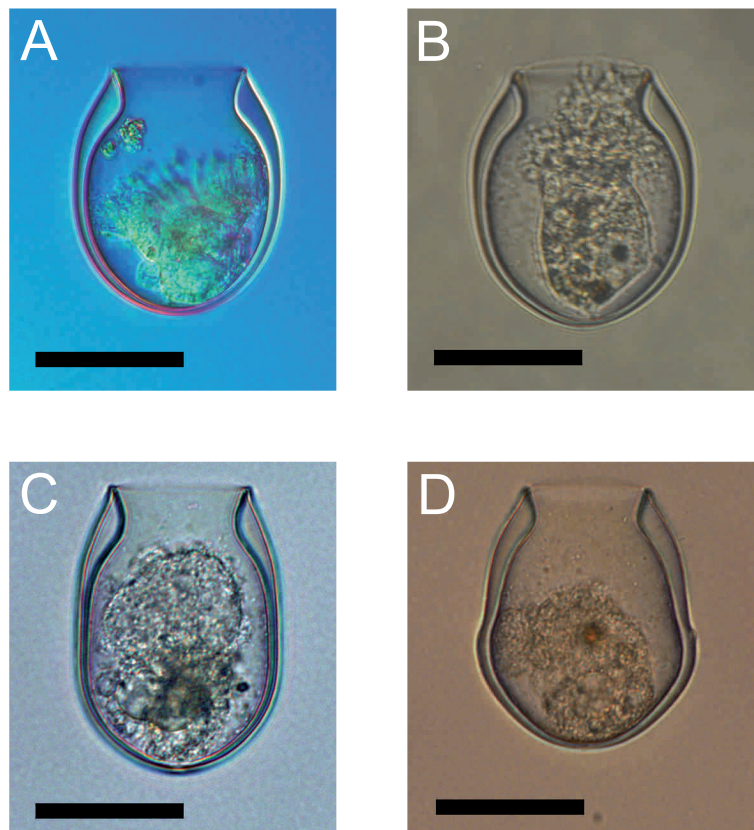


Figure 1. Bachy et al.

SSU rDNA sequence type

Position (bp)	296	304	306	331	353	545	
A	C	A	T	A	A	T	} 0.3% } 0.9% } 0.6%
B	.	.	.	G	.	C	
C	G	T	C	.	G	.	
A/B	.	.	.	A/G	.	T/C	
A/C	C/G	A/T	T/C	.	A/G	.	
A/B/C	C/G	A/T	T/C	A/G	A/G	T/C	

ITS sequence type

Position (bp)	190	215	217	228	
A	C	G	G	T	} 0.9%
B	T	A	A	C	
A/B	C/T	G/A	G/A	T/C	

Figure 2. Bachy et al.

ITS-2 Helix II

Helix III

○ Non-compensatory changes between ITS types

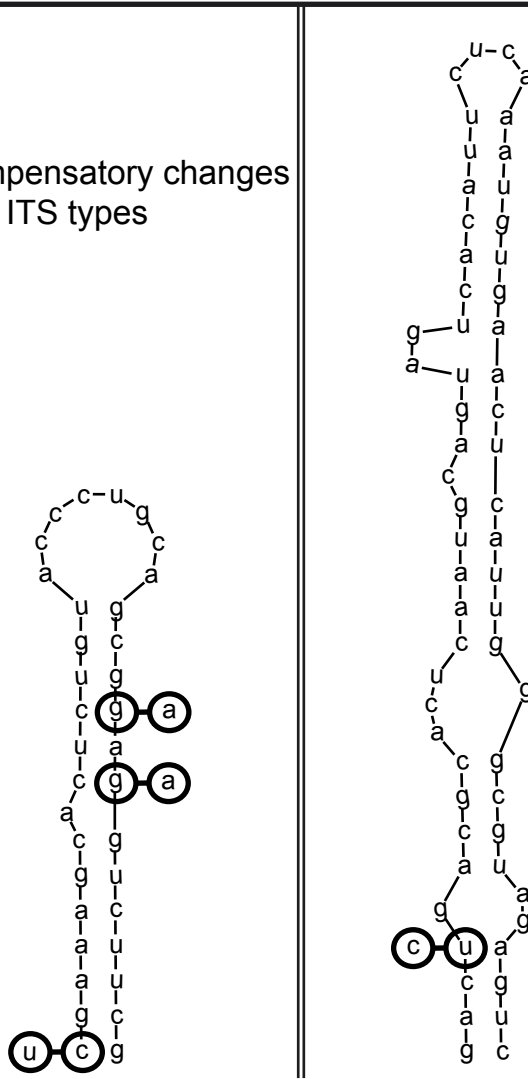


Figure 3. Bachy et al.

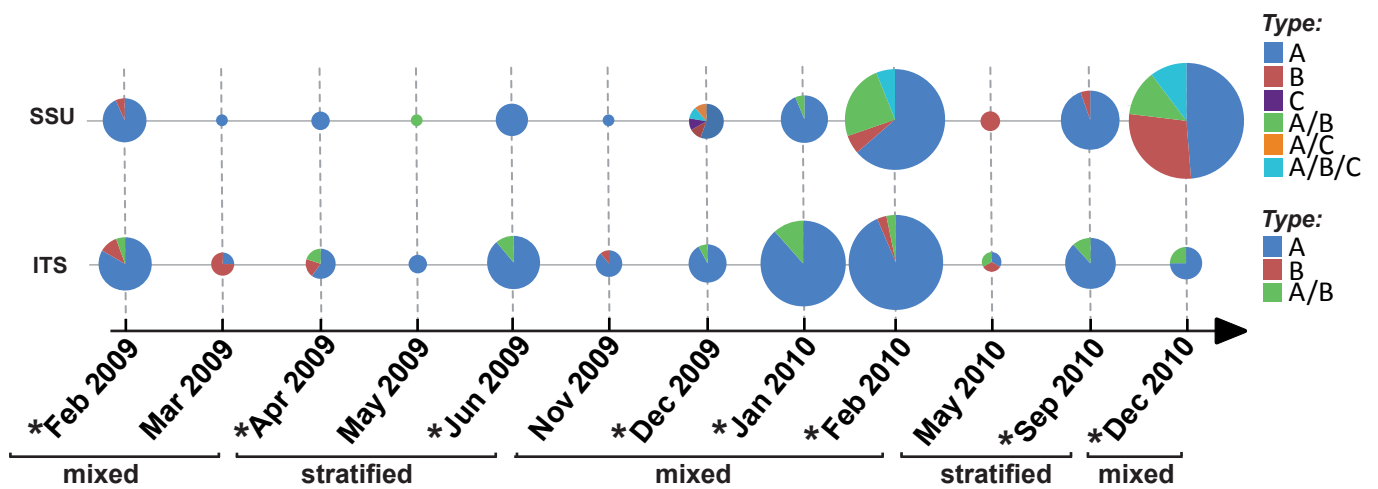


Figure 4. Bachy et al.

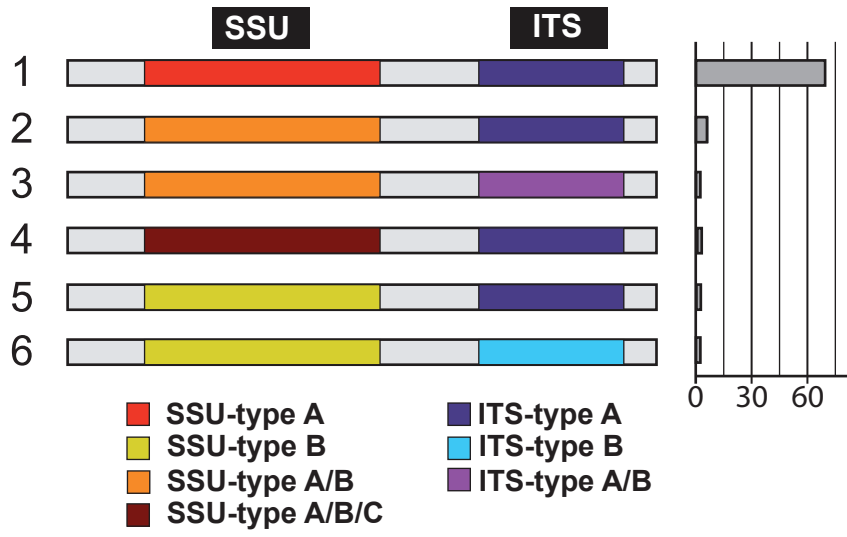


Figure 5. Bachy et al.

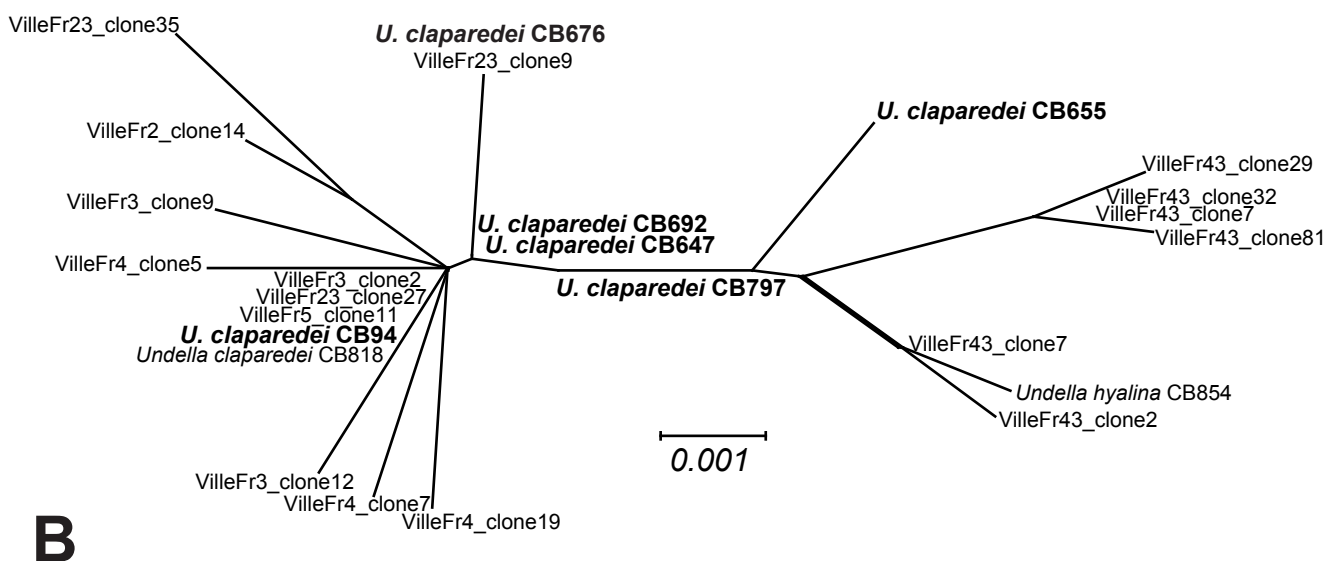
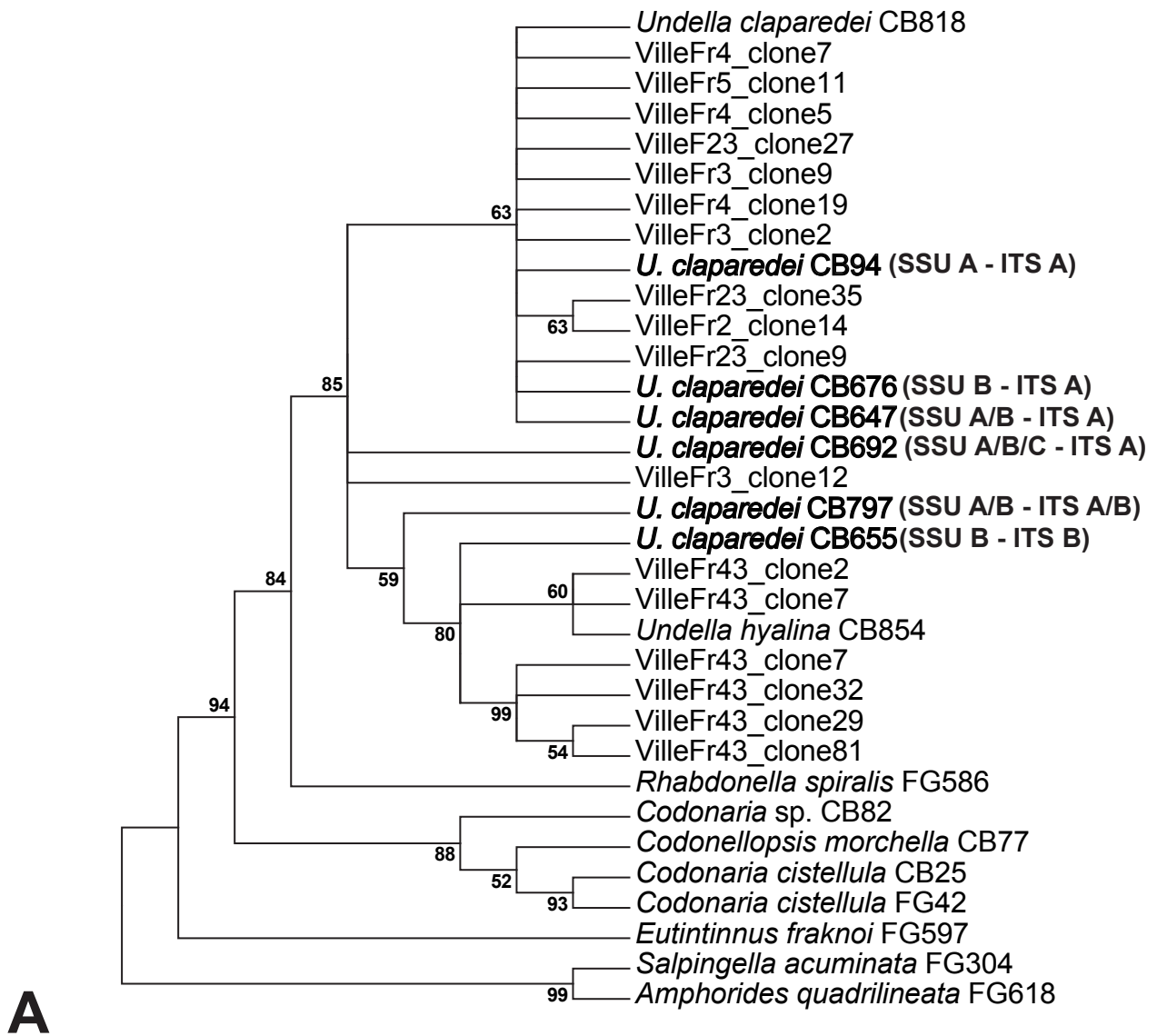


Figure 6. Bachy et al.

3. Synthèse et éléments de discussion

Nos analyses des individus de l'espèce *Undella claparedei* de la rade de Villefranche-sur-Mer au cours des deux ans ont révélé différents types de séquences de l'ADNr 18S et de l'ITS. Il n'y a pas eu de différences significatives détectées dans la morphologie de la lorica (aspect général et diamètre de l'ouverture orale) en fonction du type de séquence porté par les individus. De plus, nos analyses phylogénétiques ont groupé l'espèce *Undella hyalina*, espèce caractérisée par une lorica très longue, collectée pour l'étude de phylogénie moléculaire (voir dans l'article de la partie IV.A), avec un individu *U. claparedei* possédant les types de séquences minoritaires (type de séquence B) pour l'ADNr18S et l'ITS. Potentiellement, cette observation renforce l'idée que des morphologies différentes chez certaines espèces de tintinnidés représentent en réalité des morphotypes différents d'une même espèce. Certains individus possédaient un mélange de séquences d'ADNr 18S ou d'ITS de différents types, qui sont très probablement des signaux d'épisodes de reproduction sexuelle et d'hybridation. Cette hypothèse a été fortement soutenue par la détection de signaux de recombinaison en analysant les individus pour lesquels nous avons séquencé avec succès l'ADNr 18S et les ITS. Ces découvertes chez l'espèce *U. claparedei* tout au long du suivi nous laissent penser que ces individus sont capables de se reproduire entre eux, et donc que nous sommes en présence de diversité intra-spécifique et non de diversité cryptique. Nous avons donc pu avec confiance réaliser des analyses typiques de génétique des populations pour analyser la structuration des populations avec le temps. L'ADNr 18S a semblé plus décisif que l'ITS puisqu'il a permis de détecter une structure génétique temporelle. En outre, la présence et la dominance d'un type de séquence majoritaire pour les deux marqueurs met l'accent sur le maintien sur le site d'étude d'un génotype en particulier. Les signes de structure génétique temporelle, l'apparition d'un troisième type de séquence pour l'ADNr 18S (type de séquence C) en fin d'année 2009 et de cellules possédant un mélange de types de séquences (A/B/C) laisse supposer un phénomène d'introduction extérieure de nouveaux génotypes par le biais des courants marins. Afin de mieux qualifier les phénomènes de migration et tester cette hypothèse, il est nécessaire de collecter et séquencer des individus dans d'autres sites plus ou moins éloignés en Mer Méditerranée.

4. Essais expérimentaux additionnels et Perspectives

Avec l'objectif initial d'obtenir un maximum d'information génétique, nous avons l'idée d'amplifier d'autres marqueurs moléculaires, notamment des gènes codant pour des protéines. J'ai donc essayé d'amplifier par PCR directe les gènes nucléaires codant pour la tubuline (Tub), l'actine (Act) ou le gène mitochondrial codant pour la sous-unité 1 de la cytochrome oxydase (COI) communément employé en barcoding. Bien que j'aie pu obtenir quelques séquences d'actine, celles-ci étaient trop peu nombreuses pour pouvoir envisager une étude de génétique des populations. Il peut être envisageable pour une étude de génétique des populations avec un grand nombre de marqueurs, autrement dit une étude de génomique des populations, d'améliorer le protocole d'amplification des gènes par PCR directe. Il serait sûrement nécessaire de développer un protocole expérimental incluant une étape de WGA à partir de cellules individuelles de tintinnidés pour obtenir suffisamment de matériel génétique. Ou bien, le développement de la transcriptomique permettra probablement dans un futur proche les études de transcriptomes de cellules individuelles. Tang et collaborateurs (2009) ont récemment démontré qu'une approche de séquençage haut débit des ARN (*RNA-seq*) sur cellules individuelles était possible sur des blastomères de souris, mais des modifications considérables sont nécessaires avant que cette méthode ne puisse être appliquée sur des protistes.

Nous avons utilisé une analyse de la variance moléculaire (AMOVA) pour estimer la distribution de la diversité génétique au sein et entre les populations, et également pour estimer les indices de différenciation (F_{st}) entre les populations. Cependant, comme l'évolution concertée peut affecter les variances des fréquences des différents allèles (appelés « sequence types » dans l'article), nous sommes conscients que de telles estimations devraient être interprétées avec précaution. De plus, dans le cas des spécimens pour lesquels nous avons observé la co-occurrence supposée de trois types d'ADNr 18S différents, un doute subsiste pour savoir si deux ou trois types de séquences co-occurrent dans un génome. Nous avons donc prévu de cloner les produits des PCR réalisées sur les cellules isolées d'*Undella claparedei*, dans le but d'examiner l'étendue et la véracité de la variation intra-individuelle des séquences de l'ADNr 18S et des régions de l'ITS. En premier lieu nous clonerons les produits de PCR des individus pour lesquelles nous avons interprété la présence de doubles pics dans les chromatogrammes comme la coexistence de plusieurs haplotypes. Ces étapes de clonage offrent la possibilité de clairement identifier les différents haplotypes, mais révèlent

parfois une diversité intra-individuelle qui n'avait pas été remarquée jusqu'ici. Chez le genre *Pseudo-nitzschia* (diatomées), il a été ainsi montré non seulement la présence de différents haplotypes pour un même individu, mais des haplotypes recombinants (Casteleyn et al. 2009a) et de nouveaux SNPs non observés avec les produits des PCRs directes (D'Alelio et al. 2009).

V. Discussion Générale et Perspectives

A. Intégration de l'information morphologique et des données moléculaires chez les ciliés tintinnidés

Nos analyses ont permis de caractériser génétiquement un grand nombre de ciliés tintinnidés isolés à partir de communautés naturelles complexes. La méthode d'amplification et séquençage des marqueurs génétiques à partir des cellules individuelles s'est révélée être une stratégie adaptée pour coupler les classifications classiques basées sur des caractères morphologiques à une classification naturelle basée sur la divergence phylogénétique des tintinnidés.

Nous avons pu améliorer et étoffer l'arbre phylogénétique des tintinnidés en utilisant le gène conservé codant pour l'ADNr 18S d'un nombre conséquent et varié des morpho-espèces classiquement décrites. En utilisant la phylogénie moléculaire comme un guide de l'histoire évolutive, l'utilisation de la morphologie de la lorica comme caractère taxonomique pour délimiter les espèces a été évaluée. Découvrir si la lorica peut être utilisée pour discriminer les espèces est essentiel du fait des avantages méthodologiques offerts par cette structure (facilité de collecte et d'identification). La lorica a par d'ailleurs permis de réaliser de nombreuses études d'écologie sur la distribution et la diversité des tintinnidés. Nos résultats de phylogénie moléculaire, basés sur l'ADNr 18S et parfois la région des ITS sur un nombre divers d'espèces, ont montré que les relations de parenté des tintinnidés sont plus compliquées qu'il n'y paraissait (partie IV.A). Certains de nos résultats vont dans le sens des taxonomistes qui ont eu tendance à maintenir un nombre réduit d'espèces de tintinnidés en intégrant une variabilité morphologique au sein d'une même espèce (Bakker et Phaff 1976; Laval-Peuto et Brownlee 1986). Ainsi, l'existence de trois espèces décrites pour le genre *Rhabdonella* sur la base de la longueur différente entre les loricae n'a pas lieu d'être sur la base de l'ADNr 18S qui est identique pour les trois variantes morphologiques (*R. spiralis*, *R. hebe*, *R. elegans*) (partie IV.A). De façon plus surprenante, des loricae très différentes morphologiquement présentent des séquences d'ADNr 18S et ITS très proches voire identiques (e.g. *Cyttarocylis* et *Petalotricha*). Cependant, certaines espèces comme *Amphorides quadrilineata* semblent cacher derrière une morphologie apparemment constante une diversité génétique importante et pourrait être plutôt considérées comme un assemblage

d'espèces cryptiques. Malgré ces difficultés, nos résultats de comparaisons entre les séquences d'ADNr 18S et les morphologies correspondantes ont montré que dans la plupart des cas, les similarités et les différences sont cohérentes avec la diversité intra- et inter-spécifique des tintinnidés (article partie IV.B). Cette approche valide l'utilisation d'une valeur seuil de 99% de similarité de séquences d'ARNr 18S comme critère phylogénétique de l'espèce. Il existe néanmoins des écarts à cette tendance générale qui indique que les deux courants de pensées qui ont prévalu pour la classification des tintinnidés (*splitters* et *lumpers*) sont complémentaires, et surtout, qu'il est nécessaire de caractériser d'autres espèces de tintinnidés pour démêler au mieux les questionnements taxonomiques qui subsistent. En particulier, la résolution des liens de parenté entre les familles en position distale dans l'arbre nécessite l'utilisation d'autres marqueurs moléculaires, comme par exemple l'ADNr 28S, qui a récemment montré son utilité dans l'étude phylogénétique des tintinnidés (Santoferrara et al. 2012). Un travail extensif d'amplification de marqueurs moléculaires par PCR directe pour d'autres espèces semble être actuellement en cours dans la communauté scientifique (Saccà et al. 2012; Santoferrara et al. 2012), notamment pour des espèces ayant des répartitions géographiques restreintes aux mers polaires (Kim, communication personnelle).

