UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Departamento de Biología

Institut de Ciències del Mar
CSIC
Consejo Superior de Investigaciones Científicas

Ecology of an uncultured heterotrophic flagellate lineage: MAST-4

(Ecología de un linaje de flagelados heterotróficos no cultivados, el MAST-4)

Raquel Rodríguez Martínez

Tesis Doctoral presentada por Dª Raquel Rodríguez Martínez para obtener el grado de Doctor por la Universidad de las Palmas de Gran Canaria, Departamento de Biología, Programa en Oceanografía (Bienio 2006-2008)

Director: Ramon Massana i Molera

Universidad de Las Palmas de Gran Canaria

Institut de Ciències del Mar (ICM-CSIC)

En Barcelona a, de de

La Doctoranda

Raquel Rodríguez Martínez

El Director

Ramon Massana i Molera

Anexo I

D/Dª.....SECRETARIO/A DEL DEPARTAMENTO DE..... DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA,

CERTIFICA,

Que el Consejo de Doctores del Departamento en su sesión de fecha.....tomó el acuerdo de dar el consentimiento para su tramitación, a la tesis doctoral titulada “Ecology of an uncultured heterotrophic flagellate lineage: MAST-4” presentada por la doctoranda Dª Raquel Rodríguez Martínez y dirigida por el Doctor Ramon Massana i Molera.

Y para que así conste, y a efectos de lo previsto en el Artº 73.2 del Reglamento de Estudios de Doctorado de esta Universidad, firmo la presente en Las Palmas de Gran Canaria, a.....de.....de dos mil.....

A mi familia,
a mi tía María Jesús,
a mi sobrino Víctor
y a todos los que me habéis
acompañado y apoyado en este recorrido.

"Our days are precious but we gladly see them going
If in their place we find a thing more precious growing:
A rare, exotic plant, our gardener's heart delighting;
A child whom we are teaching, a booklet we are writing."

Hermann Hesse, *Das Glasperlenspiel*

Contents

Glossary

18S rDNA Gene coding the RNA component of the small ribosomal subunit, and widely used to identify and classify eukaryotic microorganisms.

Automated Ribosomal Interspacer Analysis (ARISA) Fingerprinting technique that describes the diversity of an assemblage based on the ITS size variation of its members.

Compensatory base changes (CBC) Mutations that occur in both nucleotides of a paired structural position of the ribosomal RNA while retaining the paired nucleotide bond.

Clone library Heterogeneous collection of cloned sequences (often 18S rDNA) derived from a complex assemblage of organisms.

Environmental molecular surveys Retrieval of genetic signatures from a complex microbial assemblage for diversity studies.

Epifluorescence microscopy Technique that allows observation by fluorescence of very small stained or autofluorescent cells retained on a filter.

Flagellates Single-celled protists, phototrophic or heterotrophic, with one or more whip-like organelles, called flagella, often used for propulsion or to create feeding currents.

Fluorescent In Situ Hybridization (FISH) Microscopic method for detection of microbial cells by labeling them with a fluorescent probe that specifically labels ribosomes.

Functional response The predation rate of a consumer as a function of food density.

Internal Transcribed Spacer (ITS) Non-coding regions separating the individual genes of the ribosomal DNA operon, with a higher degree of variation than the genic regions.

Marine alveolates (MALV) Clades without cultured representatives that belong to the eukaryote supergroup alveolates, detected in molecular surveys of marine picoplankton.

Marine stramenopiles (MAST) Clades without cultured representatives that belong to the eukaryote supergroup stramenopiles, detected in molecular surveys of marine picoplankton.

Numerical response The growth rate of a consumer as a function of food density.

Phylogenetic clade (or lineage) Set of related sequences that originate from a single common ancestor.

Picoeukaryotes Phototrophic and heterotrophic protists smaller than 3 μm.

Population All the organisms that both belong to the same group or species and live in the same geographical area.

Protists General term for eukaryotes not belonging to plants, animals, fungi, or macroalgae, generally single-celled organisms of sizes from 1 μm to more than 100 μm.

Quantitative Polymerase Chain Reaction (Q-PCR) Also called real-time polymerase chain reaction (RT-PCR**).** Molecular technique based on the quantification of a targeted DNA molecule during its amplification by PCR.

Secondary structure Basepair interactions within a single molecule of RNA decomposed into stems and loops.