Les nouvelles séquences d'ADNr 18S et de la région des ITS obtenues par PCR directe sur cellules individuelles des tintinnidés récoltés en Méditerranée m'ont donc également servi pour établir un bon cadre taxonomique et phylogénétique pour l'identification des données génétiques environnementales sur ces mêmes eaux méditerranéennes. Au préalable, les études moléculaires environnementales de diversité par amplification d'une partie de l'ADNr 18S (~800 bp), menées sur les ciliés marins, avaient révélé un grand nombre de séquences non caractérisées (e.g. Doherty et al. 2007; Tamura et al. 2011) que l'on ne pouvait attribuer qu'au rang de l'ordre (Tintinnida) ou de la famille avec des supports phylogénétiques faibles, rendant difficile l'interprétation des résultats en lien avec l'état de l'art existant. Pour pallier à cette limite, les approches combinant l'amplification de gènes de cellules individuelles et, en parallèle, d'amplification de gènes à partir d'ADN environnemental nous ont permis de lier sans ambiguïté les données moléculaires nouvelles avec les données morphologiques classiques chez les tintinnidés (parties IV.B et IV.C). Pourtant, ce type d'approche couplée est encore rare dans les études environnementales sur la diversité des protistes (e.g. Medinger et al. 2010) : son efficacité devrait encourager son application lors de l'élaboration des protocoles expérimentaux.

Dans notre objectif d'évaluation de l'efficacité des études moléculaires de diversité à partir d'ADN environnemental pour décrire la composition des communautés, les ciliés tintinnidés comme groupe modèle ont également montré leur intérêt. L'utilisation d'amorces pour amplifier spécifiquement tout ou partie de l'ADNr 18S des tintinnidés à partir d'ADN environnemental a révélé des résultats cohérents entre la diversité phylogénétique et la diversité morphologique (partie IV.B), et ce, quelle que soit la stratégie moléculaire employée (clonage et séquençage Sanger ou pyroséquençage). Ces résultats ne correspondent pas à l'idée de diversité cryptique souvent observée chez certaines espèces de protistes, qui suppose que derrière la diversité morphologique se cache une diversité génétique plus grande. Cette spécificité des tintinnidés est très probablement à rapprocher avec la morphologie de la lorica qui semble être un caractère généralement suffisant pour discriminer une majorité d'espèces de tintinnidés, même s'il existe certaines exceptions (voir ci-dessus). Malgré ce bon accord entre la diversité morphologique et moléculaire chez les tintinnidés, il existe encore quelques décalages entre les deux approches qui pourraient être liés aux effets stochastiques d'échantillonnage, aux différences dans le nombre de noyaux que possèdent les différentes espèces ou aux différents biais d'amplification. Nous sommes également conscients que ce type de comparaison à partir d'échantillons environnementaux n'est pas possible pour de nombreuses lignées de protistes pour lesquelles les espèces sont plus difficiles à identifier morphologiquement. Mais notre cas d'étude a montré que cette approche est praticable pour décrire la diversité des tintinnidés, et donc, on peut se demander si elle pourrait être efficace pour décrire la diversité d'autres groupes de protistes dans l'environnement. Avec les coccolithophoridés comme cas d'étude, la comparaison de banques de clones d'ADNr 28S avec l'observation au microscope (optique et électronique) des mêmes échantillons d'eau de mer ont montré que la diversité morphologique ne reflétait pas précisément la diversité génétique (Liu 2009, thèse de doctorat). Cependant, on observe que les tentatives de conciliation des notions d'espèce morphologique et phylogénétique dans le contexte d'analyse métagénétique se développent chez les protistes.

La génération des milliers de pyroséquences à partir d'amplicons des fragments d'ADNr 18S et d'ITS chez les tintinnidés n'a pas mis à jour de diversité insoupçonnée de ces organismes par rapport aux méthodes de clonage et séquençage Sanger et à l'observation morphologique. Toutefois, l'utilisation des données morphologiques pour comparer avec nos jeux de séquences a permis d'évaluer raisonnablement l'efficacité de différentes approches

computationnelles pour traiter les données de pyroséquençage et limiter le gonflage artificiel de leur diversité. En effet, nous avons constaté que, pour les tintinnidés, des résultats très différents, voire contradictoires, étaient obtenus en fonction de la méthode d'alignement de pyroséquences (alignement multiple versus alignement deux à deux), la méthode et le pourcentage d'identité utilisés pour le regroupement de séquences en UTOs et des *pipelines* de correction d'erreurs de séquençage, avec le nombre d'UTOs pouvant varier d'un facteur 100 pour un même marqueur et un même échantillon (partie IV.B). La part des espèces rares évolue également du simple au double, pouvant représenter jusqu'à 99,9% de la diversité spécifique. La crainte soulevée par certaines études sur la potentielle surestimation de la diversité avec les approches de pyroséquençage des amplicons s'est révélée justifiée dans notre étude. Avec l'avènement certain de ces nouvelles méthodes de séquençage, nous mettons en garde les utilisateurs de ces techniques et conseillons d'utiliser des méthodes astringentes pour corriger les erreurs de séquençage et pour regrouper les UTOs.

L'exploration de la diversité intra-spécifique de l'espèce *Undella claparedei* a montré que la diversité génétique observée pour l'ADNr 18S et l'ITS restait relativement faible en pourcentage (<1% pour chaque marqueur). De plus, des signaux indirects de reproduction sexuée ont été détectés avec l'analyse des séquences, indiquant que, dans la rade de Villefranche, les individus décrits comme faisant partie de l'espèce *U. claparedei* sont capables de se reproduire entre eux. Il apparaît donc que le concept de morpho-espèce semble bien corrélé avec les concepts phylogénétiques et biologiques de l'espèce pour *U. claparedei*. Cependant, cette observation ne prévaut pas d'une diversité génétique plus grande à une échelle géographique plus large. Il m'apparaît nécessaire d'élargir le champ d'étude à une échelle spatiale plus large, afin d'explorer la biogéographie de cette espèce.

Les clades de tintinnidés détectés dans nos échantillons marins sont divers, regroupant en plus des familles de Tintinnidiidae, Eutintinnidae, Tintinnidae, Xystonellidae, Codonellidae, Codonellopsidae, Undellidae, Cytarocylidae, *incertae sedis* affiliées aux genres *Tintinnopsis* et *Climacocylis* (*sensu* Bachy et al. 2012, partie IV.A), des groupes environnementaux ETC-1 et ETC-2. Ces clades environnementaux qui dominent les banques de clones au cours des mois de février 2009 et 2010 dans notre suivi temporel (partie IV.C), indiquent potentiellement une importance écologique des tintinnidés affiliés à ces groupes pendant cette période de l'année. Encore une fois, cette observation appuie l'idée qu'il faille

compléter la phylogénie des tintinnidés en relation avec la taxonomie morphologique. Ce travail pourra nous permettre d'intégrer les nouvelles données générées par les études de métagénétique ou de métagénomique à large échelle avec les méthodes morphologiques traditionnelles.

B. Dynamique spatio-temporelle de la diversité génétique à l'échelle de la communauté et à l'échelle d'une morpho-espèce

Un autre des objectifs principaux de cette thèse était d'étudier l'évolution de la composition des communautés de tintinnidés dans la colonne d'eau et sur le long terme (à l'échelle de l'année).

À l'échelle des communautés, malgré le fait connu que la composition des communautés planctoniques change le long de l'année, notamment en lien avec le changement des conditions environnementales, le facteur temps est rarement exploré dans les études moléculaires sur les protistes, en particulier pour des raisons pratiques. Plusieurs découvertes prêtent à l'hypothèse écologique un rôle important dans l'assemblage des espèces des protistes marins, du moins pour certains taxons. Celles-ci incluent des études récentes documentant des changements dans la structure des communautés naturelles de protistes plus rapides et marqués que l'on pensait (Countway et al. 2005; Kim et al. 2011a; Vigil et al. 2009), et l'observation que les communautés de protistes dans des sites très éloignés peuvent être plus similaires entre eux que des communautés échantillonnées dans des sites très proches (Caron et Countway 2009). Si l'on considère que les lignées de protistes sont largement distribuées sur la planète, alors des conditions environnementales similaires à différents endroits devraient sélectionner des groupes identiques ou similaires d'espèces de protistes. Le suivi des ciliés tintinnidés s'inscrit dans ce débat, et nos observations de regroupement d'échantillons par similarités (phylogénétique et d'occurrence d'UTOs) en fonction des saisons (voir partie IV.C) vont dans le sens d'un rôle de l'environnement structurant des communautés. En particulier, la communauté de tintinnidés de l'échantillon VilleFr-43 est plus proche de celle de l'échantillon Ioni-7, collectée à la même saison dans un site éloigné en

Méditerranée, que des communautés de tintinnidés détectées le reste de l'année sur le même site, à la même profondeur. Cependant, notre étude environnementale de la partie IV.C, n'a pas pu montrer un rôle structurant significatif d'un ensemble de paramètres physico-chimiques (température, salinité, PAR) mesurés, ce qui pourrait indiquer que les paramètres biotiques affectés par ces paramètres abiotiques peuvent être plus déterminants pour les tintinnidés que les paramètres mesurés jusqu'à présent. La réalisation d'analyses plus intégratives en incorporant les variations temporelles des communautés d'autres eucaryotes, des procaryotes et virales devrait permettre à l'avenir de mieux comprendre l'influence de chaque facteur.

À l'échelle de l'espèce, notre étude de la diversité génétique intra-spécifique de l'ADNr 18S et de la région ITS de la morpho-espèce *U. claparedei* le long du temps a permis d'aborder la question de savoir si l'apparente persistance d'une morpho-espèce dans un site donné reflète le maintien d'un seul génotype ou la succession de plusieurs (succession d'espèces cryptiques). Les études de diversité aux échelles locales sont peu nombreuses et ont été réalisées en particulier chez quelques espèces de diatomées pour lesquelles la culture de lignées cellulaires est maîtrisée, permettant donc une analyse de plusieurs centaines d'individus (Casteleyn et al. 2009b; Evans et al. 2005) à différentes périodes. Les études de génétique des populations au sein de ces espèces ont montré une diversité intra-spécifique élevée, mais une différenciation entre les populations faible, probablement due à un flux de gènes important entre les différentes populations étudiées. L'approche de génétique des populations employée sur nos données moléculaires, avec la supposition que nous étions en présence d'une même espèce, a montré une diversité génétique relativement faible, mais une claire différenciation génétique des populations avec le temps, du moins pour l'ADNr 18S. Cette diversité plus faible pourrait s'expliquer par les différents marqueurs utilisés en fonction des études, puisque nous avons étudié la diversité de marqueurs qui restent beaucoup plus conservés (même si l'ITS est supposé évoluer de manière neutre) comparés aux marqueurs microsatellites traditionnellement employés en génétique des populations.

L'étude de la génétique des populations d'espèces difficilement cultivables à partir des cellules isolées manuellement en est à ses débuts. Mais avec le développement des techniques moléculaires de séquençage et d'amplification des génomes, je suppose qu'il sera vite possible d'amplifier les gènes conservés codant pour les ARNr comme marqueurs spécifiques avec d'autres gènes nucléaires ou mitochondriaux plus variables codant pour des protéines

V. Discussion Générale et Perspectives

afin d'appréhender les problématiques de flux de gènes au sein de différentes populations d'un même site, dans le temps et dans l'espace. Ensuite, peut être, nous commencerons à percer les mystères des phénomènes adaptatifs et de spéciation chez les protistes.

V. Discussion Générale et Perspectives

VI. Conclusion

Cette thèse avait pour objectif principal de caractériser la diversité génétique des ciliés tintinnidés marins et de la mettre en relation avec leur diversité morphologique. Ce couplage des données morphologiques et moléculaires devrait permettre d'établir une classification naturelle pour les tintinnidés et de servir de contrôle pour la validation des études moléculaires de diversité environnementale et des suivis spatio-temporels pour étudier l'écologie de ce groupe de ciliés. Pour atteindre ces objectifs, nous avons caractérisé leur diversité par amplification des gènes ADNr 18S et ITS à partir de cellules isolées depuis leur milieu naturel et d'analyse de banques de gènes ADNr 18S et ITS, et de fingerprinting par DGGE. Ce couplage d'études moléculaires de la diversité nous a permis d'aborder de nombreuses facettes de l'étude contemporaine des protistes. Les études menées à l'aide des différentes approches présentées dans l'introduction ont permis de résoudre certaines interrogations sur la diversité des tintinnidés à l'échelle de l'ordre taxonomique (article de la partie IV.A), à l'échelle des communautés naturelles (articles des parties IV.B et IV.C) et à l'échelle intra-spécifique (article partie IV.D). Le premier volet de cette thèse nous a permis de résoudre une partie de l'histoire évolutive de l'ordre des Tintinnida, en construisant une phylogénie moléculaire représentative de la diversité des tintinnidés (voir article IV.A). Cette phylogénie m'a offert un cadre de travail solide pour identifier les séquences générées par les études moléculaires environnementales de diversité (voir articles IV.B et IV.C) et pour évaluer leur efficacité pour décrire la diversité des communautés complexes (voir article IV.B). Enfin, l'étude sur presque deux ans des communautés naturelles de tintinnidés décrites par les études moléculaires environnementales de diversité, et des populations naturelles d'*Undella claparedei* par amplification de marqueurs moléculaires par PCR directe, a permis d'appréhender pour la première fois la dynamique temporelle des tintinnidés par des approches moléculaires.

VI. Conclusion

Travailler avec les ciliés tintinnidés marins en combinant plusieurs approches moléculaires nous a révélé plusieurs résultats notables :

- ➔ L'amplification par PCR directe de l'ADNr 18S et des ITS sur cellules isolées est une méthode très efficace pour reconstruire une phylogénie à large échelle en peu de temps.
- ➔ Il apparaît que pour certaines espèces, derrière une grande diversité morphologique se cache une faible diversité génétique.
- ➔ L'étude comparative morphologique/moléculaire basée sur les tintinnidés nous révèle qu'en travaillant sur un groupe cohérent d'un point de vue taxonomique, les études moléculaires environnementales de diversité sont efficaces pour décrire la diversité identifiée sur des critères morphologiques.
- ➔ Malgré notre établissement d'une phylogénie de référence enrichie pour les espèces méditerranéennes, nous avons détecté deux UTOs, les clades environnementaux ETC-1 et ETC-2, qui sont très abondants en hiver et auxquels des espèces morphologiques n'ont pas pu être attribuées.
- ➔ En utilisant les tintinnidés comme modèle et différentes approches pour caractériser leur diversité, nous mettons en garde sur la méthode de plus en plus employée de mesure de la diversité microbienne par amplification de l'ADNr 18S (ou de l'ITS) et pyroséquençage. Une analyse rigoureuse des données de pyroséquençage, incluant un débruitage des données et un groupement des séquences en UTOs avec des paramètres ajustés en fonction du groupe d'étude, est nécessaire pour décrire justement la diversité microbienne avec cette approche moléculaire.
- ➔ La composition et l'abondance relative des communautés de tintinnidés varient avec le temps. Cette variation ne semble pas être corrélée de manière directe avec les paramètres physico-chimiques mesurés. Nous faisons l'hypothèse que les paramètres biotiques (proies, prédateurs et parasites) pourraient être les facteurs déterminants de ces variations.
- ➔ L'étude populationnelle d'*Undella claparedei* révèle une structuration temporelle des populations détectable au niveau du gène ARNr 18S. Derrière une apparente

VI. Conclusion

continuité morphologique, il existe des variations génétiques chez cette espèce qui sont le signe de phénomènes microévolutifs restant à décrire.

- ➔ Nous avons mis en évidence des signaux indirects d'épisodes de reproduction sexuelle (recombinaison au niveau des séquences, co-occurrence de différents marqueurs dans une cellule) au sein des populations naturelles d'*Undella claparedei*, dont la prévalence et l'effet sur la diversité génétique restent encore à éclaircir.

VI. Conclusion

Bibliographie

- Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle GUY, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup Ø, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MFJR** (2005) The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *J Eukaryot Microbiol* **52**:399-451
- Agatha S, Struder-Kypke MC** (in press) Systematics and evolution of tintinnid ciliates. In: Dolan JR, Agatha S, Coats DW, Montagnes DJS, Stoecker DK (eds) *Biology and Ecology of Tintinnid Ciliates: Models for Marine Plankton*. Wiley-Blackwell, Oxford
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-402
- Alverson AJ** (2008) Molecular Systematics and the Diatom Species. *Protist* **159**:339-353
- Andreoli I, Mangini L, Ferrantini F, Santangelo G, Verni F, Petroni G** (2009) Molecular phylogeny of unculturable Karyorelictea (Alveolata, Ciliophora). *Zool Scr* **38**:651-662
- Auinger BM, Pfandl K, Boenigk J** (2008) Improved Methodology for Identification of Protists and Microalgae from Plankton Samples Preserved in Lugol's Iodine Solution: Combining Microscopic Analysis with Single-Cell PCR. *Appl Environ Microbiol* **74**:2505-2510
- Baas-Becking LGM** (1934) *Geobiologie of inleiding tot de milieukunde* W.P. Van Stockum & Zoon, The Hague, the Netherlands
- Bachvaroff TR, Kim S, Guillou L, Delwiche CF, Coats DW** (2012) Molecular Diversity of the Syndinean Genus *Euduboscquella* Based on Single-Cell PCR Analysis. *Appl Environ Microbiol* **78**:334-345
- Bachy C, Gómez F, López-García P, Dolan JR, Moreira D** (2012) Molecular Phylogeny of Tintinnid Ciliates (Tintinnida, Ciliophora). *Protist*
- Bakker C, Phaff WJ** (1976) Tintinnida from coastal waters of the S.W.-Netherlands I. The genus *Tintinnopsis* Stein. *Hydrobiologia* **50**:101-111
- Baldauf SL** (2008) An overview of the phylogeny and diversity of eukaryotes. *Journal of Systematics and Evolution* **46**:263-273
- Balech E** (1959) Tintinninea del Mediterraneo. *Trab Inst Esp Oceanogr* **28**:1-88
- Barth D, Krenke S, Fokin SI, Berendonk TU** (2006) Intraspecific Genetic Variation in *Paramecium* Revealed by Mitochondrial Cytochrome c Oxidase I Sequences. *J Eukaryot Microbiol* **53**:20-25
- Bass D, Cavalier-Smith T** (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Microbiol* **54**:2393-404
- Bass D, Richards T, Matthai L, Marsh V, Cavalier-Smith T** (2007) DNA evidence for global dispersal and probable endemism of protozoa. *BMC Evol Biol* **7**:162
- Baum DA, Donoghue MJ** (1995) Choosing among alternative "phylogenetic" species concepts. *Syst Bot* **20**:560-573
- Béthoux JP, Prieur L, Bong JH** (1988) The Ligurian Current off the French Riviera. *Oceanologica Acta* **9**:59-67
- Binga EK, Lasken RS, Neufeld JD** (2008) Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* **2**:233-241
- Boenigk J, Pfandl K, Stadler P, Chatzinotas A** (2005) High diversity of the 'Spumella-like' flagellates: an investigation based on the SSU rRNA gene sequences of isolates from habitats located in six different geographic regions. *Environ Microbiol* **7**:685-697

- Bojanić N, Vidjak O, Šolić M, Krstulović N, Brautović I, Matijević S, Kušpilić G, Šestanović S, Ninčević Gladan Ž, Marasović I** (2012) Community structure and seasonal dynamics of tintinnid ciliates in Kaštela Bay (middle Adriatic Sea). *J Plankton Res* **34**:510-530
- Boltovskoy D, Dinofrio EO, Alder VA** (1990) Intraspecific variability in Antarctic tintinnids: the *Cymatocylis affinis/convallaria* species group. *J Plankton Res* **12**:403-413
- Brandt K** (1906) Die Tintinnodeen der Plankton-Expedition. Ergebnisse der Plankton-Expedition der Humboldt-Stiftung **3**:33
- Brandt K** (1907) Die Tintinnodeen der Plankton-Expedition. Ergebnisse der Plankton-Expedition der Humboldt-Stiftung **3**:1-499
- Brate J, Klaveness D, Rygh T, Jakobsen K, Shalchian-Tabrizi K** (2010) Telsonemia-specific environmental 18S rDNA PCR reveals unknown diversity and multiple marine-freshwater colonizations. *BMC Microbiol* **10**:168
- Burki F, Okamoto N, Pombert J-Fo, Keeling PJ** (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proceedings of the Royal Society B: Biological Sciences*
- Burki F, Shalchian-Tabrizi K, Pawlowski J** (2008) Phylogenomics reveals a new "megagroup" including most photosynthetic eukaryotes. *Biology Letters* **4**:366-369
- Bütschli O** (1880–1889) Protozoa. Abt. I (1880–1882) Sarkodina und Sporozoa. Abt. II (1883–1887) Mastigophora. Abt. III (1887–1889) Infusoria und System der Radiolaria. In: Bronn HG (ed) *Klassen und Ordnung des Thier-Reichs*. Winter, Leipzig, p 1-616, 617-1097, 1098-2035
- Calbet A, Landry MR** (2004) Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnol Oceanogr* **49**:7
- Campbell AS** (1942) The Oceanic Tintinnina of the Plankton Gathered during the Last Cruise of the CARNEGIE. Carnegie Institute of Washington, Publication 537
- Caron DA, Countway PD** (2009) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat Microb Ecol* **57**:227-238
- Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD, Dennett MR, Moran DM, Jones AC** (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* **75**:5797-808
- Casteleyn G, Adams NG, Vanormelingen P, Debeer A-E, Sabbe K, Vyverman W** (2009a) Natural Hybrids in the Marine Diatom *Pseudo-nitzschia pungens* (Bacillariophyceae): Genetic and Morphological Evidence. *Protist* **160**:343-354
- Casteleyn G, Evans K, Backeljau T, D'hondt S, Chepurnov V, Sabbe K, Vyverman W** (2009b) Lack of population genetic structuring in the marine planktonic diatom *Pseudo-nitzschia pungens* (Bacillariophyceae) in a heterogeneous area in the Southern Bight of the North Sea. *Mar Biol* **156**:1149-1158
- Castresana J** (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol* **17**:540-552
- Cavalier-Smith T** (1987) The Origin of Eukaryote and Archaeobacterial Cells. *Ann N Y Acad Sci* **503**:17-54
- Cavalier-Smith T** (1993) Kingdom protozoa and its 18 phyla. *Microbiological Reviews* **57**:953-994
- Cavalier-Smith T** (1998) A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* **73**:203-66
- Cavalier-Smith T** (2004) Only six kingdoms of life. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**:1251-1262
- Chatton E** (1920) Les Péridiniens parasites: morphologie, reproduction, ethologie. *Archives de Zoologie Expérimentale et Générale* **59**:1-475
- Chatton E** (1938) *Titre et travaux scientifique (1906-1937) de Edouard Chatton*, Sette, Sottano, Italy
- Claparède R-E, Lachmann J** (1868) *Etudes sur les infusoires et les rhizopodes*. H. Georg, Genève et Bale

- Copeland HF** (1938) The kingdoms of organisms. The Quaterly review of biology **13**:383-420
- Corliss JO** (1984) The kingdom Protista and its 45 phyla. Biosystems **17**:87-126
- Countway PD, Gast RJ, Dennett MR, Savai P, Rose JM, Caron DA** (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). Environ Microbiol **9**:1219-32
- Countway PD, Gast RJ, Savai P, Caron DA** (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. J Eukaryot Microbiol **52**:95 - 106
- Criscuolo A, Gribaldo S** (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol **10**:210
- D'Alelio D, Amato A, Kooistra WH, Procaccini G, Casotti R, Montresor M** (2009) Internal Transcribed Spacer Polymorphism in *Pseudo-nitzschia multistriata* (Bacillariophyceae) in the Gulf of Naples: Recent Divergence or Intraspecific Hybridization? Protist **160**:9-20
- Daday Ev** (1886) Ein kleiner Beitrag zur Kenntnis der Infusorien-Fauna des Golfes von Neapel. Mittheilungen aus der Zool. Station zu Neapel **6**:481-498
- Daday Ev** (1887) Monographie der Familie der Tintinnodeen. Mittheilungen aus der Zool. Station zu Neapel **7**:473-591
- Davis CC** (1955) The marine and freshwater plankton. Michigan State University Press, East Lansing
- Davis CC** (1981) Variations of lorica shape in the genus *Ptychocylis* (Protozoa: Tintinnina) in relation to species identification. J Plankton Res **3**:433-443
- de Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J** (1999) Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. Proceedings of the National Academy of Sciences **96**:2864-2868
- Decelle J, Suzuki N, Mahé F, de Vargas C, Not F** (2012) Molecular Phylogeny and Morphological Evolution of the Acantharia (Radiolaria). Protist **163**:435-450
- Diez B, Pedrós-Alió C, Massana R** (2001) Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing. Appl Environ Microbiol **67**:2932-2941
- Doherty M, Costas BA, McManus GB, Katz LA** (2007) Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. Aquat Microb Ecol **48**:141-154
- Doherty M, Tamura M, Costas BA, Ritchie ME, McManus GB, Katz LA** (2010) Ciliate diversity and distribution across an environmental and depth gradient in Long Island Sound, USA. Environ Microbiol **12**:886-98
- Dolan JR** (in press) Diversity and Distributions of Tintinnid Ciliates. In: Dolan JR, Agatha S, Coats DW, Montagnes DJS, Stoecker DK (eds) Biology and Ecology of Tintinnid Ciliates: Models for Marine Plankton. Wiley-Blackwell, Oxford
- Dolan JR, Claustre H, Carlotti F, Plounevez S, Moutin T** (2002) Microzooplankton diversity: relationships of tintinnid ciliates with resources, competitors and predators from the Atlantic Coast of Morocco to the Eastern Mediterranean. Deep-Sea Research part I **49**:1217-1232
- Dolan JR, Lemée R, Gasparini S, Mousseau L, Heyndrickx C** (2006) Probing diversity in the plankton: using patterns in tintinnids (planktonic marine ciliates) to identify patterns. Hydrobiologia **555**:143-157
- Duff RJ, Ball H, Lavrentyev PJ** (2008) Application of combined morphological-molecular approaches to the identification of planktonic protists from environmental samples. J Eukaryot Microbiol **55**:306-12
- Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C, Holder M, Taylor GT, Suarez P, Varela R, Epstein S** (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. ISME J

- Entz G** (1884) Über Infusorien des Golfes von Neapel. Mittheilungen aus der Zoologischen Station zu Neapel **5**:289-444
- Entz G** (1885) Zur nähren Kenntniss der Tintinnoden. Mittheilungen aus der Zoologischen Station zu Neapel **5**:185-216
- Evans KM, Kühn SF, Hayes PK** (2005) High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia pungens* (Bacillariophyceae) populations. *J Phycol* **41**:506-514
- Excoffier L, Laval G, Schneider S** (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Libertas Academica*
- Excoffier L, Smouse PE, Quattro JM** (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **131**:479-491
- Fauré-Fremiet E** (1924) Contributions à la connaissance des infusoires planktoniques. *Bulletin Biologique de France de Belgique* **58**:1-169
- Fenchel T** (2005) Cosmopolitan microbes and their 'cryptic' species. *Aquat Microb Ecol* **41**:49-54
- Finlay BJ** (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296**:1061-3
- Finlay BJ, Esteban GF, Brown S, Fenchel T, Hoef-Emden K** (2006) Multiple Cosmopolitan Ecotypes within a Microbial Eukaryote Morphospecies. *Protist* **157**:377-390
- Finlay BJ, Fenchel Tom** (2004) Cosmopolitan Metapopulations of Free-Living Microbial Eukaryotes. *Protist* **155**:237-244
- Foissner, Wilhelm** (2006) Biogeography and dispersal of micro-organisms : A review emphasizing protists. *Nencki Institute of Experimental Biology, Warszawa, POLOGNE*
- Foissner W** (1999) Protist Diversity: Estimates of the Near-Imponderable. *Protist* **150**:363-368
- Fol H** (1881) Contribution à la connaissance de la famille Tintinnodea. *Archs Sci Phys Nat* **5**:5-24+plate 1
- Gao S, Gong J, Lynn D, Lin X, Song W** (2009) An updated phylogeny of oligotrich and choreotrich ciliates (Protozoa, Ciliophora, Spirotrichea) with representative taxa collected from Chinese coastal waters. *Syst Biodivers* **7**:235-242
- Gaston KJ** (1994) *Rarity* Chapman & Hall, London
- Genetekaki E, Lynn DH** (2009) High-level genetic diversity but no population structure inferred from nuclear and mitochondrial markers of the peritrichous ciliate *Carchesium polypinum* in the Grand River basin (North America). *Appl Environ Microbiol* **75**:3187-95
- Gifford, D J, Dagg, M J** (1988) Feeding of the estuarine copepod *Acartia tonsa* dana: carnivory vs. herbivory in natural microplankton assemblages. *Allen, Lawrence, KA, ETATS-UNIS*
- Giovannoni SJ, Britschgi TB, Moyer CL, Field KG** (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**:60-63
- Gomaa F, Todorov M, Heger TJ, Mitchell EAD, Lara E** (2012) SSU rRNA Phylogeny of Arcellinida (Amoebozoa) Reveals that the Largest Arcellinid Genus, *Diffugia* Leclerc 1815, is not Monophyletic. *Protist* **163**:389-399
- Gómez F, López-García P, Moreira D** (2009) Molecular phylogeny of the ocelloid-bearing dinoflagellates erythrosideinium and warnowia (warnowiaceae, dinophyceae). *J Eukaryot Microbiol* **56**:440-445
- Gómez F, López-García P, Moreira D** (2011) Molecular phylogeny of dinophysoid dinoflagellates: the systematic position of *Oxyphysis oxytoxoides* and the *Dinophysis hastata* group (Dinophysales, Dinophyceae). *J Phycol* **47**:393-406
- Gómez F, Moreira D, Benzerara K, López-García P** (2010a) *Solenicola setigera* is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. *Environ Microbiol*:no-no

- Gómez F, Moreira D, López-García P** (2010b) Molecular Phylogeny of Noctiluroid Dinoflagellates (Noctilucales, Dinophyceae). *Protist* **161**:466-478
- Gómez F, Moreira D, López-García P** (2012) Sinophysis and Pseudophalacroma are Distantly Related to Typical Dinophysoid Dinoflagellates (Dinophysales, Dinophyceae). *J Eukaryot Microbiol* **59**:188-190
- Guillou L, Nézan E, Cueff V, Erard-Le Denn E, Cambon-Bonavita M-A, Gentien P, Barbier G** (2002) Genetic Diversity and Molecular Detection of Three Toxic Dinoflagellate Genera (Alexandrium, Dinophysis, and Karenia) from French Coasts. *Protist* **153**:223-238
- Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Meth B, DeSantis TZ, The Human Microbiome C, Petrosino JF, Knight R, Birren BW** (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**:494-504
- Hada Y** (1937) The fauna of Akkeshi Bay IV. The pelagic ciliata. *Journal of the Faculty of Science Hokkaido Imperial University, Series 6, Zoology* **5**:147-216
- Haeckel E** (1873) Ueber einige pelagische Infusorien. *Jenaische Zeitschrift für Medizin und Naturwissenschaft* **7**:561-568
- Haeckel EHPA** (1866) *Generelle morphologie der organismen*. G. Reimer, Berlin
- Harada A, Ohtsuka S, Horiguchi T** (2007) Species of the parasitic genus *Duboscquella* are members of the enigmatic Marine Alveolate Group I. *Protist* **158**:337-47
- Heinbokel JF** (1978) Studies on the functional role of tintinnids in the Southern California Bight. II. Grazing rates of field populations. *Mar Biol* **47**:191-197
- Heinbokel JF, Beers JR** (1979) Studies on the functional role of tintinnids in the Southern California Bight. III. Grazing impact of natural assemblages. *Mar Biol* **52**:23-32
- Heiss AA, Keeling PJ** (2006) The Phylogenetic Position of the Oxymonad *Saccinobaculus* Based on SSU rRNA. *Protist* **157**:335-344
- Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R** (2010) Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J*
- Hillis DM, Dixon MT** (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly review of biology* **66**:411-453
- Huelsenbeck JP, Ronquist F** (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754-755
- Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM** (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**:R143
- Huse SM, Welch DM, Morrison HG, Sogin ML** (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**:1889-1898
- Iglesias-Rodríguez MD, Schofield OM, Batley J, Medlin LK, Hayes PK** (2006) Intraspecific genetic diversity in the marine coccolithophore *Emiliania Huxleyi* (prymnesiophyceae) : The use of microsatellite analysis in marine phytoplankton population studies. *J Phycol* **42**:11
- Jobb G, von Haeseler A, Strimmer K** (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* **4**:18
- Jørgensen E** (1924) Mediterranean Tintinnidae. Report on the Danish oceanographical expeditions 1908–10 to the Mediterranean and adjacent seas. *Biology* **2**:1-110
- Katoh K, Misawa K, Kuma K, Miyata T** (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059-3066
- Katz LA, McManus GB, Snoeyenbos-West OLO, Griffin A, Pirog K, Costas B, Foissner W** (2005) Reframing the ‘Everything is everywhere’ debate: evidence for high gene flow and diversity in ciliate morphospecies. *Aquat Microb Ecol* **41**:55-65

- Katz LA, Snoeyenbos-West O, Doerder FP** (2006) Patterns of Protein Evolution in Tetrahymena thermophila: Implications for Estimates of Effective Population Size. *Mol Biol Evol* **23**:608-614
- Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW** (2005) The tree of eukaryotes. *Trends Ecol Evol* **20**:670-6
- Kim D, Countway P, Gast R, Caron D** (2011a) Rapid Shifts in the Structure and Composition of a Protistan Assemblage During Bottle Incubations Affect Estimates of Total Protistan Species Richness. *Microb Ecol* **62**:383-398
- Kim E, Harrison JW, Sudek S, Jones MD, Wilcox HM, Richards TA, Worden AZ, Archibald JM** (2011b) Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proc Natl Acad Sci U S A*
- Kim SY, Yang EJ, Gong JUN, Choi JK** (2010) Redescription of *Favella ehrenbergii* (Claparède and Lachmann, 1858) Jörgensen, 1924 (Ciliophora: Choreotrichia), with Phylogenetic Analyses Based on Small Subunit rRNA Gene Sequences. *J Eukaryot Microbiol* **57**:460-467
- Kofoed CA, Campbell AS** (1929) A conspectus of the marine and freshwater ciliata belonging to the suborder Tintinnoinea, with descriptions of new species principally from the Agassiz expedition to the Eastern Tropical Pacific 1904–1905. *Univ Calif Publ Zool*
- Kofoed CA, Campbell AS** (1939) The Ciliata: the Tintinnoinea. *Bull Mus Comp Zool Harvard*
- Krabberød AK, Bråte J, Dolven JK, Ose RF, Klaveness D, Kristensen T, Bjørklund KR, Shalchian-Tabrizi K** (2012) Radiolaria Divided into Polycystina and Spasmaria in Combined 18S and 28S rDNA Phylogeny. *PLoS ONE* **6**:e23526
- Kunin V, Engelbrektson A, Ochman H, Hugenholtz P** (2010) Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**:118-123
- Laackmann H** (1907) Antarktische Tintinnen. *Zool Anz* **31**:235-239
- Laackmann H** (1908) Ungeschlechtliche und geschlechtliche Fortpflanzung der Tintinnen. *Wissenschaftliche Meeresuntersuchungen, Abteilung Kiel* **10**:13-38+Plates1-3
- Laackmann H** (1910) Die Tintinnodean der Deutschen Sudpolar-expedition 1901-1903. **11**:340-396
- Lara E, Heger TJ, Ekelund F, Lamentowicz M, Mitchell EAD** (2008) Ribosomal RNA Genes Challenge the Monophyly of the Hyalospheniidae (Amoebozoa: Arcellinida). *Protist* **159**:165-176
- Lara E, Moreira D, Vereshchaka A, Lopez-Garcia P** (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonads. *Environ Microbiol* **11**:47-55
- Laval-Peuto M** (1977) Reconstruction d'une lorica de forme *Coxliella* par le trophonte nu de *Favella ehrenbergii* (Ciliata, Tintinnina). *C R Acad Sci* **284**:547-550
- Laval-Peuto M** (1981) Construction of the lorica in Ciliata Tintinnina. In vivo study of *Favella ehrenbergii*: variability of the phenotypes during the cycle, biology, statistics, biometry. *Protistologica* **17**:249-272
- Laval-Peuto M, Brownlee DC** (1986) Identification and systematics of the Tintinnina (Ciliophora): evaluation and suggestions for improvement. *Institut océanographique, Paris, FRANCE*
- Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, Østerås M, Farinelli L, Pawlowski J** (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proceedings of the National Academy of Sciences* **108**:13177-13182
- Leeuwenhoek Av** (1676) *Philosophical Transactions of the Royal Society of London*
- Li Z, Yi Z, Yang J, Gong J, Clamp JC, Al-Rasheid KAS, Al-Arifi S, Al-Khedhairi AA, Song W** (2009) Phylogenetic investigation on five genera of tintinnid ciliates (Ciliophora, Choreotrichia), based on the small subunit ribosomal RNA gene sequences. *Progress in Natural Science* **19**:1097–1101
- Liu H** (2009) Evolution, diversity, and biogeography in pelagic calcifying protists. *The State University of New Jersey, New Brunswick, New Jersey*, p 169

- Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, Not F, de Vargas C** (2009) Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci U S A* **106**:12803-8
- Lohmann H** (1908) Untersuchungen zur Feststellung des vollständigen Gehaltes des Meeres an Plankton. *Wissenschaftliche Meeresuntersuchungen Kiel* **10**:131-370
- Long EO, Dawid IB** (1980) Repeated genes in eukaryotes. *Annu Rev Biochem* **49**:727-764
- López-García P, Moreira D** (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Res Microbiol* **159**:67-73
- López-García P, Rodriguez-Valera F, Pedrós-Alió C, Moreira D** (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**:603-607
- Lowe CD, Day A, Kemp SJ, Montagnes DJS** (2005) There are High Levels of Functional and Genetic Diversity in *Oxyrrhis marina*. *J Eukaryot Microbiol* **52**:250-257
- Lynn DH** (2008) *The Ciliated Protozoa. Characterization, Classification, and Guide to the Literature.* Springer
- Margalef R** (1957) Fitoplancton de las costas de Blanes (Gerona) de agosto de 1952 a junio de 1956. *Investigaciones Pesqueras* **8**:89-95
- Martin DP, Williamson C, Posada D** (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**:260-2
- Massana R** (2011) Eukaryotic Picoplankton in Surface Oceans. *Annu Rev Microbiol* **65**:91-110
- Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, Valentin K, Pedrós-Alió C** (2004) Phylogenetic and Ecological Analysis of Novel Marine Stramenopiles. *Appl Environ Microbiol* **70**:3528-3534
- Massana R, Guillou L, Diez B, Pedrós-Alió C** (2002) Unveiling the Organisms behind Novel Eukaryotic Ribosomal DNA Sequences from the Ocean. *Appl Environ Microbiol* **68**:4554-4558
- Maynard Smith J** (1992) Analyzing the mosaic structure of genes. *J Mol Evol* **34**:126-129
- Mayr E** (1948) *The Bearing of the New Systematics on Genetical Problems The Nature of Species* Adv Genet. Academic Press, p 205-237
- McManus GB, Katz LA** (2009) Molecular and morphological methods for identifying plankton: what makes a successful marriage? *J Plankton Res* **31**:1119–1129
- Medinger R, Nolte V, Pandey RV, Jost S, Ottenwalder B, Schlotterer C, Boenigk J** (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol Ecol* **19 Suppl 1**:32-40
- Meunier A** (1910) *Microplancton des Mers de Barents et de Kara. Duc d'Orléans. Campagne Arctique de 1907.* Imprimerie scientifique Charles Bulens, Brussels
- Mills EL** (1989) *Biological Oceanography. An Early History, 1870-1960.* Cornell University Press
- Montagnes DJS, Lynn DH** (1987) A quantitative Protargol stain (QPS) for ciliates: Method description and test of its quantitative nature. *Mar Microb Food Webs* **2**:83-93
- 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, López-García P** (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol* **10**:31-38
- Müller OF** (1776) *Zoologica Danicae Prodrumus seu Animalium Daniae et Norvegiae indigenarum characteres, nomine, et synonyma imprimis popularium*
- Nanney DL, Park C, Preparata R, Simon EM** (1998) Comparison of Sequence Differences in a Variable 23S rRNA Domain among Sets of Cryptic Species of Ciliated Protozoa. *J Eukaryot Microbiol* **45**:91-100
- Nebel M, Pfabel C, Stock A, Dunthorn M, Stoeck T** (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. *Environmental Microbiology Reports* **3**:154-158

- Nival P, Nival S, Thiriot A** (1975) Influence des conditions hivernales sur les productions phyto-et zooplanctoniques en Méditerranée Nord-Occidentale. V. Biomasse et production zooplanctonique — relations phyto-zooplancton. *Mar Biol* **31**:249-270
- Not F, Valentin K, Romari K, Lovejoy C, Massana R, Tobe K, Vaultot D, Medlin LK** (2007) Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315**:253-5
- Orias E, Hashimoto N, Chau M-F, Higashinakagawa T** (1991) PCR Amplification of *Tetrahymena* rDNA Segments Starting with Individual Cells. *J Eukaryot Microbiol* **38**:306-311
- Patterson DJ** (1999) The Diversity of Eukaryotes. *The American Naturalist* **154**:96-124
- Paulmier G** (1995) Les tintinnides (Ciliophora, Oligotrichida, Tintinnina) des côtes Françaises de la Manche et de l'Atlantique. *Annales de la Société des Sciences Naturelles de la Charente-Maritime* **8**:453-487
- Paulmier G** (1997) Tintinnides (Ciliophora, Oligotrichida, Tintinnina) de l'Atlantique Boréal, de l'Océan Indien et quelques mers adjacentes: Méditerranée, Mer Caraïbe, Mer Rouge. Inventaire et distribution. Observations basée sur les loricas. Rapport IFREMER DRV/RH/97-17:191 pp
- Pawlowski J, Fahrni J, Lecroq B, Longet D, Cornelius N, Excoffier L, Cedhagen T, Gooday AJ** (2007) Bipolar gene flow in deep-sea benthic foraminifera. *Mol Ecol* **16**:4089-4096
- Pawlowski J, Holzmann M, Berney C, Fahrni J, Cedhagen T, Bowser SS** (2002) Phylogeny of Allogromiid foraminifera inferred from SSU rRNA gene sequences. *The Journal of Foraminiferal Research* **32**:334-343
- Pecher WT, Robledo JAF, Vasta GR** (2004) Identification of a Second rRNA Gene Unit in the *Perkinsus andrewsi* Genome. *J Eukaryot Microbiol* **51**:234-245
- Pedrós-Alió C** (2007) Dipping into the Rare Biosphere. *Science* **315**:192-193
- Philippe H** (1993) MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* **21**:5264-72
- Pierce RW, Turner JT** (1993) Global biogeography of marine tintinnids. *Mar Ecol Prog Ser* **94**:11-26
- Potvin M, Lovejoy C** (2009) PCR-Based Diversity Estimates of Artificial and Environmental 18S rRNA Gene Libraries. *J Eukaryot Microbiol* **56**:174-181
- Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT** (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Meth* **6**:639-641
- Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ** (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**:38
- Rambaut A** (2009) FIGTREE <http://tree.bio.ed.ac.uk/>.
- Rappe, Michael S, Giovannoni, Stephen J** (2003) The uncultured microbial majority. *Annual Reviews*, Palo Alto, CA, ETATS-UNIS
- Rassoulzadegan F** (1979) Évolution annuelle des ciliés pélagiques en Méditerranée nord-occidentale. II. Ciliés oligotriches. Tintinnides (Tintinnina). *Investigación Pesqueras* **43**:417-448
- Reeder J, Knight R** (2009) The 'rare biosphere': a reality check. *Nat Meth* **6**:636-637
- Rodriguez F, Derelle E, Guillou L, Le Gall F, Vaultot D, Moreau H** (2005) Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol* **7**:853-9
- Romari K, Vaultot D** (2004) Composition and Temporal Variability of Picoeukaryote Communities at a Coastal Site of the English Channel from 18S rDNA Sequences. *Limnol Oceanogr* **49**:784-798
- Rynearson TA, Armbrust EV** (2005) Maintenance of clonal diversity during a spring bloom of the centric diatom *Ditylum brightwellii*. *Mol Ecol* **14**:1631-1640

- Saccà A, Strüder-Kypke MC, Lynn DH** (2012) Redescription of *Rhizodomus tagatzi* (Ciliophora: Spirotrichea: Tintinnida), Based on Morphology and Small Subunit Ribosomal RNA Gene Sequence. *J Eukaryot Microbiol*:n/a-n/a
- Santoferrara LF, Alder VA** (2009) Morphological variability, spatial distribution and abundance of *Helicostomella* species (Ciliophora: Tintinnina) in relation to environmental factors (Argentine shelf; 40-55°S). *Sci Mar* **73**:701-716
- Santoferrara LF, McManus GB, Alder VA** (2012) Utility of Genetic Markers and Morphology for Species Discrimination within the Order Tintinnida (Ciliophora, Spirotrichea). *Protist*
- Sawyer S** (1989) Statistical tests for detecting gene conversion. *Mol Biol Evol* **6**:526-538
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF** (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**:7537-7541
- Schrank FvP** (1803) *Fauna boica. Durchgedachte Geschichte der in Baiern einheimischen und zahmen Thiere*
- Shendure J, Ji H** (2008) Next-generation DNA sequencing. *Nat Biotech* **26**:1135-1145
- Shimodaira H** (2002) An Approximately Unbiased Test of Phylogenetic Tree Selection. *Syst Biol* **51**:492-508
- Simpson AGB, Roger AJ** (2002) Eukaryotic Evolution: Getting to the Root of the Problem. *Curr Biol* **12**:R691-R693
- Sitran R, Bergamasco A, Decembrini F, Guglielmo L** (2007) Temporal succession of tintinnids in the northern Ionian Sea, Central Mediterranean. *J Plankton Res* **29**:495–508
- Slapeta J, López-García P, Moreira D** (2006) Global Dispersal and Ancient Cryptic Species in the Smallest Marine Eukaryotes. *Mol Biol Evol* **23**:23-29
- Smayda T** (1980) Phytoplankton species succession. In: Morris I (ed) *The physiological ecology of phytoplankton*. University of California Press, Berkeley, p 493-570
- Snoeyenbos-West OLO, Salcedo T, McManus GB, Katz LA** (2002) Insights into the diversity of choreotrich and oligotrich ciliates (Class: Spirotrichea) based on genealogical analyses of multiple loci. *Int J Syst Evol Microbiol* **52**:1901-1913
- Sogin ML** (1989) Evolution of Eukaryotic Microorganisms and Their Small Subunit Ribosomal RNAs. *Am Zool* **29**:487-499
- Sogin ML, Elwood HJ, Gunderson JH** (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proceedings of the National Academy of Sciences* **83**:1383-1387
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ** (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* **103**:12115 - 12120
- Stanier RY, Niel CB** (1962) The concept of a bacterium. *Arch Microbiol* **42**:17-35
- Stoeck T, Behnke A, Christen R, Amaral-Zettler L, Rodriguez-Mora M, Chistoserdov A, Orsi W, Edgcomb V** (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biology* **7**:72
- Stoeck T, Epstein S** (2009) Protists and the rare biosphere. *Crystal Ball. Environ Microbiol Reports* **1**:3 - 26
- Strüder-Kypke MC, Lynn DH** (2003) Sequence analyses of the small subunit rRNA gene confirm the paraphyly of oligotrich ciliates sensu lato and support the monophyly of the subclasses Oligotrichia and Choreotrichia (Ciliophora, Spirotrichea). *Journal of Zoology* **260**:87-97
- Strüder-Kypke MC, Lynn DH** (2008) Morphological versus molecular data - Phylogeny of tintinnid ciliates (Ciliophora, Choreotrichia) inferred from small subunit rRNA gene sequences. *Denisia* **23**:417-424