Taxon Systematic unit to design a hierarchical level in the classification of organisms.

Acronyms

ARISA	Automated Ribosomal Interspacer Analysis
CBC	Compensatory Base Change
CCTH	Cryptophytes, Centrohelids, Telonemids plus Haptophytes
DAPI	4,6-diamidino-2-phenylindole
FISH	Fluorescent In Situ Hybridization
FLB	Fluorescently labeled bacteria
Fst	Fixation Index
HF	Heterotrophic Flagellates
ITS	Internal Transcribed Spacer
Ks	Half-saturation constant
MALV	Marine Alveolates
MAST	Marine Stramenopiles
MJ	Median-Joining
OTU	Operational taxonomic Unit
PF	Phototrophic Flagellates
PCR	Polymerase Chain Reaction
Q-PCR	Quantitative Polymerase Chain Reaction
RAS	Rhizaria, Alveolates and Stramenopiles
rDNA	Ribosomal Deoxyribonucleic Acid
SAG	Single amplified genome

Summary

Heterotrophic flagellates are fundamental in marine microbial food webs. They play a key role in channeling bacteria to higher trophic levels as well as in nutrient recycling. Despite this recognized ecological importance, and contrasting with the significant advances achieved with marine bacteria and archaea, little is known on the diversity of marine heterotrophic flagellates. This heterogeneous group of cells is weakly captured by microscopic and culturing approaches, and has been only recently targeted by environmental molecular surveys. This thesis is about the study of the ecology of one specific uncultured lineage, the MAST-4 (Marine Stramenopiles clade-4) group, which can be viewed as a model heterotrophic flagellate. MAST-4 consists of tiny cells (2-3 μm) that are widely distributed in the photic zone of the oceans with temperatures above ~5°C. On average, it accounts for ~9% of marine heterotrophic flagellates. It actively grazes on bacteria and other picosized cells being well adapted to the typical bacterioplankton abundances and preferring bacteria in a good physiological state. MAST-4 displays low genetic variability, being formed by just five main clades, each representing at least one biological species. This indicates that this successful group has suffered low evolutionary diversification. The most represented MAST-4 clades did not reveal geographical barriers, whereas temperature was the main factor influencing the distribution patterns. Different clades of MAST-4 seemed to have physiological adaptations that allowed them to establish and dominate under different environmental regimes. This ecotypic differentiation could partly explain the success of this flagellate, at the level of global distribution and abundance, allowing it to exploit the whole spectrum of habitat variability.

Resumen

Los flagelados heterotróficos son un componente fundamental en las redes microbianas marinas. Juegan un papel clave en la canalización de las bacterias hacia niveles tróficos superiores y en el reciclaje de nutrientes inorgánicos. A pesar de esta reconocida importancia ecológica, a diferencia de los grandes avances alcanzados en el estudio de las bacterias y arqueas marinas, todavía se conoce bastante poco de la diversidad de los flagelados heterotróficos marinos. La identificación de este grupo heterogéneo de células se ha visto limitada por la dificultad de su detección por microscopía o su aislamiento en cultivo. El desarrollo de los recientes análisis moleculares ambientales ha permitido comenzar a caracterizarlos. Esta tesis aborda el estudio de la ecología de un linaje específico no cultivado, el grupo MAST-4 (estramenópilos marinos clado-4), el cual podría considerarse como un flagelado heterotrófico modelo. El linaje MAST-4 está formado por células pequeñas (2-3 μm) ampliamente distribuidas en la zona fótica de los océanos con temperaturas superiores a ≈5°C. Como promedio, representa el ≈9% de los flagelados heterotróficos marinos. Es un depredador activo de bacterias y otras células muy pequeñas y parece estar bien adaptado a las abundancias típicas del bacterioplancton, prefiriendo las bacterias en buen estado fisiológico. El taxón MAST-4 muestra una variabilidad genética baja, se compone solo de cinco clados principales, cada uno representando al menos una especie biológica. Esto indica que este grupo tan exitoso ha experimentado una diversificación evolutiva limitada. Los clados más representados del MAST-4 no revelaron barreras geográficas, en cambio, se observó cómo la temperatura era el principal factor estudiado que influenciaba la distribución de los patrones. Distintos clados del MAST-4 parecían tener adaptaciones fisiológicas que les permitían establecerse y dominar bajo diferentes condiciones ambientales. Esta diferenciación ecotípica podría explicar parcialmente el éxito de este flagelado, a nivel de distribución y abundancia global, permitiéndole explotar todo el espectro de la variabilidad del hábitat marino.