- Takano Y, Hagino K, Tanaka Y, Horiguchi T, Okada H** (2006) Phylogenetic affinities of an enigmatic nanoplankton, *Braarudosphaera bigelowii* based on the SSU rDNA sequences. *Mar Micropaleontol* **60**:145-156
- Tamura M, Katz LA, McManus GB** (2011) Distribution and diversity of oligotrich and choreotrich ciliates across an environmental gradient in a large temperate estuary. *Aquat Microb Ecol* **64**:51-67
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA** (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Meth* **6**:377-382
- Utermöhl H** (1958) Zur Vervollkommnung der quantitativen Phytoplankton-Methodik
- Van Hoek AHAM, Van Alen TA, Vogels GD, Hackstein JHP** (2006) Contribution by the methanogenic endosymbionts of anaerobic ciliates to methane production in dutch freshwater sediments. *Acta Protozool* **45**:10
- Vigil P, Countway PD, Rose J, Lonsdale DJ, Gobler C, Caron DA** (2009) Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. *Aquat Microb Ecol* **54**:83-100
- Walker G, Dorrell RG, Schlacht A, Dacks JB** (2011) Eukaryotic systematics: a user's guide for cell biologists and parasitologists. *Parasitology* **138**:1638-1663
- Wasik A, Mikolajczyk E** (1994) Annual cycle of tintinnids in Admiralty Bay with an emphasis on seasonal variability in *Cymatocylis affinis/convallaria* lorica morphology. *J Plankton Res* **16**:1-8
- Whittaker RH** (1969) New Concepts of Kingdoms of Organisms. *Science* **163**:150-160
- Woese CR, Fox GE** (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* **74**:5088-5090
- Woese CR, Kandler O, Wheelis ML** (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**:4576-4579
- Wu S, Zhu Z, Fu L, Niu B, Li W** (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12**:444
- Yuasa T, Takahashi O, Dolven JK, Mayama S, Matsuoka A, Honda D, Bjorklund KR** (2006) Phylogenetic position of the small solitary phaeodarians (Radiolaria) based on 18S rDNA sequences by single cell PCR analysis. *Mar Micropaleontol* **59**:104-114
- Zuckerkindl E, Pauling L** (1965) Molecules as documents of evolutionary history. *J Theor Biol* **8**:357-366

Table des Annexes

Annexe 1. Liste des specimens de tintinnidés isolés pendant le suivi temporel dans la rade de Villefranche-sur-Mer au point B+.....p 220

Annexe 2. Liste des échantillons d'eau marine collectés à la bouteille Niskin pendant le suivi temporel dans la rade de Villefranche-sur-Mer au point B+.....p 240

Annexe 3. Article sur la diversité des eucaryotes au pôle Nord dans la neige, la banquise et la colonne d'eau à la fin de la nuit polaire.....p 242

Annexes

Annexe 1. Liste des spécimens de tintinnidés isolés pendant le suivi temporel dans la rade de Villefranche-sur-Mer au point B+

Espèce/genre	Numérotation	Date
Undella claparedei	CB1	18/02/2009
Undella claparedei	CB2	18/02/2009
Undella claparedei	CB3	18/02/2009
Undella claparedei	CB4	18/02/2009
Stenosemella nivalis	CB5	18/02/2009
Stenosemella nivalis	CB6	18/02/2009
Stenosemella nivalis	CB7	18/02/2009
Codonella nationalis	CB8	18/02/2009
Undella claparedei	CB9	18/02/2009
Undella claparedei	CB10	18/02/2009
Undella claparedei	CB11	18/02/2009
Codonella nationalis	CB12	18/02/2009
Petalotricha ampulla	CB13	18/02/2009
Codonella aspera	CB14	18/02/2009
Codonella nationalis	CB15	18/02/2009
Undella claparedei	CB16	18/02/2009
Undella claparedei	CB17	18/02/2009
Undella claparedei	CB18	18/02/2009
Petalotricha ampulla	CB19	18/02/2009
Codonellopsis morchella	CB20	18/02/2009
Codonella aspera	CB21	18/02/2009
Petalotricha ampulla	CB22	18/02/2009
Undella claparedei	CB23	18/02/2009
Undella claparedei	CB24	18/02/2009
Codonella nationalis	CB25	18/02/2009
Codonella nationalis	CB26	18/02/2009
Undella claparedei	CB27	18/02/2009
Codonella nationalis	CB28	18/02/2009
Codonella aspera	CB29	18/02/2009
Undella claparedei	CB30	18/02/2009
Undella claparedei	CB31	18/02/2009
Stenosemella nivalis	CB32	18/02/2009
Stenosemella nivalis	CB33	18/02/2009
Stenosemella nivalis	CB34	18/02/2009
Stenosemella nivalis	CB35	18/02/2009
Stenosemella nivalis	CB36	18/02/2009

Annexes

Codonella aspera	CB37	18/02/2009
Codonella aspera	CB38	18/02/2009
Codonella aspera	CB39	18/02/2009
Codonella aspera	CB40	18/02/2009
Codonella nationalis	CB41	18/02/2009
Codonella nationalis	CB42	18/02/2009
Stenosemella nivalis	CB43	18/02/2009
Stenosemella nivalis	CB44	18/02/2009
Codonella nationalis	CB45	18/02/2009
Stenosemella nivalis	CB46	18/02/2009
Stenosemella nivalis	CB47	18/02/2009
Stenosemella nivalis	CB48	18/02/2009
Codonella aspera	CB49	18/02/2009
Codonella nationalis	CB50	18/02/2009
Codonella aspera	CB51	18/02/2009
Codonella aspera	CB52	18/02/2009
Codonella nationalis	CB53	18/02/2009
Codonella aspera	CB54	18/02/2009
Cyttarocyliis sp.	CB55	18/02/2009
Codonella nationalis	CB56	18/02/2009
Rhabdonella spiralis	CB57	18/02/2009
Codonellopsis morchella	CB58	18/02/2009
Petalotricha ampulla	CB59	18/02/2009
Cyttarocyliis sp.	CB60	18/02/2009
Codonellopsis morchella	CB61	18/02/2009
Stenosemella nivalis	CB62	18/02/2009
Codonella aspera	CB64	18/02/2009
Codonella nationalis	CB65	18/02/2009
Stenosemella nivalis	CB66	18/02/2009
Dictyocysta sp.	CB67	18/02/2009
Cyttarocyliis sp.	CB68	18/02/2009
codonella amphorella	CB69	18/02/2009
Dictyocysta sp.	CB70	18/02/2009
Undella marsupialis	CB71	18/02/2009
Cyttarocyliis sp.	CB72	18/02/2009
Stenosemella nivalis	CB73	20/02/2009
Codonella nationalis	CB74	20/02/2009
Codonella nationalis	CB75	20/02/2009
Petalotricha ampulla	CB76	20/02/2009
Codonellopsis morchella	CB77	20/02/2009
Stenosemella nivalis	CB78	20/02/2009
Stenosemella nivalis	CB79	20/02/2009
Stenosemella nivalis	CB80	20/02/2009
Codonella aspera	CB81	20/02/2009

Annexes

Codonella nationalis	CB82	20/02/2009
Petalotricha ampulla	CB83	20/02/2009
Cyttarocyliis sp.	CB84	20/02/2009
Codonella nationalis	CB85	20/02/2009
Xystonella sp. longicauda	CB86	18/02/2009
Stenosemella nivalis	CB88	18/02/2009
Salpingella acuminata	CB89	18/02/2009
Undella marsupialis	CB90	18/02/2009
Undella claparedei	CB91	18/02/2009
Undella claparedei	CB92	18/02/2009
Undella claparedei	CB93	18/02/2009
Undella claparedei	CB94	18/02/2009
Undella claparedei	CB95	18/02/2009
Undella claparedei	CB96	18/02/2009
Undella claparedei	CB97	18/02/2009
Undella claparedei	CB98	18/02/2009
Undella claparedei	CB99	18/02/2009
Undella claparedei	CB100	18/02/2009
Undella claparedei	CB101	18/02/2009
Undella claparedei	CB102	18/02/2009
Undella claparedei	CB103	18/02/2009
Undella claparedei	CB104	18/02/2009
Undella claparedei	CB105	18/02/2009
Undella claparedei	CB106	18/02/2009
Undella claparedei	CB107	18/02/2009
Undella claparedei	CB108	18/02/2009
Undella claparedei	CB109	18/02/2009
Undella claparedei	CB110	18/02/2009
Undella claparedei	CB111	18/02/2009
Undella claparedei	CB112	18/02/2009
Undella claparedei	CB113	18/02/2009
Undella claparedei	CB114	18/02/2009
Undella claparedei	CB115	18/02/2009
Undella claparedei	CB116	18/02/2009
Stenosemella nivalis	CB117	18/03/2009
Stenosemella nivalis	CB118	18/03/2009
Stenosemella nivalis	CB119	18/03/2009
Stenosemella nivalis	CB120	18/03/2009
Stenosemella nivalis	CB121	18/03/2009
Undella claparedei	CB122	18/03/2009
Stenosemella nivalis	CB123	18/03/2009
Stenosemella nivalis	CB124	18/03/2009
Stenosemella nivalis	CB125	18/03/2009
Stenosemella nivalis	CB126	18/03/2009

Annexes

Stenosemella nivalis	CB127	18/03/2009
Stenosemella nivalis	CB128	18/03/2009
Undella claparedei	CB129	18/03/2009
Stenosemella nivalis	CB130	18/03/2009
Stenosemella nivalis	CB131	18/03/2009
Stenosemella nivalis	CB132	18/03/2009
Stenosemella nivalis	CB133	18/03/2009
Stenosemella nivalis	CB134	18/03/2009
Stenosemella nivalis	CB135	18/03/2009
Stenosemella nivalis	CB136	18/03/2009
Stenosemella nivalis	CB137	18/03/2009
Stenosemella nivalis	CB138	18/03/2009
Stenosemella nivalis	CB139	18/03/2009
Stenosemella nivalis	CB140	18/03/2009
dadayella ganymedes	CB141	18/03/2009
Rhabdonella spiralis	CB142	18/03/2009
Codonella nationalis	CB143	18/03/2009
Codonella nationalis	CB144	18/03/2009
Codonella nationalis	CB145	18/03/2009
Salpingella acuminata	CB146	18/03/2009
Codonella aspera	CB147	18/03/2009
Codonella nationalis	CB148	18/03/2009
Codonella nationalis	CB149	18/03/2009
Codonella aspera	CB150	18/03/2009
Codonellopsis morchella	CB151	18/03/2009
Codonellopsis morchella	CB152	18/03/2009
Codonellopsis morchella	CB153	18/03/2009
Codonellopsis morchella	CB154	18/03/2009
Codonellopsis morchella	CB155	18/03/2009
Codonellopsis morchella	CB156	18/03/2009
Codonellopsis morchella	CB157	18/03/2009
Codonellopsis morchella	CB158	18/03/2009
Codonellopsis morchella	CB159	18/03/2009
Undella claparedei	CB160	18/03/2009
codonella amphorella	CB161	18/03/2009
Undella claparedei	CB162	18/03/2009
Stenosemella nivalis	CB163	18/03/2009
Codonella nationalis	CB164	18/03/2009
Codonella nationalis	CB165	18/03/2009
Undella claparedei	CB166	18/03/2009
Codonella nationalis	CB167	18/03/2009
Undella claparedei	CB168	18/03/2009
Codonella nationalis	CB169	18/03/2009
codonella amphorella	CB170	18/03/2009

Annexes

Codonella nationalis	CB171	18/03/2009
Stenosemella nivalis	CB172	18/03/2009
Stenosemella nivalis	CB173	18/03/2009
Codonella nationalis	CB174	18/03/2009
Undella claparedei	CB175	18/03/2009
Dictyocysta sp.	CB176	18/03/2009
Undella claparedei	CB177	18/03/2009
Codonaria sp.	CB178	18/03/2009
Codonella nationalis	CB179	18/03/2009
Codonella nationalis	CB180	18/03/2009
Codonella nationalis	CB181	18/03/2009
Codonella nationalis	CB182	18/03/2009
Codonella aspera	CB183	18/03/2009
Codonella nationalis	CB184	18/03/2009
Codonella nationalis	CB185	18/03/2009
Stenosemella nivalis	CB186	18/03/2009
Codonella aspera	CB187	18/03/2009
Stenosemella nivalis	CB188	18/03/2009
Codonella nationalis	CB189	18/03/2009
Codonella nationalis	CB190	18/03/2009
Undella claparedei	CB191	16/04/2009
Undella claparedei	CB192	16/04/2009
Undella claparedei	CB193	16/04/2009
Stenosemella nivalis	CB194	16/04/2009
Undella claparedei	CB195	16/04/2009
Undella claparedei	CB196	16/04/2009
Undella marsupialis	CB197	16/04/2009
Undella claparedei	CB198	16/04/2009
Codonella aspera	CB199	16/04/2009
Codonella aspera	CB200	16/04/2009
Stenosemella nivalis	CB201	16/04/2009
Stenosemella nivalis	CB202	16/04/2009
Codonella aspera	CB203	16/04/2009
Codonella nationalis	CB204	16/04/2009
Undella claparedei	CB205	16/04/2009
Undella claparedei	CB206	16/04/2009
Undella claparedei	CB207	16/04/2009
Undella claparedei	CB208	16/04/2009
Undella claparedei	CB209	16/04/2009
Undella claparedei	CB210	16/04/2009
Undella claparedei	CB211	16/04/2009
Undella claparedei	CB212	16/04/2009
Stenosemella nivalis	CB213	16/04/2009
Stenosemella nivalis	CB214	16/04/2009

Annexes

Undella claparedei	CB215	16/04/2009
Stenosemella nivalis	CB216	16/04/2009
Undella claparedei	CB217	16/04/2009
Stenosemella nivalis	CB218	16/04/2009
Undella claparedei	CB219	16/04/2009
Undella claparedei	CB220	16/04/2009
Codonella nationalis	CB221	16/04/2009
Stenosemella nivalis	CB222	16/04/2009
Undella claparedei	CB223	16/04/2009
Undella claparedei	CB224	16/04/2009
Undella claparedei	CB225	16/04/2009
Codonella nationalis	CB226	16/04/2009
Stenosemella nivalis	CB227	16/04/2009
Undella claparedei	CB228	16/04/2009
Undella claparedei	CB229	16/04/2009
Undella claparedei	CB230	16/04/2009
Undella claparedei	CB231	16/04/2009
Stenosemella nivalis	CB232	16/04/2009
Codonella aspera	CB233	16/04/2009
Codonella nationalis	CB234	16/04/2009
Codonella aspera	CB235	16/04/2009
Stenosemella nivalis	CB236	16/04/2009
Stenosemella nivalis	CB237	16/04/2009
Stenosemella nivalis	CB238	16/04/2009
Undella claparedei	CB239	16/04/2009
Undella claparedei	CB240	16/04/2009
Undella claparedei	CB241	16/04/2009
Stenosemella nivalis	CB242	16/04/2009
Codonella aspera	CB243	16/04/2009
Stenosemella nivalis	CB244	16/04/2009
Undella claparedei	CB245	16/04/2009
Stenosemella nivalis	CB246	27/05/2009
Stenosemella nivalis	CB247	27/05/2009
Salpingella acuminata	CB248	27/05/2009
Undella claparedei	CB249	27/05/2009
Rhabdonella spiralis	CB250	27/05/2009
Xystonellopsis sp.	CB251	27/05/2009
Undella marsupialis	CB252	27/05/2009
Rhabdonella spiralis	CB254	27/05/2009
Undella marsupialis	CB256	27/05/2009
Undella claparedei	CB258	27/05/2009
Undella claparedei	CB259	27/05/2009
Undella marsupialis	CB260	27/05/2009
Undella claparedei	CB261	27/05/2009

Annexes

Undella claparedei	CB262	27/05/2009
Undella claparedei	CB263	27/05/2009
Undella marsupialis	CB264	27/05/2009
Xystonellopsis sp.	CB265	27/05/2009
Favella sp.	CB267	29/05/2009
Rhabdonella spiralis	CB268	29/05/2009
Rhabdonella spiralis	CB269	29/05/2009
Salpingella acuminata	CB270	29/05/2009
Undella claparedei	CB271	25/06/2009
Undella claparedei	CB272	25/06/2009
Undella claparedei	CB273	25/06/2009
Undella claparedei	CB274	25/06/2009
Undella claparedei	CB275	25/06/2009
Undella claparedei	CB276	25/06/2009
Xystonella sp.	CB277	25/06/2009
Undella claparedei	CB278	25/06/2009
Undella claparedei	CB279	25/06/2009
Undella claparedei	CB280	25/06/2009
Eutintinnus sp.	CB281	25/06/2009
Undella claparedei	CB282	25/06/2009
Rhabdonella spiralis	CB283	25/06/2009
Undella claparedei	CB284	25/06/2009
Undella claparedei	CB285	25/06/2009
Undella claparedei	CB286	25/06/2009
Undella claparedei	CB287	25/06/2009
Rhabdonella spiralis	CB288	25/06/2009
Undella claparedei	CB289	25/06/2009
Undella claparedei	CB290	25/06/2009
Undella claparedei	CB291	25/06/2009
Xystonella sp.	CB292	25/06/2009
Undella claparedei	CB293	25/06/2009
Undella claparedei	CB294	25/06/2009
Undella claparedei	CB295	25/06/2009
Salpingella acuminata	CB296	25/06/2009
Eutintinnus sp.	CB297	25/06/2009
Eutintinnus sp.	CB298	25/06/2009
Codonella aspera	CB299	25/06/2009
Undella claparedei	CB300	25/06/2009
Rhabdonella spiralis	CB301	25/06/2009
Salpingella acuminata	CB302	25/06/2009
Salpingella acuminata	CB303	25/06/2009
Eutintinnus sp.	CB304	25/06/2009
Undella claparedei	CB305	25/06/2009
Rhabdonella spiralis	CB306	25/06/2009

Annexes

Undella claparedei	CB307	25/06/2009
Undella claparedei	CB308	25/06/2009
Undella claparedei	CB309	25/06/2009
Salpingella acuminata	CB310	25/06/2009
Rhabdonella spiralis	CB311	25/06/2009
Undella claparedei	CB312	25/06/2009
Rhabdonella spiralis	CB313	25/06/2009
Undella hyalina	CB314	25/06/2009
Rhabdonella spiralis	CB315	25/06/2009
Salpingella acuminata	CB316	25/06/2009
Salpingella acuminata	CB317	25/06/2009
Undella claparedei	CB318	25/06/2009
Xystonella sp. climacocylis	CB319	25/06/2009
Xystonella sp.	CB320	25/06/2009
Undella claparedei	CB321	25/06/2009
Rhabdonella spiralis	CB322	25/06/2009
Rhabdonella spiralis	CB323	25/06/2009
Rhabdonella spiralis	CB324	25/06/2009
Eutintinnus sp.	CB325	25/06/2009
Undella marsupialis	CB326	25/06/2009
Salpingella acuminata	CB327	25/06/2009
Salpingella acuminata	CB328	25/06/2009
Undella marsupialis	CB329	25/06/2009
Undella claparedei	CB330	25/06/2009
codonella	CB331	25/06/2009
Rhabdonella spiralis	CB332	25/06/2009
Petalotricha ampulla	CB333	25/06/2009
Salpingella acuminata	CB334	25/06/2009
Xystonella sp.	CB335	25/06/2009
Salpingella acuminata	CB336	25/06/2009
Salpingella acuminata	CB337	25/06/2009
Rhabdonella spiralis	CB339	24/07/2009
Rhabdonella spiralis	CB340	24/07/2009
Eutintinnus sp.	CB341	24/09/2009
Eutintinnus sp.	CB342	24/09/2009
Eutintinnus sp.	CB343	24/09/2009
Tintinnopsis sp	CB344	24/09/2009
Eutintinnus sp.	CB345	24/09/2009
Eutintinnus sp.	CB346	24/09/2009
Eutintinnus sp.	CB347	24/09/2009
Eutintinnus sp.	CB348	24/09/2009
Xystonella sp.	CB349	24/09/2009
Eutintinnus apertus	CB350	24/09/2009
Salpingella acuminata	CB351	24/09/2009

Annexes

Eutintinnus sp.	CB352	24/09/2009
Eutintinnus sp.	CB353	24/09/2009
Eutintinnus sp.	CB354	24/09/2009
Dadayella ganymedes	CB355	24/09/2009
Rhabdonella spiralis	CB356	24/09/2009
Tintinnopsis sp	CB357	24/09/2009
Tintinnopsis sp	CB358	24/09/2009
Tintinnopsis sp	CB359	24/09/2009
Eutintinnus sp.	CB360	24/09/2009
Salpingella acuminata	CB361	24/09/2009
Codonellopsis morchella	CB362	22/10/2009
Salpingella acuminata	CB363	22/10/2009
Salpingella acuminata	CB364	22/10/2009
Eutintinnus sp.	CB365	22/10/2009
Xystonella sp.	CB366	22/10/2009
Salpingella acuminata	CB367	22/10/2009
Eutintinnus sp.	CB368	22/10/2009
Codonella sp.	CB369	22/10/2009
Codonellopsis morchella	CB370	22/10/2009
Salpingella acuminata	CB371	22/10/2009
Xystonella sp.	CB372	22/10/2009
Codonellopsis morchella	CB373	22/10/2009
Salpingella acuminata	CB374	22/10/2009
Codonella sp.	CB375	22/10/2009
Salpingella acuminata	CB376	22/10/2009
Codonellopsis morchella	CB377	22/10/2009
Dictyocysta sp.	CB378	22/10/2009
Salpingella acuminata	CB379	22/10/2009
Codonella sp.	CB380	22/10/2009
Codonellopsis morchella	CB381	22/10/2009
Xystonellopsis	CB382	22/10/2009
Codonella	CB383	22/10/2009
Eutintinnus sp.	CB384	22/10/2009
Xystonella sp.	CB385	22/10/2009
Xystonella sp.	CB386	22/10/2009
Salpingella acuminata	CB387	22/10/2009
Xystonella sp.	CB388	22/10/2009
Eutintinnus sp. apertus	CB389	22/10/2009
Salpingella acuminata	CB390	22/10/2009
Codonella	CB391	22/10/2009
Codonella	CB392	22/10/2009
Codonellopsis morchella	CB393	22/10/2009
Codonella	CB394	22/10/2009
Salpingella acuminata	CB395	22/10/2009

Annexes

Codonellopsis morchella	CB396	22/10/2009
Codonella	CB397	22/10/2009
Codonaria sp./codonella	CB398	22/10/2009
Codonella	CB399	22/10/2009
Codonellopsis morchella	CB400	22/10/2009
Xystonella sp.	CB401	22/10/2009
Codonellopsis morchella	CB402	22/10/2009
Codonella	CB403	22/10/2009
Codonella	CB404	22/10/2009
Codonella	CB405	22/10/2009
Salpingella acuminata	CB406	22/10/2009
Codonella	CB407	22/10/2009
Codonellopsis morchella	CB408	23/10/2009
Codonellopsis morchella	CB409	23/10/2009
Codonellopsis morchella	CB410	23/10/2009
Codonella	CB411	23/10/2009
Undella marsupialis	CB412	23/10/2009
Undella marsupialis	CB413	23/10/2009
Salpingella acuminata	CB414	23/10/2009
Salpingella acuminata	CB415	23/10/2009
Climacocylis	CB416	23/10/2009
Salpingella acuminata	CB417	23/10/2009
Codonella	CB418	23/10/2009
Undella claparedei	CB419	18/11/2009
Undella claparedei	CB420	18/11/2009
Undella claparedei	CB421	18/11/2009
Undella claparedei	CB422	18/11/2009
Eutintinnus sp.	CB423	18/11/2009
Codonellopsis morchella	CB424	18/11/2009
Undella claparedei	CB425	18/11/2009
Codonellopsis morchella	CB426	18/11/2009
Undella claparedei	CB427	18/11/2009
Eutintinnus sp.	CB428	18/11/2009
codonella	CB429	18/11/2009
Undella claparedei	CB430	18/11/2009
Eutintinnus sp.	CB431	18/11/2009
Amphorides	CB432	18/11/2009
Undella claparedei	CB433	18/11/2009
Eutintinnus sp.	CB434	18/11/2009
Undella claparedei	CB435	18/11/2009
Amphorides	CB436	18/11/2009
Salpingella acuminata	CB437	18/11/2009
Undella marsupialis	CB438	18/11/2009
Undella marsupialis	CB439	18/11/2009

Annexes

Undella claparedei	CB440	18/11/2009
Codonellopsis morchella	CB441	18/11/2009
Codonellopsis morchella	CB442	18/11/2009
Undella claparedei	CB443	18/11/2009
Codonellopsis morchella	CB444	18/11/2009
Codonellopsis morchella	CB445	18/11/2009
Eutintinnus sp.	CB446	19/11/2009
Eutintinnus sp.	CB447	19/11/2009
Dictyocysta sp.	CB448	19/11/2009
Dictyocysta sp.	CB449	19/11/2009
Dictyocysta sp.	CB450	19/11/2009
Rhabdonella spiralis	CB451	19/11/2009
Dictyocysta sp.	CB452	19/11/2009
Dictyocysta sp.	CB453	19/11/2009
Eutintinnus sp.	CB454	19/11/2009
Eutintinnus sp.	CB455	19/11/2009
Eutintinnus sp.	CB456	19/11/2009
Dictyocysta sp.	CB457	19/11/2009
Eutintinnus sp.	CB458	19/11/2009
Eutintinnus sp.	CB459	19/11/2009
Dictyocysta sp.	CB460	19/11/2009
Dictyocysta sp.	CB461	19/11/2009
codonella	CB462	19/11/2009
Eutintinnus sp.	CB463	19/11/2009
codonella	CB464	19/11/2009
Dictyocysta sp.	CB465	19/11/2009
codonella	CB466	19/11/2009
Dictyocysta sp.	CB467	19/11/2009
Dictyocysta sp.	CB468	19/11/2009
codonella	CB469	19/11/2009
codonella	CB470	19/11/2009
codonella	CB471	19/11/2009
codonella	CB472	19/11/2009
codonella	CB473	19/11/2009
codonella	CB474	19/11/2009
Dictyocysta sp.	CB475	19/11/2009
Dictyocysta sp.	CB476	19/11/2009
Salpingella acuminata	CB477	18/11/2009
Undella claparedei	CB478	18/11/2009
Salpingella acuminata	CB479	18/11/2009
Undella claparedei	CB480	18/11/2009
Undella claparedei	CB481	18/11/2009
Undella claparedei	CB482	16/12/2009
Undella claparedei	CB483	16/12/2009

Annexes

Undella claparedei	CB484	16/12/2009
Undella claparedei	CB485	16/12/2009
Undella claparedei	CB486	16/12/2009
Undella claparedei	CB487	16/12/2009
Undella claparedei	CB488	16/12/2009
Undella claparedei	CB489	16/12/2009
Dictyocysta sp.	CB490	16/12/2009
Undella claparedei	CB491	16/12/2009
Undella claparedei	CB492	16/12/2009
Undella claparedei	CB493	16/12/2009
Undella claparedei	CB494	16/12/2009
Undella claparedei	CB495	16/12/2009
Undella claparedei	CB496	16/12/2009
Undella claparedei	CB497	16/12/2009
Undella claparedei	CB498	16/12/2009
Undella claparedei	CB499	16/12/2009
Undella claparedei	CB500	16/12/2009
Undella claparedei	CB501	16/12/2009
Undella claparedei	CB502	16/12/2009
Undella claparedei	CB503	16/12/2009
Petalotricha ampulla	CB504	16/12/2009
Undella claparedei	CB505	16/12/2009
Undella marsupialis	CB506	16/12/2009
Undella claparedei	CB507	16/12/2009
Undella claparedei	CB508	16/12/2009
Undella claparedei	CB509	16/12/2009
Undella claparedei	CB510	16/12/2009
Petalotricha ampulla	CB511	16/12/2009
Epiplocylyis sp.	CB512	16/12/2009
Undella claparedei	CB513	16/12/2009
Undella claparedei	CB514	16/12/2009
Undella marsupialis	CB515	16/12/2009
Undella marsupialis	CB516	16/12/2009
Undella claparedei	CB517	16/12/2009
Undella claparedei	CB518	16/12/2009
Undella claparedei	CB519	16/12/2009
Undella claparedei	CB520	16/12/2009
Undella claparedei	CB521	16/12/2009
Undella claparedei	CB522	16/12/2009
Undella claparedei	CB523	16/12/2009
Undella claparedei	CB524	16/12/2009
Undella claparedei	CB525	16/12/2009
Undella claparedei	CB526	16/12/2009
Undella claparedei	CB527	16/12/2009

Undella claparedei	CB528	16/12/2009
Undella claparedei	CB529	16/12/2009
Undella claparedei	CB530	16/12/2009
Undella claparedei	CB531	16/12/2009
Undella claparedei	CB532	16/12/2009
Undella claparedei	CB533	17/12/2009
Undella claparedei	CB534	17/12/2009
Undella claparedei	CB535	17/12/2009
Undella claparedei	CB536	17/12/2009
Undella claparedei	CB537	17/12/2009
Undella claparedei	CB538	17/12/2009
Undella claparedei	CB539	16/12/2009
Undella claparedei	CB540	16/12/2009
Undella claparedei	CB541	16/12/2009
Undella claparedei	CB542	16/12/2009
Undella claparedei	CB543	16/12/2009
Undella claparedei	CB544	16/12/2009
Undella claparedei	CB545	16/12/2009
Undella claparedei	CB546	16/12/2009
Undella claparedei	CB547	16/12/2009
Undella claparedei	CB548	16/12/2009
Epiplocylys sp.	CB549	16/12/2009
Epiplocylys sp.	CB550	16/12/2009
Salpingella acuminata	CB551	16/12/2009
Salpingella acuminata	CB552	16/12/2009
Salpingella acuminata	CB553	16/12/2009
Undella marsupialis	CB554	16/12/2009
Amphorides quadrilineata	CB555	16/12/2009
Codonella nationalis	CB556	16/12/2009
Epiplocylys sp.	CB557	16/12/2009
Tintinnopsis campanulla	CB558	16/12/2009
Codonella nationalis	CB559	16/12/2009
Petalotricha ampulla	CB560	16/12/2009
Codonellopsis orthoceras	CB561	16/12/2009
Petalotricha ampulla	CB562	16/12/2009
Petalotricha ampulla	CB563	16/12/2009
Codonellopsis orthoceras	CB564	16/12/2009
Codonellopsis morchella	CB565	16/12/2009
Undella claparedei	CB566	27/01/2010
Undella claparedei	CB567	27/01/2010
Undella claparedei	CB568	27/01/2010
Undella claparedei	CB569	27/01/2010
Undella marsupialis	CB570	27/01/2010
Undella claparedei	CB571	27/01/2010

Undella claparedei	CB572	27/01/2010
Undella claparedei	CB573	27/01/2010
Undella claparedei	CB574	27/01/2010
Undella claparedei	CB575	27/01/2010
Undella claparedei	CB576	27/01/2010
Undella claparedei	CB577	27/01/2010
Undella claparedei	CB578	27/01/2010
Undella claparedei	CB579	27/01/2010
Undella claparedei	CB580	27/01/2010
Undella claparedei	CB581	27/01/2010
Undella claparedei	CB582	27/01/2010
Undella claparedei	CB583	27/01/2010
Codonella aspera	CB584	27/01/2010
Undella claparedei	CB585	27/01/2010
Undella claparedei	CB586	27/01/2010
Undella claparedei	CB587	27/01/2010
Undella claparedei	CB588	27/01/2010
Undella claparedei	CB589	27/01/2010
Undella claparedei	CB590	27/01/2010
Undella claparedei	CB591	27/01/2010
Undella claparedei	CB592	27/01/2010
Undella claparedei	CB593	27/01/2010
Undella claparedei	CB594	27/01/2010
Undella claparedei	CB595	27/01/2010
Undella claparedei	CB596	27/01/2010
Undella claparedei	CB597	27/01/2010
Undella claparedei	CB598	27/01/2010
Undella claparedei	CB599	27/01/2010
Undella claparedei	CB600	27/01/2010
Undella claparedei	CB601	27/01/2010
Undella claparedei	CB602	27/01/2010
Undella claparedei	CB603	27/01/2010
Undella claparedei	CB604	27/01/2010
Dictyocysta sp.	CB605	27/01/2010
Undella claparedei	CB606	27/01/2010
Undella claparedei	CB607	27/01/2010
Undella claparedei	CB608	27/01/2010
Undella claparedei	CB609	27/01/2010
Codonella nationalis	CB610	27/01/2010
Codonella nationalis	CB611	27/01/2010
Codonaria sp.	CB612	27/01/2010
Codonella nationalis	CB613	27/01/2010
Codonella nationalis	CB614	27/01/2010
Codonaria sp.	CB615	27/01/2010

Codonella nationalis	CB616	27/01/2010
Codonella nationalis	CB617	27/01/2010
Codonella nationalis	CB618	27/01/2010
Codonella nationalis	CB619	27/01/2010
Codonella nationalis	CB620	27/01/2010
Codonella nationalis	CB621	27/01/2010
Codonella nationalis	CB622	27/01/2010
Codonella nationalis	CB623	28/01/2010
Codonella nationalis	CB624	28/01/2010
Codonella nationalis	CB625	28/01/2010
Codonella nationalis	CB626	28/01/2010
Codonella nationalis	CB627	28/01/2010
Codonella nationalis	CB628	28/01/2010
Codonella nationalis	CB629	28/01/2010
Codonella nationalis	CB630	28/01/2010
Codonella nationalis	CB631	28/01/2010
Codonella nationalis	CB632	28/01/2010
Codonella nationalis	CB633	28/01/2010
Codonella nationalis	CB634	28/01/2010
Undella claparedei	CB635	24/02/2010
Undella claparedei	CB636	24/02/2010
Undella claparedei	CB637	24/02/2010
Undella claparedei	CB638	24/02/2010
Undella claparedei	CB639	24/02/2010
Undella claparedei	CB640	24/02/2010
Undella claparedei	CB641	24/02/2010
Undella claparedei	CB642	24/02/2010
Undella claparedei	CB643	24/02/2010
Undella claparedei	CB644	24/02/2010
Undella claparedei	CB645	24/02/2010
Undella claparedei	CB646	24/02/2010
Undella claparedei	CB647	24/02/2010
Undella claparedei	CB648	24/02/2010
Undella claparedei	CB649	24/02/2010
Undella claparedei	CB650	24/02/2010
Undella claparedei	CB651	24/02/2010
Undella claparedei	CB652	24/02/2010
Undella claparedei	CB653	24/02/2010
Undella claparedei	CB654	24/02/2010
Undella claparedei	CB655	24/02/2010
Undella claparedei	CB656	24/02/2010
Undella claparedei	CB657	24/02/2010
Undella claparedei	CB658	24/02/2010
Undella claparedei	CB659	24/02/2010

Annexes

Undella claparedei	CB660	24/02/2010
Undella claparedei	CB661	24/02/2010
Undella claparedei	CB662	24/02/2010
Undella claparedei	CB663	24/02/2010
Undella claparedei	CB664	24/02/2010
Undella claparedei	CB665	24/02/2010
Undella claparedei	CB666	24/02/2010
Undella claparedei	CB667	24/02/2010
Undella claparedei	CB668	24/02/2010
Undella claparedei	CB669	24/02/2010
Undella claparedei	CB670	24/02/2010
Undella claparedei	CB671	24/02/2010
Undella claparedei	CB672	24/02/2010
Undella claparedei	CB673	24/02/2010
Undella claparedei	CB674	24/02/2010
Undella claparedei	CB675	24/02/2010
Undella claparedei	CB676	24/02/2010
Undella claparedei	CB677	24/02/2010
Undella claparedei	CB678	24/02/2010
Undella claparedei	CB679	24/02/2010
Undella claparedei	CB680	24/02/2010
Undella claparedei	CB681	24/02/2010
Undella claparedei	CB682	24/02/2010
Undella claparedei	CB683	24/02/2010
Undella claparedei	CB684	24/02/2010
Undella claparedei	CB685	24/02/2010
Undella claparedei	CB686	24/02/2010
Undella claparedei	CB687	24/02/2010
Undella claparedei	CB688	24/02/2010
Undella claparedei	CB689	24/02/2010
Undella claparedei	CB690	24/02/2010
Undella claparedei	CB691	24/02/2010
Undella claparedei	CB692	24/02/2010
Undella claparedei	CB693	24/02/2010
Undella claparedei	CB694	24/02/2010
Undella claparedei	CB695	24/02/2010
Undella claparedei	CB696	24/02/2010
Undella claparedei	CB697	24/02/2010
Undella claparedei	CB698	24/02/2010
Undella claparedei	CB699	24/02/2010
Undella claparedei	CB700	24/02/2010
Undella claparedei	CB701	24/02/2010
Undella claparedei	CB702	26/05/2010
Undella claparedei	CB703	26/05/2010

Undella claparedei	CB704	26/05/2010
Undella claparedei	CB705	26/05/2010
Undella claparedei	CB706	26/05/2010
Undella claparedei	CB707	26/05/2010
Undella claparedei	CB708	26/05/2010
Undella claparedei	CB709	26/05/2010
Undella claparedei	CB710	26/05/2010
Undella claparedei	CB711	26/05/2010
Undella claparedei	CB712	26/05/2010
Undella claparedei	CB713	26/05/2010
Undella claparedei	CB714	26/05/2010
Undella claparedei	CB715	26/05/2010
Undella claparedei	CB716	26/05/2010
Undella claparedei	CB717	26/05/2010
Undella claparedei	CB718	26/05/2010
Undella claparedei	CB719	26/05/2010
Codonella aspera	CB720	27/05/2010
Codonella aspera	CB721	27/05/2010
Codonella aspera	CB722	27/05/2010
Codonella aspera	CB723	27/05/2010
Codonella aspera	CB724	27/05/2010
Codonella aspera	CB725	27/05/2010
Codonella aspera	CB726	27/05/2010
Codonella aspera	CB727	27/05/2010
Codonella aspera	CB728	27/05/2010
Codonella aspera	CB729	27/05/2010
Codonella aspera	CB730	27/05/2010
Codonella aspera	CB731	27/05/2010
Codonella aspera	CB732	27/05/2010
Codonella aspera	CB733	27/05/2010
Codonella aspera	CB734	27/05/2010
Codonella aspera	CB735	27/05/2010
Codonella aspera	CB736	27/05/2010
Codonella aspera	CB737	27/05/2010
Undella claparedei	CB738	29/09/2010
Undella claparedei	CB739	29/09/2010
Undella claparedei	CB740	29/09/2010
Undella claparedei	CB741	29/09/2010
Undella claparedei	CB742	29/09/2010
Undella claparedei	CB743	29/09/2010
Undella claparedei	CB744	29/09/2010
Undella claparedei	CB745	29/09/2010
Undella claparedei	CB746	29/09/2010
Undella claparedei	CB747	29/09/2010

Annexes

Undella claparedei	CB748	29/09/2010
Undella claparedei	CB749	29/09/2010
Undella claparedei	CB750	29/09/2010
Undella claparedei	CB751	29/09/2010
Undella claparedei	CB752	29/09/2010
Undella claparedei	CB753	29/09/2010
Undella claparedei	CB754	29/09/2010
Undella claparedei	CB755	29/09/2010
Undella claparedei	CB756	29/09/2010
Undella claparedei	CB757	29/09/2010
Undella marsupialis	CB758	29/09/2010
Rhabdonella spiralis	CB759	29/09/2010
Undella claparedei	CB760	29/09/2010
Undella claparedei	CB761	29/09/2010
Undella claparedei	CB762	29/09/2010
Undella hyalina	CB763	29/09/2010
Codonellopsis orthoceras	CB764	29/09/2010
Undella claparedei	CB765	29/09/2010
Undella claparedei	CB766	29/09/2010
Undella claparedei	CB767	29/09/2010
Dictyocysta mitra	CB768	29/09/2010
Undella claparedei	CB769	29/09/2010
Salpingella acuminata	CB770	29/09/2010
Salpingella acuminata	CB771	29/09/2010
Dadayella ganymedes	CB772	29/09/2010
Salpingella acuminata	CB773	29/09/2010
Undella claparedei	CB774	29/09/2010
Parundella sp.	CB775	29/09/2010
Xystonellopsis sp.	CB776	29/09/2010
dadayella ganymedes	CB777	29/09/2010
Eutintinnus sp.	CB778	29/09/2010
Xystonella sp.	CB779	29/09/2010
Undella claparedei	CB780	29/09/2010
Eutintinnus sp. apertus	CB781	01/10/2010
Eutintinnus sp. apertus	CB782	01/10/2010
Eutintinnus sp. corta	CB783	01/10/2010
codonella corta	CB784	01/10/2010
Eutintinnus sp. apertus	CB785	01/10/2010
Undella claparedei	CB786	01/12/2010
Undella claparedei	CB787	01/12/2010
Dictyocysta mitra	CB788	01/12/2010
Undella claparedei	CB789	01/12/2010
Undella claparedei	CB790	01/12/2010
Undella claparedei	CB791	01/12/2010

Annexes

Undella claparedei	CB792	01/12/2010
Undella claparedei	CB793	01/12/2010
Epipocyclis sp.	CB794	01/12/2010
Undella claparedei	CB795	01/12/2010
Undella claparedei	CB796	01/12/2010
Undella claparedei	CB797	01/12/2010
Undella claparedei	CB798	01/12/2010
Undella claparedei	CB799	01/12/2010
Undella claparedei	CB800	01/12/2010
Undella claparedei	CB801	01/12/2010
Undella claparedei	CB802	01/12/2010
Undella claparedei	CB803	01/12/2010
Undella claparedei	CB804	01/12/2010
Undella claparedei	CB805	01/12/2010
Undella hyalina	CB806	01/12/2010
Rhabdonella spiralis	CB807	01/12/2010
Undella claparedei	CB808	01/12/2010
Undella claparedei	CB809	01/12/2010
Undella hyalina	CB810	01/12/2010
Undella claparedei	CB811	01/12/2010
Undella claparedei	CB812	01/12/2010
Undella claparedei	CB813	01/12/2010
Undella claparedei	CB814	01/12/2010
Undella claparedei	CB815	01/12/2010
Undella claparedei	CB816	01/12/2010
Undella claparedei	CB817	01/12/2010
Undella claparedei	CB818	01/12/2010
Undella claparedei	CB819	01/12/2010
Undella claparedei	CB820	01/12/2010
Undella claparedei	CB821	01/12/2010
Undella claparedei	CB822	01/12/2010
Undella claparedei	CB823	01/12/2010
Undella hyalina	CB824	01/12/2010
Undella claparedei	CB825	01/12/2010
Undella claparedei	CB826	01/12/2010
Undella hyalina	CB827	01/12/2010
Petalotricha ampulla	CB828	01/12/2010
Undella claparedei	CB829	01/12/2010
Petalotricha ampulla	CB830	01/12/2010
Undella claparedei	CB831	01/12/2010
Undella claparedei	CB832	01/12/2010
Undella hyalina	CB833	01/12/2010
Undella marsupialis	CB835	02/12/2010
Cyttarocylis sp.	CB836	02/12/2010

Annexes

Petalotricha ampulla	CB837	02/12/2010
Xystonellopsis	CB838	02/12/2010
Undella claparedei	CB839	02/12/2010
Undella claparedei	CB840	02/12/2010
Salpingella acuminata	CB841	02/12/2010
Undella claparedei	CB842	02/12/2010
Undella claparedei	CB843	02/12/2010
Undella claparedei	CB844	02/12/2010
Undella claparedei	CB845	02/12/2010
Undella hyalina	CB846	02/12/2010
Undella claparedei	CB847	02/12/2010
Undella claparedei	CB848	02/12/2010
Undella claparedei	CB849	02/12/2010
Dictyocysta lepidans	CB850	02/12/2010
Undella claparedei	CB851	02/12/2010
Undella claparedei	CB852	02/12/2010
Undella claparedei	CB853	02/12/2010
Undella hyalina	CB854	02/12/2010
Undella claparedei	CB855	02/12/2010
Undella claparedei	CB856	02/12/2010
Undella claparedei	CB857	02/12/2010
Undella claparedei	CB858	02/12/2010
Undella claparedei	CB859	02/12/2010
Undella claparedei	CB860	02/12/2010
Undella claparedei	CB861	02/12/2010
Undella claparedei	CB862	02/12/2010
Undella claparedei	CB863	02/12/2010
Undella claparedei	CB864	02/12/2010
Undella claparedei	CB865	02/12/2010
Xystonellopsis sp.	CB866	02/12/2010
Codonella amphorella	CB867	02/12/2010
Salpingella acuminata	CB868	02/12/2010
Tintinnidium sp.	CB869	02/12/2010
Dictyocysta sp.	CB870	02/12/2010
Codonella nationalis	CB871	02/12/2010
Codonella aspera	CB872	02/12/2010
Cyttarocylys sp.	CB873	02/12/2010
Undella claparedei	CB874	02/12/2010
Stenosemella nivalis	CB875	02/12/2010
Stenosemella nivalis	CB876	02/12/2010
Cyttarocylys sp.	CB877	02/12/2010

Annexe 2. Liste des échantillons d'eau marine collectés à la bouteille Niskin pendant le suivi temporel dans la rade de Villefranche-sur-Mer au point B+, avec les analyses moléculaires de diversité employées pour chaque échantillon.

Nom de l'échantillon	Profondeur (mètres)	Date	Amplification, clonage et séquençage Sanger	DGGE	Amplification et pyroséquençage
VilleFr-1	5	18-févr-09	X	X	
VilleFr-2	25	18-févr-09	X	X	
VilleFr-3	50	18-févr-09	X	X	
VilleFr-4	100	18-févr-09	X	X	
VilleFr-5	160	18-févr-09	X	X	
VilleFr-6	5	18-mars-09		X	
VilleFr-7	25	18-mars-09		X	
VilleFr-8	50	18-mars-09	X	X	
VilleFr-9	100	18-mars-09		X	
VilleFr-10	160	18-mars-09		X	
VilleFr-11	5	17-avr-09		X	
VilleFr-12	25	17-avr-09		X	
VilleFr-13	50	17-avr-09		X	
VilleFr-14	100	17-avr-09		X	
VilleFr-15	160	17-avr-09		X	
VilleFr-16	5	27-mai-09		X	
VilleFr-17	25	27-mai-09		X	
VilleFr-18	50	27-mai-09	X	X	
VilleFr-19	100	27-mai-09		X	
VilleFr-20	160	27-mai-09		X	
VilleFr-21	5	25-juin-09		X	
VilleFr-22	25	25-juin-09	X	X	
VilleFr-23	50	25-juin-09	X	X	
VilleFr-24	100	25-juin-09	X	X	
VilleFr-25	160	25-juin-09	X	X	
VilleFr-26	5	24-juil-09		X	
VilleFr-27	25	24-juil-09		X	
VilleFr-28	50	24-juil-09	X	X	
VilleFr-29	100	24-juil-09		X	
VilleFr-30	160	24-juil-09		X	
VilleFr-31	5	24-sept-09	X	X	
VilleFr-32	25	24-sept-09	X	X	
VilleFr-33	50	24-sept-09	X	X	
VilleFr-34	100	24-sept-09	X	X	
VilleFr-35	160	24-sept-09		X	
VilleFr-36	5	22-oct-09		X	
VilleFr-37	25	22-oct-09		X	

Annexes

VilleFr-38	50	22-oct-09		x	
VilleFr-39	100	22-oct-09		x	
VilleFr-40	160	22-oct-09		x	
VilleFr-41	5	18-nov-09		x	
VilleFr-42	25	18-nov-09		x	
VilleFr-43	50	18-nov-09	x	x	x
VilleFr-44	100	18-nov-09	x	x	
VilleFr-45	160	18-nov-09		x	
VilleFr-46	5	16-déc-09		x	
VilleFr-47	25	16-déc-09		x	
VilleFr-48	50	16-déc-09	x	x	
VilleFr-49	100	16-déc-09		x	
VilleFr-50	160	16-déc-09		x	
VilleFr-51	5	27-janv-10		x	
VilleFr-52	25	27-janv-10		x	
VilleFr-53	50	27-janv-10	x	x	
VilleFr-54	100	27-janv-10		x	
VilleFr-55	160	27-janv-10		x	
VilleFr-56	5	24-févr-10		x	
VilleFr-57	25	24-févr-10		x	
VilleFr-58	50	24-févr-10	x	x	
VilleFr-59	100	24-févr-10		x	
VilleFr-60	160	24-févr-10		x	
VilleFr-61	5	26-mai-10			
VilleFr-62	25	26-mai-10			
VilleFr-63	50	26-mai-10	x		
VilleFr-64	100	26-mai-10			
VilleFr-65	160	26-mai-10			
VilleFr-66	5	26-sept-10			
VilleFr-67	25	26-sept-10			
VilleFr-68	50	26-sept-10	x		
VilleFr-69	100	26-sept-10			
VilleFr-70	160	26-sept-10			
VilleFr-71	5	01-déc-10			
VilleFr-72	25	01-déc-10			
VilleFr-73	50	01-déc-10	x		
VilleFr-74	100	01-déc-10			
VilleFr-75	160	01-déc-10			

Annexe 3. Article réalisé durant mon doctorat sur la diversité des eucaryotes dans des échantillons collectés à la fin de la nuit polaire Arctique sur un site proche du pôle Nord. L'amplification de l'ADNr 18S des eucaryotes m'a donné l'occasion de décrire la diversité le long d'un gradient vertical partant de la couche épaisse de neige qui recouvre la banquise, en passant par la banquise pluri-annuelle à trois profondeurs différentes, jusqu'à des échantillons de la colonne d'eau sous la banquise.



Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the North Pole at the end of the polar night

Charles Bachy¹, Purificación López-García¹, Alexander Vereshchaka² and David Moreira^{1*}

¹ Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, Orsay, France

² Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

Edited by:

Jennifer F. Biddle, University of Delaware, USA

Reviewed by:

Peter D. Countway, Bigelow Laboratory for Ocean Sciences, USA

Daniel Vaultot, Station Biologique de Roscoff, France

*Correspondence:

David Moreira, Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, 91405 Orsay Cedex, France.
e-mail: david.moreira@u-psud.fr

Our knowledge about the microorganisms living in the high Arctic Ocean is still rudimentary compared to other oceans mostly because of logistical challenges imposed by its inhospitable climate and the presence of a multi-year ice cap. We have used 18S rRNA gene libraries to study the diversity of microbial eukaryotes in the upper part of the water column (0–170 m depth), the sea ice (0–1.5 m depth) and the overlying snow from samples collected in the vicinity of the North Pole (N88°35', E015°59') at the very end of the long polar night. We detected very diverse eukaryotes belonging to Alveolata, Fungi, Amoebozoa, Viridiplantae, Metazoa, Rhizaria, Heterokonta, and Telonemia. Different alveolates (dinoflagellates and Marine Alveolate Groups I and II species) were the most abundant and diverse in gene libraries from water and sea ice, representing 80% of the total number of clones and operational taxonomic units. Only contaminants and/or species from continental ecosystems were detected in snow, suggesting wind- and animal- or human-mediated cosmopolitan dispersal of some taxa. By contrast, sea ice and seawater samples harbored a larger and more similar inter-sample protist diversity as compared with snow. The North Pole was found to harbor distinctive eukaryotic communities along the vertical gradient with an unparalleled diversity of core dinoflagellates, largely dominant in libraries from the water column, as compared to other oceanic locations. In contrast, phototrophic organisms typical of Arctic sea ice and plankton, such as diatoms and prasinophytes, were very rare in our samples. This was most likely due to a decrease of their populations after several months of polar night darkness and to the presence of rich populations of diverse grazers. Whereas strict phototrophs were scarce, we identified a variety of likely mixotrophic taxa, which supports the idea that mixotrophy may be important for the survival of diverse protists through the long polar night.

Keywords: North Pole, arctic, plankton, protist diversity, sea ice, dinoflagellates, alveolates

INTRODUCTION

One of the most characteristic features of the Arctic Ocean is its permanent sea ice cover, which is a present source of concern because of its steady decline linked to global warming (Lemke et al., 2007). The logistical constraints to sample the northernmost sites on Earth due to harsh environmental conditions are still an obstacle to biodiversity and ecological studies in such high latitudes, especially during maximum sea ice cover periods (Comiso, 2003). Despite these limitations, it is known that sea ice offers an exclusive habitat for a diverse microbial community in polar environments (see review in Mock and Thomas, 2005). A special attention has been paid to dominant species, mostly unicellular algae and heterotrophic bacteria. A variety of protist species have also been reported in Arctic sea ice and seawater by traditional morphology-based methods (e.g., Horner, 1985; Lizotte, 2003; Olli et al., 2007). Diatoms, which significantly contribute to the primary production in the Arctic Ocean (see review by Arrigo, 2003), are by far the best studied protist group in these waters (Palmisano and Garrison, 1993; Ikävalko and Thomsen, 1997; Melnikov, 1997). However, a wide diversity of other eukaryotes is also present in Arctic sea ice and underlying waters. Marine microorganisms may be trapped during ice formation and remain inactive

and frozen, but also active microorganisms live in brine-channel systems within the ice, being subjected to strong physical and chemical constraints (e.g., temperature, salinity, light, nutrients) which vary both spatially and temporally (Eicken, 2003). These parameters reach extreme values during winter, not only within the ice but also in the ice-covered Arctic Ocean that, with an average depth of approximately 1000 m, is an environment characterized by very limited light and extremely low temperatures. The sea ice is typically 2 m thick and the snow cover up to 28 cm thick (Richter-Menge et al., 2006). However, the information about microbial communities associated with the multi-year ice cover is very limited (Thomas et al., 1995; Perovich et al., 1999; Werner et al., 2007), especially during winter and in central areas of the Arctic Ocean, where the water column depths are the highest (4179 m at the North Pole; maximum depth 5450 m at the European basin).

During the last decade, environmental studies of eukaryotic diversity based on polymerase chain reaction (PCR) amplification, cloning, and sequencing of the 18S rRNA gene have revealed a wide diversity of protists in a variety of poorly explored habitats (López-García et al., 2001; Moon-Van Der Staay et al., 2001; Moreira and López-García, 2002; Richards and Bass, 2005). So far,

only a few molecular studies of protist diversity have been carried out in very high-latitude oceanic regions. Gast et al. (2004) reported the first molecular survey of the microbial eukaryotes from sea ice environments in the Antarctic Ross Sea, showing a large genetic diversity and a certain degree of habitat specificity. In the high Arctic, molecular methods were used to study protist diversity in nearshore fast ice and waters (Lovejoy et al., 2006; Terrado et al., 2009; Eddie et al., 2010), fjords (Luo et al., 2009; Piquet et al., 2010), sediments (Stoeck et al., 2007; Tian et al., 2009) and, very recently, snow (Harding et al., 2011). However, the diversity and vertical distribution of microbial eukaryotic communities in Arctic offshore environments remains to be studied.

Samples collected during the Barneo-2008 expedition (March–April 2008) offered us the opportunity to explore the highest latitudes on Earth at the very end of the polar night. Here, we report the diversity and vertical distribution of microbial eukaryotes of

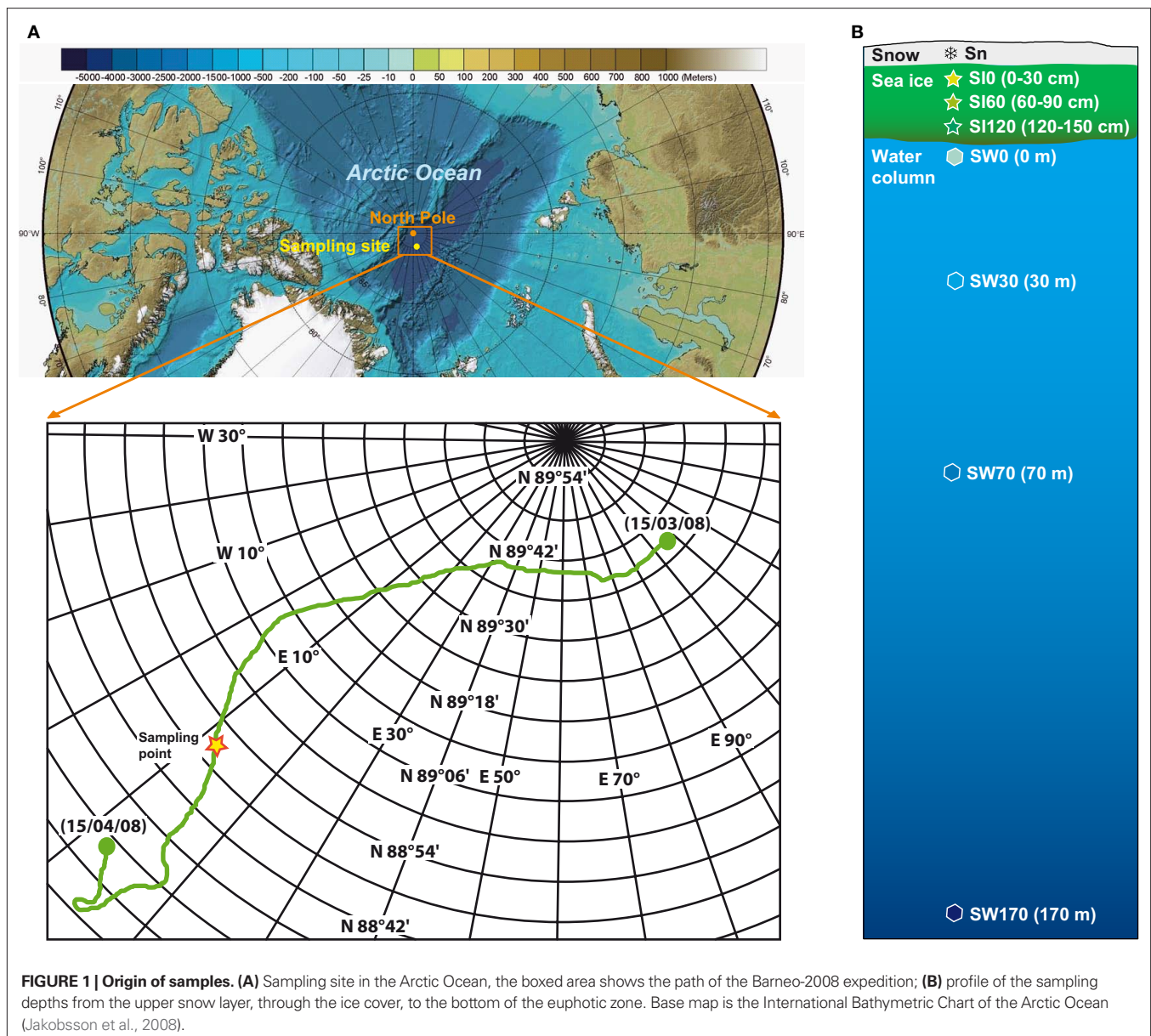
sizes comprised between 0.2 and 200 μm in the multi-year sea ice, the underlying waters and the snow covering the sea ice from a sampling site located at close proximity to the geographic North Pole.

MATERIALS AND METHODS

SAMPLE COLLECTION

Samples were collected during the Barneo-2008 expedition¹, a Russian initiative organized as part of the 2007–2008 International Polar Year. This expedition was based on the installation of a scientific camp on a large ice floe of several kilometers of diameter drifting in the North Pole region. Scientists occupied the camp for 1 month (from March 15th to April 15th) while the ice floe freely drifted from the original location, N89°36', E118°49', to one more to the South, N88°28', E12°47' (Figure 1A), Because of logistic

¹<http://barneo.ru/2008e.htm>



reasons and to wait until the very end of the polar night, samples were taken at an intermediate location (N88°35', E015°59') between April 4th and 11th. These samples corresponded to different depths along a vertical profile (**Figure 1B**), from the snow cover (1 sample of 4 l), through the 1.5-m thick multi-year sea ice (three samples of 4 l each, at depths 0–30, 60–90, and 120–150 cm) to the epipelagic zone of the underlying water column (four samples at 0, 30, 70, and 170 m depth). Snow was collected with a sterile bucket, sea ice with a sterile ice driller (11 cm of diameter), and seawater with a 5-l Niskin bottle. Seawater was collected directly under the sea ice through the hole opened during the sea ice sampling. Snow and ice were melted at 4°C under sterile conditions. After passing water samples through a 200- μ m mesh, samples of approximately 4 (snow and ice) and 1 l (seawater) were filtered through a GTTP Millipore 0.22 μ m-pore-diameter filter. Filters containing the cells were fixed in absolute ethanol and conserved at –20°C following a protocol already tested for plankton samples (López-García et al., 2003, 2007). DNA was purified from filters cut in small pieces using the Ultraclean MoBio Soil DNA kit (MoBio, Solana Beach, CA, USA). Nucleic acids were resuspended in 10 mM Tris–HCl, pH 8.

PCR AMPLIFICATION, CLONING, AND SEQUENCING

Near full-length 18S rDNA fragments were amplified by PCR using the eukaryotic-specific primers 82F (GAA ACT GCG AAT GGC TC) and 1498R (CAC CTA CGG AAA CCT TGT TA; López-García et al., 2003). PCR was carried out under the following conditions: 35 cycles (denaturation at 94°C for 15 s, annealing at 50°C for 30 s, extension at 72°C for 2 min) preceded by 3 min denaturation at 94°C, and followed by 10 min extension at 72°C. 18S rDNA clone libraries were constructed using the Topo TA cloning system (Invitrogen) following the instructions provided by the manufacturer. Positive inserts of expected size were selected from each library and sequenced with the reverse primer. We obtained a total of 1000 high-quality partial sequences (>700 bp) which served for a preliminary phylogenetic analysis and identification of operational taxonomic units (OTUs). For each library, we completely sequenced at least one clone per OTU (defined here as clusters of sequences having >98% identity) to obtain complete sequences representative of the whole taxonomic diversity found. 175 complete sequences were thus retained after exclusion of 14 chimeras detected with KeyDNATools² (Guillou et al., 2008). Sequences were submitted to GenBank (accession numbers HQ438100–HQ438190 and JF826314–JF826397)

OTU IDENTIFICATION AND BIODIVERSITY ESTIMATES

High-quality partial 18S rDNA sequences were aligned using MAFFT (Katoh et al., 2002) and a distance matrix was generated in Phylip format. The resulting matrix was used as input for MOTHRU³ (Schloss et al., 2009) in order to group sequences in OTUs at different identity levels and to compare communities using heatmaps.

PHYLOGENETIC ANALYSES

The 175 representative clones that were completely sequenced, together with their most similar sequences (both known species and/or environmental clones) identified by BLAST (Altschul

et al., 1997) in GenBank⁴, were aligned using MAFFT (Katoh et al., 2002). The multiple sequence alignment was then manually edited using the program ED from the MUST package (Philippe, 1993). A neighbor-joining (NJ) tree was constructed for all the taxa in order to choose representative subsets of sequences for further phylogenetic analyses. The first three datasets corresponded to a variety of ciliate, Marine Alveolate Group I (MAG I) and II (Syndiniales) and dinoflagellate sequences, which were found to be the most abundant and diverse. The two other datasets corresponded to metazoan, fungal, amoebozoan, green plant, rhizarian, and heterokont sequences, spanning the rest of the eukaryotic diversity found. These datasets were analyzed by maximum likelihood (ML) using TREEFINDER (Jobb et al., 2004) applying a general time reversible (GTR) model of sequence evolution with a 4-rate categories Gamma law and invariant sites to accommodate for among-site rate variation. ML bootstrap proportions were inferred using 1000 replicates. Bayesian inference was also conducted using MrBayes (Huelsenbeck and Ronquist, 2001), starting with a random tree, ran for one million generations in four chains and excluding the first 3500 generations as “burn in.” The ML and Bayesian analyses produced congruent topologies.

RESULTS

MICROBIAL EUKARYOTIC DIVERSITY NEAR THE NORTH POLE

Our clone libraries were clearly dominated by Alveolata, especially by the core dinoflagellates (Dinokaryota) and their close relatives Marine Alveolate Groups I (MAG I) and II (Syndiniales; **Figure 2**). Sequences affiliating to core dinoflagellates were the most abundant in all libraries from the water column, especially in our deepest seawater sample (W170, 170 m depth), where they represented up to 80% of clones. They were also dominant in sea ice, except for the sample SI60 where MAG I and II sequences were equally abundant and represented almost 40% of clones in this sample. In fact, MAG I and II were after dinoflagellates the most represented lineages in almost all the seawater and sea ice samples, though with varying proportions (**Figure 2**). Ciliate, heterokonts (stramenopiles), green algal, rhizarian, and fungal clones were detected in weak proportions in seawater and sea ice samples (from 1 to 7%). We noticed the presence of sequences affiliated to animals in seawater and snow, reaching remarkable proportions in the surface seawater sample SW0 (29% of clones of this library). A few sequences were identified as Amoebozoa in the deep seawater sample SW170 and as Telonemia in the deepest sea ice sample SI120. The composition of high-rank taxa in the snow sample was radically different from other samples with a large majority of fungal (72%) and green plant (20%) clones, strongly suggesting contamination (see below). The clone library from the deepest sea ice sample was the richest in terms of high-rank phylogenetic taxa diversity, whereas that from the snow sample was the poorest.

Using the criterion of >98% sequence identity for the definition of OTUs, we observed a specific richness of, collectively, 98 OTUs in the water column, 62 OTUs in the sea ice, and only 4 OTUs in the snow. More precisely, the observed specific richness was similar between the pelagic samples and the bottom sea ice sample and slightly lower in the upper and middle sea ice layers (**Table 1**). In addition to these diversity estimations, we carried out a statistical

²<http://www.keydnatools.com/>

³<http://www.mothur.org/>

⁴<http://www.ncbi.nlm.nih.gov/>

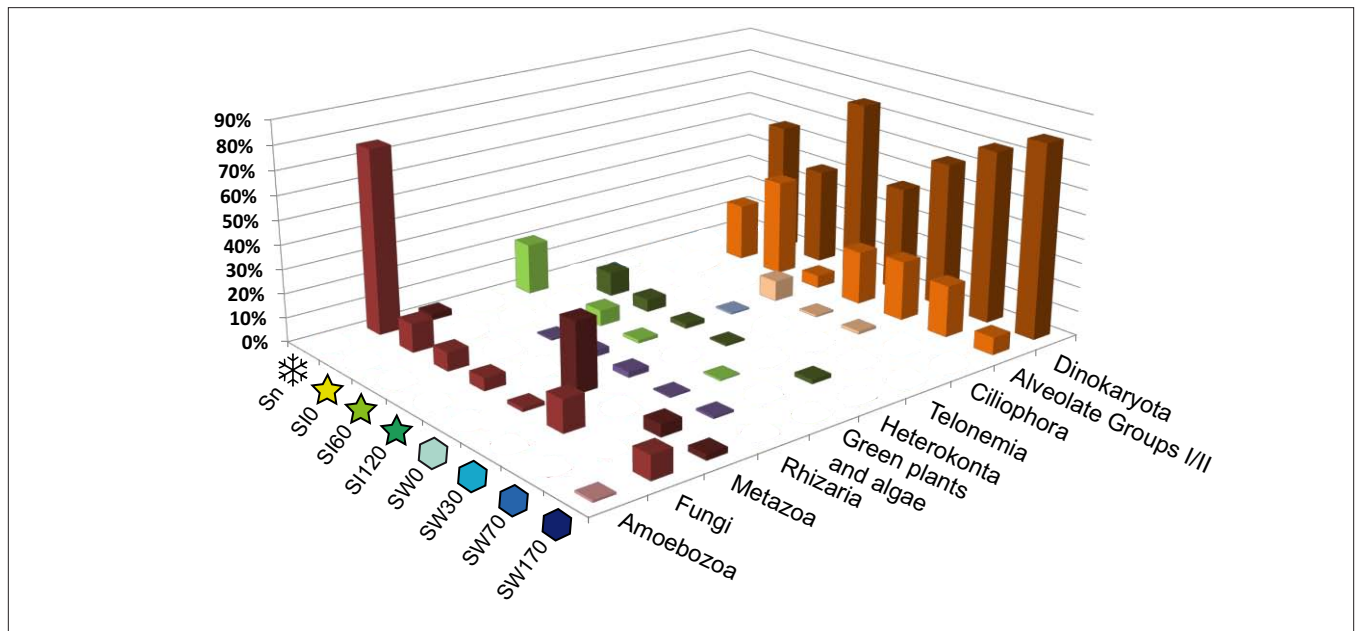


FIGURE 2 | Relative 18S rDNA abundance of major eukaryotic groups in gene libraries from samples near the North Pole. For sample names, see Figure 1.

Table 1 | Number of OTUs and clones in the North Pole clone libraries.

	OTUs 98%	OTUs 95%	Number of clones
Snow	4	4	38
SI0	19	14	130
SI60	25	21	155
SI120	33	24	146
Sea ice*	62	45	431
SW0	34	24	133
SW30	44	31	154
SW70	30	17	118
SW170	29	19	125
Seawater*	98	60	530

*Total values for all sea ice and water samples combined.

comparison of the overall eukaryotic diversity among samples using two different levels of sequence identity: 98% identity as a proxy for species boundaries and at 95% identity as a proxy for genera. At 98% sequence identity, the seawater samples SW30 and W170 as well as the sea ice samples SI0 and SI60 were the closest among them, differing significantly from the rest (Figure 3). At 95% sequence identity the differences between seawater samples and the deepest sea ice sample (SI120) were negligible (Figure 3), suggesting that differences in protist community composition between these samples concentrate at the species, but not genus, level. Likewise, the upper and middle sea ice samples (SI0 and SI60) appeared to be similar. Therefore, our analysis supports the partition of our samples into three groups in terms of their diversity of protist species and genera: (i) snow, (ii) the upper and middle sea ice layer, and (iii) the deepest sea ice layer and all the seawater samples (with the deepest sea ice layer being intermediate between the seawater and the upper ice layers, as expected).

DIVERSITY OF ALVEOLATA

As mentioned above, the alveolates were by far the most abundant and diverse group in both plankton and sea ice 18S rDNA libraries. A few phylotypes related to marine ciliates were found at the interface of sea ice and marine waters (Figure 4) but, as has been observed in many other marine plankton samples, alveolate phylotypes belonging to dinoflagellates and the parasitic MAG I and Syndiniales (MAG II) were dominant in libraries, representing more than 78% of the 1000 partially sequenced clones. Seven seawater OTUs grouped with Syndiniales representatives, some of which had already been detected in other marine locations (López-García et al., 2007; Not et al., 2007; Guillou et al., 2008), and two clones (SI120-63 and SW30-39) branched within a well-supported cluster grouping *Amoebophrya* species (Figure 5). The Syndiniales are a dinoflagellate order composed of small marine parasites that appear to be highly diverse and abundant in environmental studies (López-García et al., 2001; Moon-Van Der Staay et al., 2001; Moreira and López-García, 2003). They can infect a variety of hosts, from marine animals to protists, though many of them are specific parasites of other dinoflagellates (e.g., *Amoebophrya* spp., Chambouvet et al., 2008). The presence of a large diversity of core dinoflagellates in our samples (see below) could explain, at least partially, the abundance and diversity of Syndiniales observed in the same samples. The MAG I, which have also been retrieved abundantly in various marine environments (see López-García et al., 2001; Moon-Van Der Staay et al., 2001; Massana et al., 2004; Romari and Vaultot, 2004; Dolven et al., 2007; Harada et al., 2007; Not et al., 2007), encompassed a larger genetic diversity than Syndiniales in our samples, being identified in all samples from 170 m depth to top of sea ice (Figure 5). We detected two large MAG I clades, one encompassing *Duboscquella* and *Ichthyodinium* sequences, and the other without any described species so far, but only environmental sequences coming from both coastal and oceanic ecosystems (Lin et al., 2006; Not et al., 2007; Behnke et al., 2010).