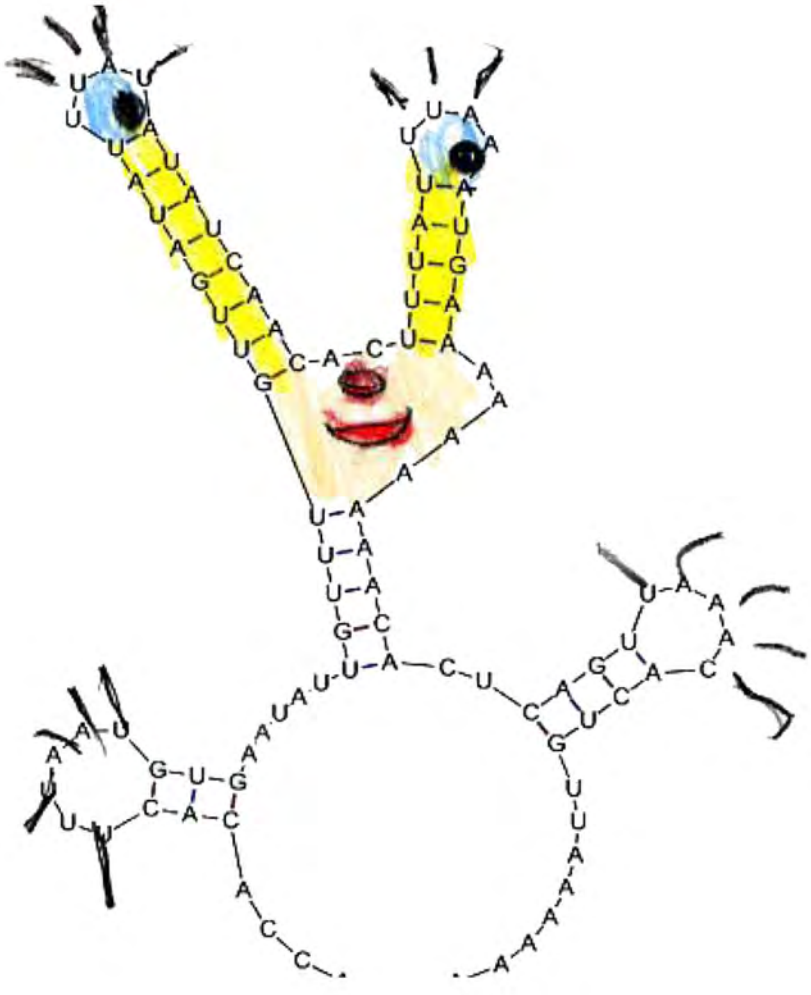
Resum

Els flagel·lats heterotròfics són un component fonamental de les xarxes tròfiques microbianes marines. Juguen un paper clau en la canalització de la biomassa bacteriana cap a nivells tròfics superiors i en el reciclatge dels nutrients inorgànics. Malgrat la reconeguda importància ecològica dels flagel·lats heterotròfics, i a diferència dels importants avenços aconseguits amb els bacteris i els arqueobacteris marins, encara hi ha un gran desconeixement sobre la seva diversitat al mar. La identificació d'aquest grup heterogeni de cèl·lules s'ha vist limitada per la dificultat de la seva detecció per microscòpia i del seu aïllament en cultiu. El desenvolupament dels recents anàlisis moleculars ha permès començar a caracteritzar-los. Aquesta tesi tracta sobre l'ecologia d'un llinatge específic no cultivat, MAST-4 (Marine Stramenopile-4), que pot ser vist com un flagel·lat heterotròfic model. El grup MAST-4 es compon de cèl·lules petites (2-3 micres) que estan àmpliament distribuïdes a la zona fòtica dels oceans. De mitjana, representen el ~9% dels flagel·lats heterotròfics marins. S'alimenten activament de bacteris i altres cèl·lules molt petites, i semblen estar ben adaptats a l'abundància típica del bacterioplàncton, preferint bacteris en un bon estat fisiològic. El grup MAST-4 mostra una baixa variabilitat genètica, essent format per només cinc clades principals, cadascun representant almenys una espècie biològica. Això indica que aquest grup tant exitós ha experimentat una limitada diversificació evolutiva. Els clades de MAST-4 més representats no pateixen barreres geogràfiques, tot i que la temperatura sí que limita i condiciona la seva distribució. Diferents subtipus de MAST-4 semblen tenir adaptacions fisiològiques que els permeten establir-se i dominar en diferents règims ambientals. Aquesta diferenciació ecotípica podria explicar en part l'èxit d'aquest flagel·lat, a nivell de la seva distribució i abundància global, permetent-li explotar tot l'espectre de la variabilitat de l'hàbitat marí.