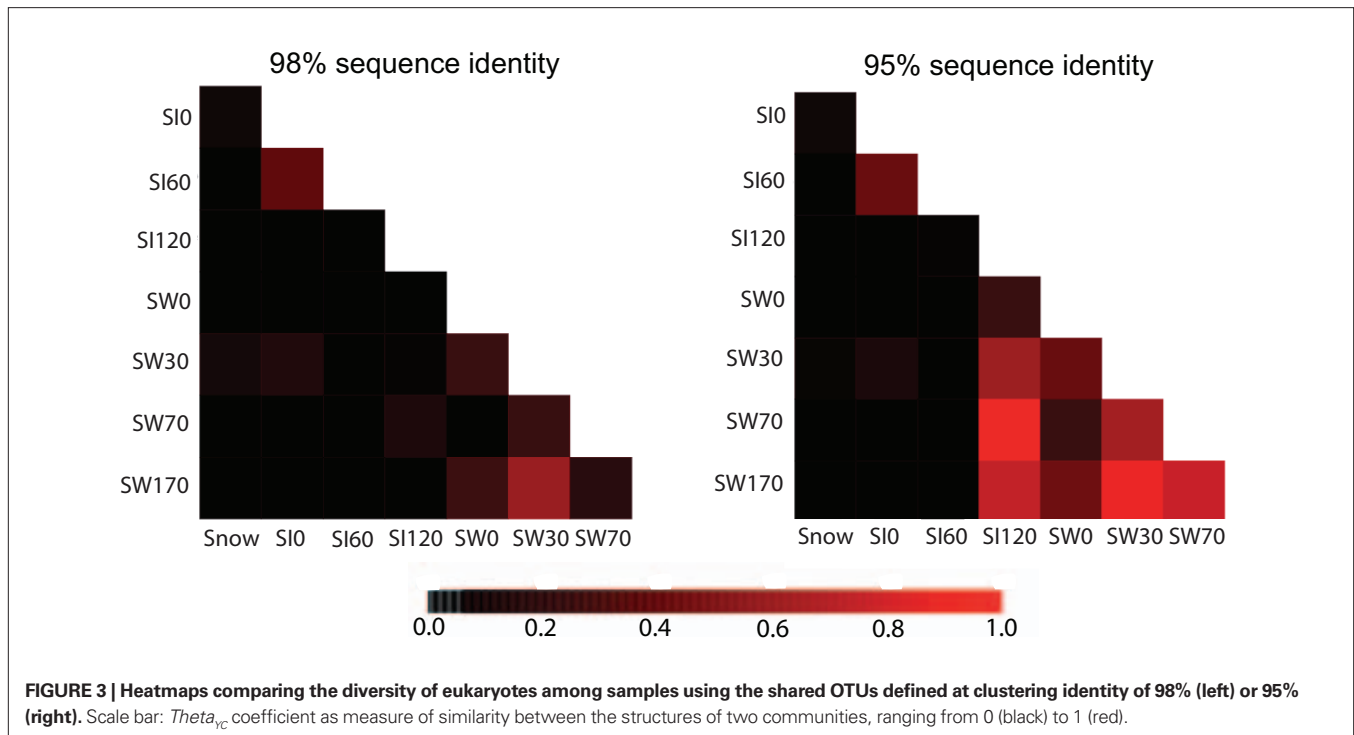


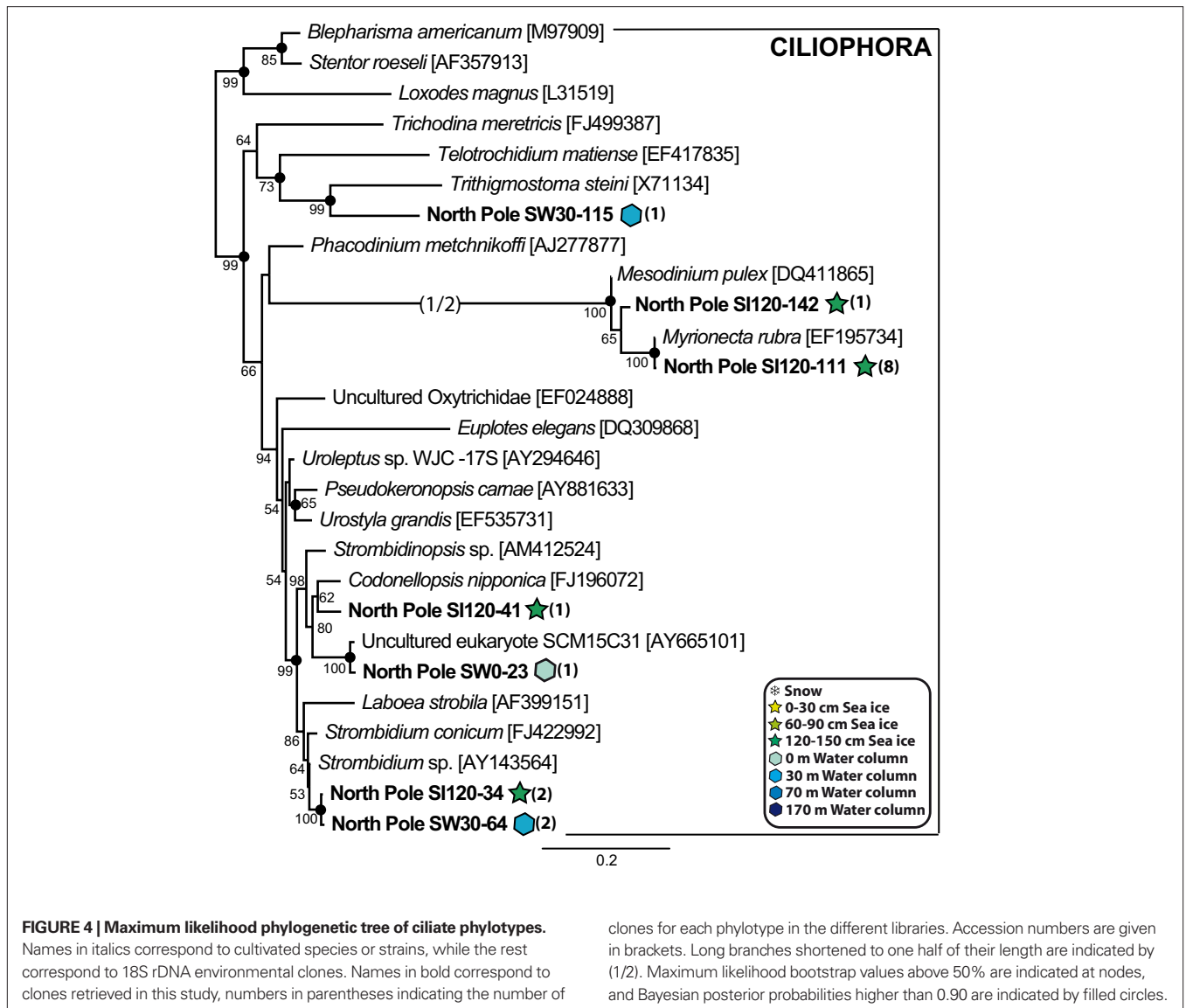
FIGURE 3 | Heatmaps comparing the diversity of eukaryotes among samples using the shared OTUs defined at clustering identity of 98% (left) or 95% (right). Scale bar: $\Theta_{v,c}$ coefficient as measure of similarity between the structures of two communities, ranging from 0 (black) to 1 (red).

The core dinoflagellate phylotypes showed a very remarkable diversity (Figure 6), in agreement with several studies based on traditional methods showing the dominance of small flagellates, especially dinoflagellates, in the central Arctic Ocean during summer (Booth and Horner, 1997; Gradinger, 1999; Olli et al., 2007). Therefore, it seems that dinoflagellates are also diverse and abundant at the end of the polar night, in both the sea ice and seawater. Several of our dinoflagellate phylotypes grouped with cultured species usually observed in cold water environments, such as *Protoperdinium* sp. (phylotypes detected in the seawater samples), *Scrippsiella hangoei*, *Polarella glacialis* and *Gymnodinium* spp. (phylotypes detected in both the seawater and sea ice samples), and *Woloszynskia halophila* (phylotypes detected in the SI120 sea ice sample). Other phylotypes, such as SW30-129, SW70-41, SW30-6, SI120-47, SW70-35, or SW0-135 had other environmental sequences as very close relatives, some of which came from cold waters, e.g., SIF 1E11 and W159F8, coming from a Norwegian fjord and the Ross Sea, respectively (Gast et al., 2006; Behnke et al., 2010). Interestingly, an OTU (represented by SI0-102) detected in seawater and sea ice formed a divergent environmental clade with no close relatives in GenBank. A similar case was shown by several seawater sequences (represented by SW170-19) 99% identical to the Sargasso Sea environmental sequence SCM27C4, which formed a fast evolving cluster characterized by a very long basal branch. Intriguingly, five phylotypes (SW30-106, SI60-101, SI60-19, SI60-129, and SW0-75) formed a group well nested within dinokaryotic dinoflagellates without any closely related cultured species or environmental sequence. In fact, our results indicated that a large part of the core dinoflagellate phylotypes found in all the water and sea ice samples was related to undescribed species. Nevertheless, the two most abundant phylotypes, SW70-6 and

SW70-12, were closely related to the well known heterotrophic species *Gyrodinium spirale* and *G. rubrum*, respectively. Both phylotypes were frequent at all depths in the water column, specially the *G. rubrum*-like SW70-12, with a total of 88 occurrences (Figure 6), namely ~9% of all our sequences, which was a remarkable value for a single species.

OTHER EUKARYOTIC TAXA

In contrast with the large diversity of alveolates, we only retrieved a limited variety of other eukaryotic groups in our North Pole samples, including several phylotypes belonging to the Metazoa, the Fungi, and the Amoebozoa. Within the Fungi, Basidiomycota, found in the snow, all the sea ice horizons and the water column, were the most diverse, whereas Ascomycota were only detected in both sea ice and seawater (Figure 7). Most of the ascomycete phylotypes were very closely related to *Penicillium chrysogenum*, a common airborne fungus (Wu et al., 2003). The marine fungi are well known to be heterotrophs that play an essential role as decomposers of organic matter (Kohlmeyer and Kohlmeyer, 1979; Newell, 1996), and their diversity in our samples suggested a similar potential role in high-latitude environments. We identified three animal phylotypes, one copepod and one annelid in the water samples. We also detected a human sequence in the snow sample, which corresponded obviously to a contaminant likely introduced during sampling and that we kept in our analyses to illustrate the fact that microbial diversity in the snow was very low, and possibly entirely composed of inactive organisms or dispersal forms, namely exogenous contaminant sequences. This was clearly the case for the snow phylotype Sn-26, closely related to the temperate-climate terrestrial plant *Pilea cadierei*, and also for the sea ice phylotype SI60-38, very similar to *Pinus* sequences (Figure 8). These two phylotypes most



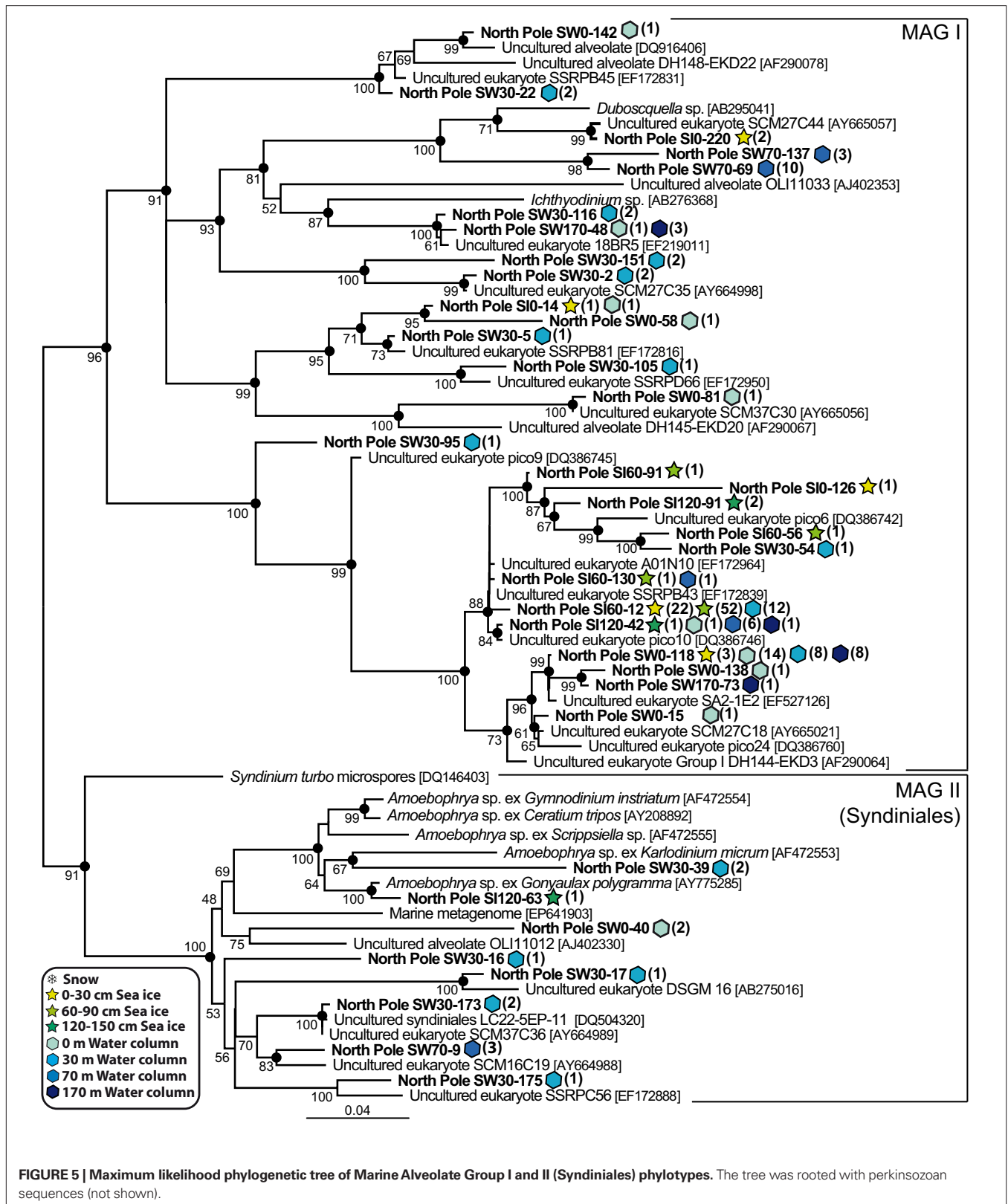
likely corresponded to pollen contaminants, pollen grains being well known to be able to traverse long distances with the help of wind currents (e.g., Campbell et al., 1999).

The diversity of heterokonts (or stramenopiles) was relatively large, with seven North Pole phylotypes (Figure 8). Only three of them corresponded to presumable photosynthetic species: SW70-32 and SW0-77, which branched among diatoms very close to *Actinocyclus curvatus* and *Thalassiosira antarctica*, two typical sea ice diatom species (Hasle and Heimdal, 1998), and SI120-23, which was very similar to an environmental clone from an anoxic Norwegian fjord, NIF-1D10, branching with the dictyochophyte *Dictyocha speculum*. The other four heterokont phylotypes were related to likely heterotrophic species: SI60-86 clustered within the environmental clade MAST-12 characterized by Kolodziej and Stoeck (2007), SI120-114 with *Pirsonia punctigera*, a parasitoid nanoflagellate that infects planktonic diatoms (Kühn et al., 2004), and SI0-36 and SI60-27 with the colorless chrysophytes *Spumella* sp. and *Ochromonas* sp. The rest of phylotypes showed a similar trend,

a predominance of heterotrophic organisms; the only apparent exception being the phylotype SI120-29 affiliated to the prasinophyte green algae, commonly detected in Arctic samples (Lovejoy et al., 2006, 2007). Other eukaryotes in our samples (Figure 8) included one amoebozoan (SW170-108) in the deepest marine sample (170 m depth), the sea ice phylotype SI120-102, very similar to the flagellate *Telonema subtile*, and three phylotypes belonging to the Rhizaria, including two radiolaria (SW30-113 and SW70-42) found at different depths of the water column, and a cercozoan from sea ice (SI60-153) related to the species *Cryothecomonas aestivalis*, which is a diatom predator isolated from the North Sea (Drebes et al., 1996).

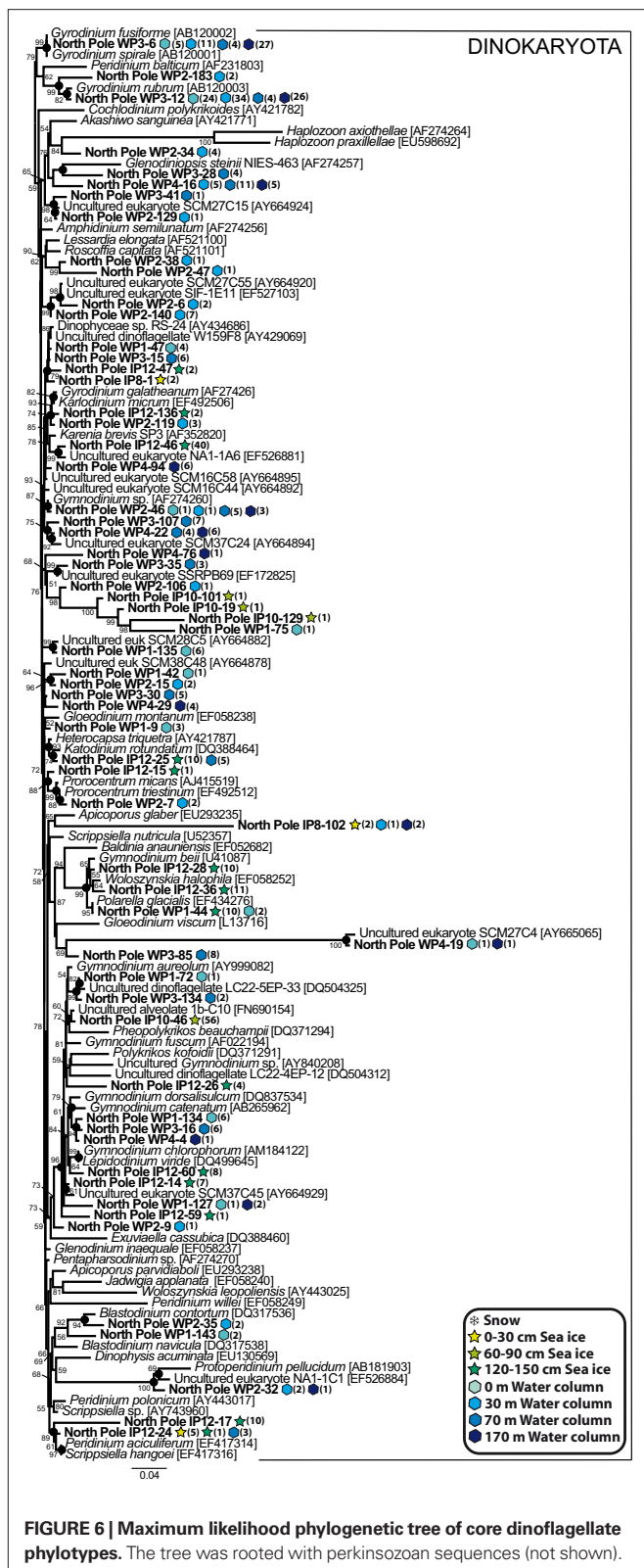
DISCUSSION

The taxonomic affiliation of the eukaryotic sequences associated to the snowpack sample was markedly different from that of the underlying sea ice and seawater. Snow was characterized by a clear dominance of fungal, fundamentally basidiomycete, and land plant



sequences (Figure 2). This could be easily explained by the high production of easily wind-dispersive spores and pollen grains by these organisms. Undoubtedly, these eukaryotic sequences did not

correspond to active organisms and, hence, our study attests to their cosmopolitan dispersal even to very remote regions. This was in agreement with a recent study of the diversity of microbes in



High Arctic snow samples showing the significant role played by wind in the aerial transport of microorganisms in this environment (Harding et al., 2011). In contrast with snow, the diversity of micro-

bial eukaryotes in sea ice and seawater was much more similar in terms of relative abundances of phyla, with typical marine protist lineages being present in sea ice, most notably core dinoflagellates, MAG I and Syndiniales. This comforted the traditional idea that organisms in the sea ice mostly derive from the trapping of marine protists during ice formation (Garrison et al., 1983; Lizotte, 2003). Whether these organisms are active in briny ice creeks, inactive waiting for ice to melt or both in varying proportions, remains to be determined, though previous studies have shown that sea ice can host complex active ecological webs (Brierley and Thomas, 2002; Lizotte, 2003).

The abundance of Syndiniales and, especially, MAG I phylotypes in North Pole sea ice and water samples is in line with numerous observations of these groups in different water masses and oceanic regions (Moreira and López-García, 2002; Harada et al., 2007; Guillou et al., 2008), indicating that parasitism is an important component also in these polar marine ecosystems. More precisely, the simultaneous abundance of very diverse core dinoflagellate species in the same samples supports the hypothesis that, at least partly, a number of these parasites may infect some of these dinoflagellate species. Nevertheless, we cannot discard that a proportion of the parasites detected in sea ice might correspond to dormant forms inactive during winter in the absence of their hosts, so that sea ice might serve as a seed bank of parasites waiting for favorable environmental conditions. More surprisingly, we detected only two diatom phylotypes in our samples (representing only 3 sequences among the 1000 analyzed in this work), which was not concordant with previous taxonomical and molecular studies made on Arctic environments showing a high abundance of diverse diatom species (Lizotte, 2003; Gast et al., 2004; Lovejoy et al., 2006; Eddie et al., 2010). Although it is known that methodological biases in clone libraries can lead to artificial low frequencies of prasinophytes and diatoms (Potvin and Lovejoy, 2009), this observation could be the result of the severe light limitation during polar winter that would decrease primary production and/or to possible grazing pressure by heterotrophs limiting algal accumulation. In fact, we obtained evidence supporting these two possibilities. On the one hand, we detected several phylotypes in closely related to typical diatom predators such as *Pirsonia*, *Cryothecomonas* and, notably, the dinoflagellate *Gyrodinium spirale*, a voracious diatom predator (Sherr and Sherr, 2007) which was one of the most abundant organisms found in our samples (see above). On the other hand, other phototrophic protists typical of Arctic regions, such as the prasinophyte green algae (Lovejoy et al., 2007), were almost completely absent from our libraries, in agreement with the very low chlorophyll concentration measured in the water column (<0.03 µg l⁻¹, V.I. Gagarin, personal communication). Likewise, among the dinoflagellate phylotypes for which we could infer their lifestyle based on their close proximity to known species, only one, SW0-44, related to the common sea ice species *Polarella glacialis* (Montresor et al., 1999), was a typical phototroph. It appeared that our samples were dominated by non-photosynthetic species much more extensively than commonly found in marine 18S rRNA gene libraries (Vaulot et al., 2002). The simplest explanation for this result can be that our samples were collected at the very end of the 6-month polar night, an extremely challenging period for phototrophic species that most likely decrease their density dramatically because of cell death and grazing pressure.

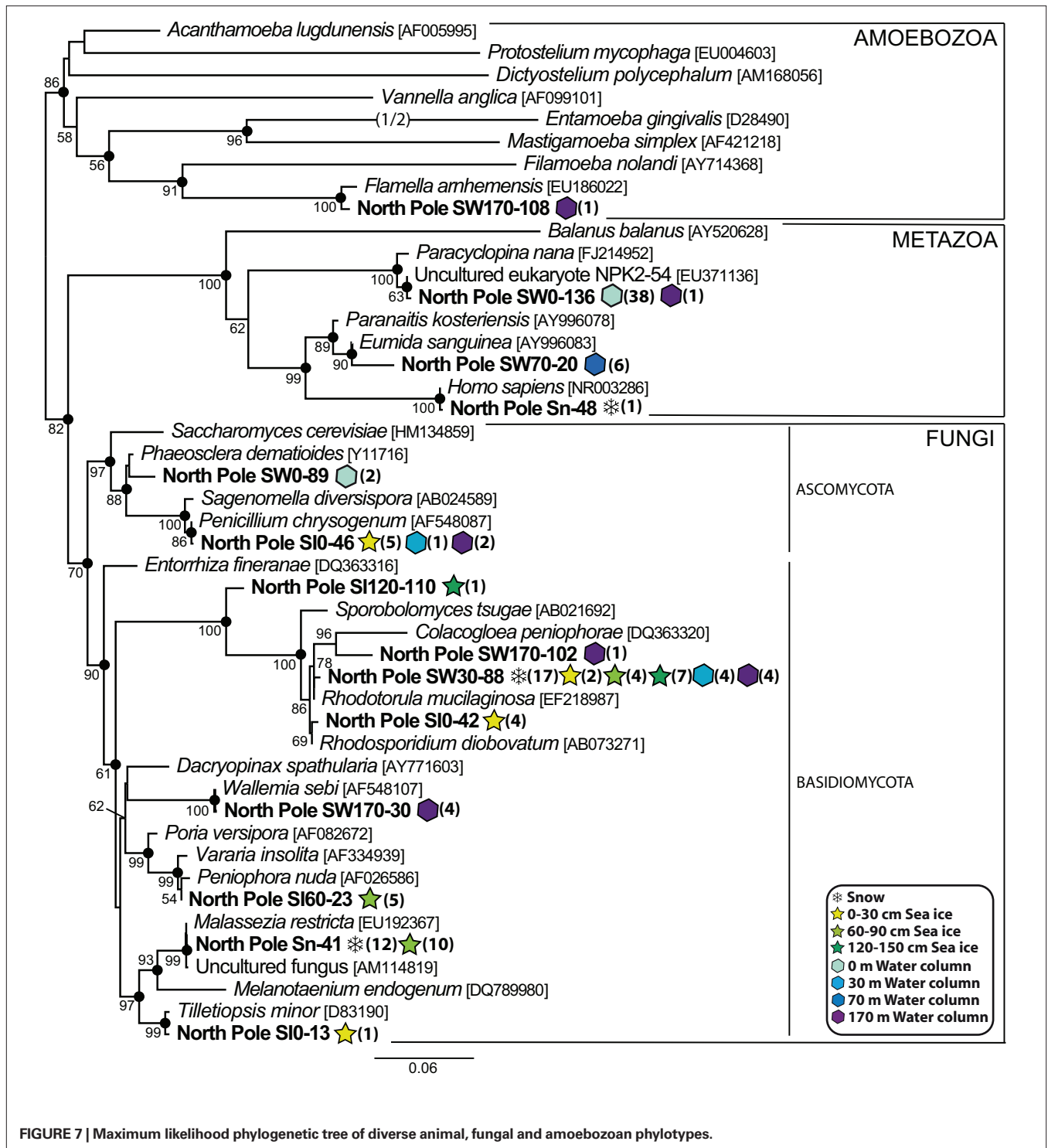
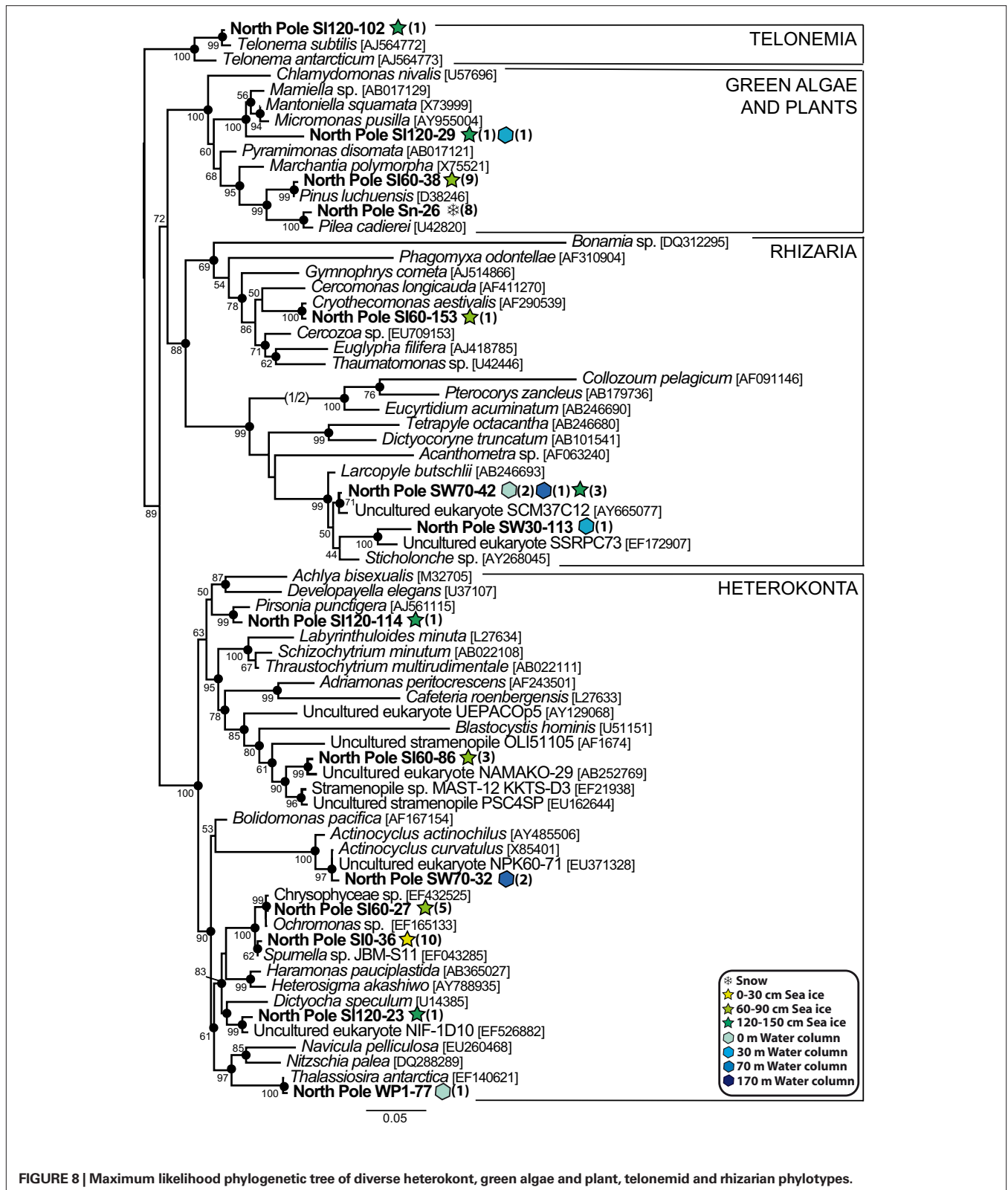


FIGURE 7 | Maximum likelihood phylogenetic tree of diverse animal, fungal and amoebozoan phylotypes.

The notable scarcity of phototrophic species in our libraries was compensated by a relatively large frequency of potential mixotrophic organisms. In fact, mixotrophy, a nutritional mode that combines autotrophy and heterotrophy, has been proposed to be a much more widespread strategy in marine and freshwater ecosystems than previously thought (Sanders, 1991; Jones, 2000), with a prominent role particularly in polar

environments (Moorthi et al., 2009). Likely mixotrophic candidates in our samples were several of the dinoflagellate phylotypes, especially those closely related to species of the genera *Gyrodinium*, *Gymnodinium*, *Blastodinium*, *Katodinium*, *Prorocentrum*, and *Woloszynskia* (Sherr and Sherr, 2007; Yoo et al., 2009; Kang et al., 2011), some of them at relatively high frequencies (Figure 6), as well as the ciliate phylotypes closely



related to *Myrionecta rubra* and *Strombidium conicum*, which possess plasts obtained from partial digestion of algal prey (Gustafson et al., 2000; Johnson and Stoecker, 2005; Myung

et al., 2006). These observations suggest that mixotrophy was probably more prevalent than strict phototrophy in our samples. As explained above, this was probably due to the particular

nature of those samples, collected after several months of darkness (because of the polar night strongly accentuated by the shading effect of the snowpack and sea ice cover). Under that strong environmental constraint, not only typical heterotrophs but also mixotrophs are advantaged over the phototrophic species. The atypical character of our samples, collected at the end of the polar night, made possible to detect with relatively high frequencies organisms that are otherwise rare in samples collected mostly in other seasons, when typical phototrophs largely dominate. Future temporal studies with samples taken

regularly all along the polar night will be useful to better understand how this shift in the composition of the microbial community occurs.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Arrigo, K. R. (2003). "Primary production in sea ice," in *Sea Ice – An Introduction to its Physics, Chemistry, Biology and Geology*, eds D. N. Thomas and G. S. Dieckmann (Oxford: Blackwell Publishing), 143–183.
- Behnke, A., Barger, K. J., Bunge, J., and Stoeck, T. (2010). Spatio-temporal variations in protistan communities along an O/HS gradient in the anoxic Framvaren Fjord (Norway). *FEMS Microbiol. Ecol.* 72, 89–102.
- Booth, B. C., and Horner, R. A. (1997). Microalgae on the Arctic Ocean section, 1994: species abundance and biomass. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 44, 1607–1622.
- Brierley, A. S., and Thomas, D. N. (2002). Ecology of southern ocean pack ice. *Adv. Mar. Biol.* 43, 171–276.
- Campbell, I. D., McDonald, K., Flannigan, M. D., and Kringayark, J. (1999). Long-distance transport of pollen into the Arctic. *Nature* 399, 29–30.
- Chambouvet, A., Morin, P., Marie, D., and Guillou, L. (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* 322, 1254–1257.
- Comiso, J. C. (2003). "Large scale characteristics and variability of the global sea ice cover," in *Sea ice – An Introduction to its Physics, Chemistry, Biology and Geology*, eds D. N. Thomas and G. S. Dieckmann (Oxford: Blackwell Publishing), 112–142.
- Dolven, J. K., Lindqvist, C., Albert, V. A., Bjorklund, K. R., Yuasa, T., Takahashi, O., and Mayama, S. (2007). Molecular diversity of alveolates associated with neritic north atlantic radiolarians. *Protist* 158, 65–76.
- Drebes, G., Kühn, S., Gmelch, A., and Schnepf, E. (1996). *Cryothecomona aestivalis* sp. nov., a colourless nanoflagellate feeding on the marine centric diatom *Guinardia delicatula*; (Cleve) Hasle. *Helgol. Mar. Res.* 50, 497–515.
- Eddie, B., Juhl, A., Krembs, C., Baysinger, C., and Neuer, S. (2010). Effect of environmental variables on eukaryotic microbial community structure of land-fast Arctic sea ice. *Environ. Microbiol.* 12, 797–809.
- Eicken, H. (2003). "From the microscopic, to the macroscopic, to the regional scale," in *Sea Ice – An Introduction to its Physics, Chemistry, Biology and Geology*, eds D. N. Thomas and G. S. Dieckmann (Oxford: Blackwell Publishing), 22–81.
- Garrison, D. L., Ackley, S. F., and Buck, K. R. (1983). A physical mechanism for establishing algal populations in frazil ice. *Nature* 306, 363–365.
- Gast, R. J., Dennett, M. R., and Caron, D. A. (2004). Characterization of protistan assemblages in the Ross Sea, Antarctica, by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 70, 2028–2037.
- Gast, R. J., Moran, D. M., Beaudoin, D. J., Blythe, J. N., Dennett, M. R., and Caron, D. A. (2006). Abundance of a novel dinoflagellate phylotype in the Ross sea, Antarctica. *J. Phycol.* 42, 233–242.
- Gradinger, R. (1999). Vertical fine structure of the biomass and composition of algal communities in Arctic pack ice. *Mar. Biol.* 133, 745–754.
- Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., Scanlan, D. J., and Worden, A. Z. (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ. Microbiol.* 10, 3349–3365.
- Gustafson, D. E., Stoecker, D. K., Johnson, M. D., Van Heukelem, W. F., and Sneider, K. (2000). Cryptophyte algae are robbed of their organelles by the marine ciliate *Mesodinium rubrum*. *Nature* 405, 1049–1052.
- Harada, A., Ohtsuka, S., and Horiguchi, T. (2007). Species of the parasitic genus *Dubosquella* are members of the enigmatic marine Alveolate group I. *Protist* 158, 337–347.
- Harding, T., Jungblut, A. D., Lovejoy, C., and Vincent, W. F. (2011). Microbes in high Arctic snow and implications for the cold biosphere. *Appl. Environ. Microbiol.* doi:10.1128/AEM.02611-10. [Epub ahead of print].
- Hasle, G. R., and Heimdal, B. R. (1998). The net phytoplankton in Kongsfjorden, Svalbard, July 1988, with general remarks on species composition of arctic phytoplankton. *Polar Res.* 17, 31–52.
- Horner, R. (1985). *Sea Ice Biota*. Boca Raton, FL: CRC Press.
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Ikävalko, J. U., and Thomsen, H. A. (1997). Baltic sea ice biota (March 1994): a study of the protistan community. *Eur. J. Protistol.* 33, 229–243.
- Jakobsson, M., Macnab, R., Mayer, L., Anderson, R., Edwards, M., Hatzky, J., Schenke, H. W., and Johnson, P. (2008). An improved bathymetric portrayal of the Arctic Ocean: implications for ocean modeling and geological, geophysical and oceanographic analyses. *Geophys. Res. Lett.* 35, L07602.
- Jobb, G., Von Haeseler, A., and Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* 4, 18. doi: 10.1186/1471-2148-4-18
- Johnson, M. D., and Stoecker, D. K. (2005). Role of feeding in growth and photophysiology of *Myrionecta rubra*. *Aquat. Microb. Ecol.* 39, 303–312.
- Jones, R. I. (2000). Mixotrophy in planktonic protists: an overview. *Freshw. Biol.* 45, 219–226.
- Kang, N. S., Jeong, H. J., Yoo, Y. D., Yoon, E. Y., Lee, K. H., Lee, K., and Kim, G. (2011). Mixotrophy in the newly described phototrophic dinoflagellate *Woloszynskia cincta* from Western Korean waters: feeding mechanism, prey species and effect of prey concentration. *J. Eukaryot. Microbiol.* 58, 152–170.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
- Kohlmeyer, J., and Kohlmeyer, E. (1979). *Marine Mycology: The Higher Fungi*. New York, NY: Academic Press.
- Kolodziej, K., and Stoeck, T. (2007). Cellular Identification of a novel uncultured marine stramenopile (MAST-12 Clade) small-subunit rRNA gene sequence from a Norwegian estuary by use of fluorescence in situ hybridization-scanning electron microscopy. *Appl. Environ. Microbiol.* 73, 2718–2726.
- Kühn, S., Medlin, L., and Eller, G. (2004). Phylogenetic position of the parasitoid nanoflagellate *Pirsonia* inferred from nuclear-encoded small subunit ribosomal DNA and a description of *Pseudopirsonia* n. gen. and *Pseudopirsonia mucosa* (Drebes) comb. nov. *Protist* 155, 143–156.
- Lenke, P., Ren, J., Alley, R. B., Allison, I., Carrasco, J., Flato, G., Fujii, Y., Kaser, G., Mote, P., Thomas, R. H., and Zhang, T. (2007). "Observations: changes in snow, ice and frozen ground," in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Cambridge: Cambridge University Press), 337–383.
- Lin, S., Zhang, H., Hou, Y., Miranda, L., and Bhattacharya, D. (2006). Development of a dinoflagellate-oriented PCR primer set leads to detection of picoplanktonic dinoflagellates from Long Island Sound. *Appl. Environ. Microbiol.* 72, 5626–5630.
- Lizotte, M. P. (2003). "The microbiology of sea ice," in *Sea Ice – An Introduction to its Physics, Chemistry, Biology and Geology*, eds D. N. Thomas and G. S. Dieckmann (Oxford: Blackwell Publishing), 184–210.
- López-García, P., Philippe, H., Gail, F., and Moreira, D. (2003). Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc. Natl. Acad. Sci. U.S.A.* 100, 697–702.

- López-García, P., Rodríguez-Valera, F., Pedros-Alio, C., and Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* 409, 603–607.
- López-García, P., Vereshchaka, A., and Moreira, D. (2007). Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. *Environ. Microbiol.* 9, 546–554.
- Lovejoy, C., Massana, R., and Pedrós-Alió, C. (2006). Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl. Environ. Microbiol.* 72, 3085–3095.
- Lovejoy, C., Vincent, W. F., Bonilla, S., Roy, S., Martineau, M. J., Terrado, R., Potvin, M., Massana, R., and Pedrós-Alió, C. (2007). Distribution, phylogeny, and growth of cold-adapted picoplankton in Arctic seas. *J. Phycol.* 43, 78–89.
- Luo, W., Li, H., Cai, M., and He, J. (2009). Diversity of microbial eukaryotes in Kongsfjorden, Svalbard. *Hydrobiologia* 636, 233–248.
- Massana, R., Balagué, V., Guillou, L., and Pedrós-Alió, C. (2004). Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol. Ecol.* 50, 231–243.
- Melnikov, I. A. (1997). *The Arctic Sea Ice Ecosystem*. Amsterdam: Gordon and Breach Science Publishers.
- Mock, T., and Thomas, D. N. (2005). Recent advances in sea-ice microbiology. *Environ. Microbiol.* 7, 605–619.
- Montesoro, M., Procaccini, G., and Stoeker, D. K. (1999). *Polarella glacialis*, gen. nov., sp. nov. (Dinophyceae): Suessiaceae are still alive! *J. Phycol.* 35, 186–197.
- Moon-Van Der Staay, S. Y., De Wachter, R., and Vault, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409, 607–610.
- Moorthi, S., Caron, D. A., Gast, R. J., and Sanders, R. W. (2009). Mixotrophy: a widespread and important ecological strategy for planktonic and sea-ice nanoflagellates in the Ross Sea, Antarctica. *Aquat. Microb. Ecol.* 54, 269–277.
- Moreira, D., and López-García, P. (2002). The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol.* 10, 31–38.
- Moreira, D., and López-García, P. (2003). Are hydrothermal vents oases for parasitic protists? *Trends Parasitol.* 19, 556–558.
- Myung, G., Yih, W., Kim, H. S., Park, J. S., and Cho, B. C. (2006). Ingestion of bacterial cells by the marine photosynthetic ciliate *Myrionecta rubra*. *Aquat. Microb. Ecol.* 44, 175–180.
- Newell, S. Y. (1996). Established and potential impacts of eukaryotic mycelial decomposers in marine/terrestrial ecotones. *J. Exp. Mar. Biol. Ecol.* 200, 187–206.
- Not, F., Gausling, R., Azam, F., Heidelberg, J. F., and Worden, A. Z. (2007). Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ. Microbiol.* 9, 1233–1252.
- Olli, K., Wassmann, P., Reigstad, M., Ratkova, T. N., Arashkevich, E., Pasternak, A., Matrai, P. A., Knulst, J., Tranvik, L., Klais, R., and Jacobsen, A. (2007). The fate of production in the central Arctic Ocean – top-down regulation by zooplankton expatriates? *Prog. Oceanogr.* 72, 84–113.
- Palmisano, A. C., and Garrison, D. L. (1993). “Microorganisms in Antarctic sea ice,” in *Antarctic Microbiology*, ed. E. I. Friedman (New York, NY: Wiley-Liss), 167–218.
- Perovich, D. K., Andreas, E. L., Curry, J. A., Eiken, H., Fairall, C. W., Grenfell, T. C., Guest, P. S., Intrieri, J., Kadko, D., Lindsay, R. W., McPhee, M. G., Morison, J., Moritz, R. E., Paulson, C. A., Pegau, W. S., Persson, P. O. G., Pinkel, R., Richter, M. J. A., Stanton, T., Stern, H., Sturm, M., Tucker, W. B., and Uttal, T. (1999). *Year on Ice Gives Climate Insights*. Washington, DC: American Geophysical Union.
- Philippe, H. (1993). MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res.* 21, 5264–5272.
- Piquet, A., Scheepens, J., Bolhuis, H., Wiencke, C., and Buma, A. (2010). Variability of protistan and bacterial communities in two Arctic fjords (Spitsbergen). *Polar Biol.* 33, 1521–1536.
- Potvin, M., and Lovejoy, C. (2009). PCR-based diversity estimates of artificial and environmental 18S rRNA gene libraries. *J. Eukaryot. Microbiol.* 56, 174–181.
- Richards, T. A., and Bass, D. (2005). Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr. Opin. Microbiol.* 8, 240–252.
- Richter-Menge, J. A., Perovich, D. K., Elder, B. C., Claffey, K., Rigor, I., and Ortmeier, M. (2006). “Ice mass-balance buoys: a tool for measuring and attributing changes in the thickness of the Arctic sea-ice cover,” in *Annals of Glaciology*, eds P. Langhorne and V. Squire (Cambridge: International Glaciological Society), 205–210.
- Romari, K., and Vault, D. (2004). Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol. Oceanogr.* 49, 784–798.
- Sanders, R. W. (1991). Mixotrophic protists in marine and freshwater ecosystems. *J. Eukaryot. Microbiol.* 38, 76–81.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., and Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.
- Sherr, E. B., and Sherr, B. F. (2007). Heterotrophic dinoflagellates: a significant component of microzooplankton biomass and major grazers of diatoms in the sea. *Mar. Ecol. Prog. Ser.* 352, 187–197.
- Stoeker, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V., and Epstein, S. (2007). Protistan diversity in the Arctic: a case of paleoclimate shaping modern biodiversity? *PLoS ONE* 2, e278. doi: 10.1371/journal.pone.0000728
- Terrado, R., Vincent, W. F., and Lovejoy, C. (2009). Mesopelagic protists: diversity and succession in a coastal Arctic ecosystem. *Aquat. Microb. Ecol.* 56, 25–40.
- Thomas, D. N., Lara, R. J., Eicken, H., Kattner, G., and Skoog, A. (1995). Dissolved organic matter in Arctic multi-year sea ice during winter: major components and relationship to ice characteristics. *Polar Biol.* 15, 477–483.
- Tian, F., Yu, Y., Chen, B., Li, H., Yao, Y.-F., and Guo, X.-K. (2009). Bacterial, archaeal and eukaryotic diversity in Arctic sediment as revealed by 16S rRNA and 18S rRNA gene clone libraries analysis. *Polar Biol.* 32, 93–103.
- Vault, D., Romari, K., and Not, F. (2002). Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends Microbiol.* 10, 266–267.
- Werner, I., Ikävalko, J., and Schünemann, H. (2007). Sea-ice algae in Arctic pack ice during late winter. *Polar Biol.* 30, 1493–1504.
- Wu, Z., Tsumura, Y., Blomquist, G., and Wang, X. R. (2003). 18S rRNA gene variation among common airborne fungi, and development of specific oligonucleotide probes for the detection of fungal isolates. *Appl. Environ. Microbiol.* 69, 5389–5397.
- Yoo, Y. D., Jeong, H. J., Kim, M. S., Kang, N. S., Song, J. Y., Shin, W., Kim, K. Y., and Lee, K. (2009). Feeding by phototrophic red-tide dinoflagellates on the ubiquitous marine diatom *Skeletonema costatum*. *J. Eukaryot. Microbiol.* 56, 413–420.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2011; accepted: 28 April 2011; published online: 11 May 2011.

Citation: Bachy C, López-García P, Vereshchaka A and Moreira D (2011) Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the North Pole at the end of the polar night. *Front. Microbio.* 2:106. doi: 10.3389/fmicb.2011.00106

This article was submitted to *Frontiers in Extreme Microbiology*, a specialty of *Frontiers in Microbiology*.

Copyright © 2011 Bachy, López-García, Vereshchaka and Moreira. This is an open-access article subject to a non-exclusive license between the authors and *Frontiers Media SA*, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other *Frontiers* conditions are complied with.