“How inappropriate to call this planet Earth when it is quite clearly Ocean.”

Sir Arthur C. Clarke

General Introduction



Eukaryotic tree of life

Eukaryotes are only one of the three domains of life, along with Bacteria and Archaea. One of the reasons of our large curiosity in them is that they include the organisms we can see. Our understanding of eukaryote biology, ecology and evolution is dominated by the study of land plants, animals and fungi. However, these are only three isolated fragments of the full diversity of existing eukaryotes. The majority of eukaryotes, in terms of main lineages, number of different taxa and also total numbers of cells, consist of predominantly unicellular lineages. A surprising number of these lineages are poorly characterized. Nonetheless, knowledge of the morphological, functional and ecological diversity of microbial eukaryotes is fundamental to our understanding of eukaryote biology and the underlying forces that shaped it (Baldauf 2008).

During the second half of the 20th century, molecular developments provided a systematic way to relate all living organisms through DNA sequence comparisons, initially using the small-subunit ribosomal RNA gene: 16S rDNA in prokaryotes and 18S rDNA in eukaryotes (Woese 1987). Using the sequences from this gene, the eukaryotic phylogenetic tree of life appeared divided into only a few supergroups (Adl et al 2005, Baldauf 2003). Including several revisions and updates (Baldauf 2008), virtually all eukaryotes can now be assigned to one of the supergroups, which form a crown radiation without clear ranking among them and with an uncertain root (figure I.1). Few morphological or ultrastructural characters connect the diverse lineages within each supergroup, but the phylogenetic signatures are robust (Jürgens and Massana 2008). Although the configuration of supergroups varies, the general consensus includes (1) Unikonts, (2) Archaeplastida, (3) Rhizaria + Alveolates + Stramenopiles (RAS), (4) Excavates and (5) Cryptophytes, Centrohelids, Telonemids plus Haptophytes (CCTH). Unikonts include all eukaryotes thought to be primitively uniflagellate, that is, Opisthokonts (including animals, fungi and some protists such as choanoflagellates) and Amoebozoa (Cavalier-Smith 2002). Archaeoplastida is the group in which eukaryotic photosynthesis first arose and includes green algae and land plants (Adl et al 2005, Archibald and Keeling 2005). The RAS group was recently proposed (Burki et al 2007, Hackett et al 2007) to unite three very heterogeneous supergroups, the rhizaria (cercozoans, radiolaria and foraminifera), the alveolates (dinoflagellates, ciliates and apicomplexa) and the stramenopiles (see below). The Excavates are formed by two distinct groups, the mitochondriate excavates that include Euglenids, Heterolobosea and core Jakobids, and the amitochondriate excavates (i. e. Diplomonads and Parabasalids), a collection of highly derived taxa with simplified internal cell structure and lacking aerobic mitochondria. CCTH is a newly described supergroup (Burki et al 2009) proposed to relate several important but difficult to locate phylogenetic lineages, such as haptophytes, cryptophytes and telonemids (Shalchian-Tabrizi et al 2006). In addition, CCTH also include the katablepharids (known to be relatives of Cryptophytes) (Okamoto and Inouye 2005) and perhaps the picobiliphytes, a novel phytoplanktonic class initially unrelated to any supergroup (Not et al 2007). A consequence of the

molecular framework is that many incertae sedis protists (Patterson and Zöffel 1991) are finding their phylogenetic position in the eukaryotic tree. In addition, sequences of the 18S rRNA gene of cultured representatives are crucial in placing protists within this phylogenetic context (Cavalier-Smith and Chao 2003, Scheckenbach et al 2005).

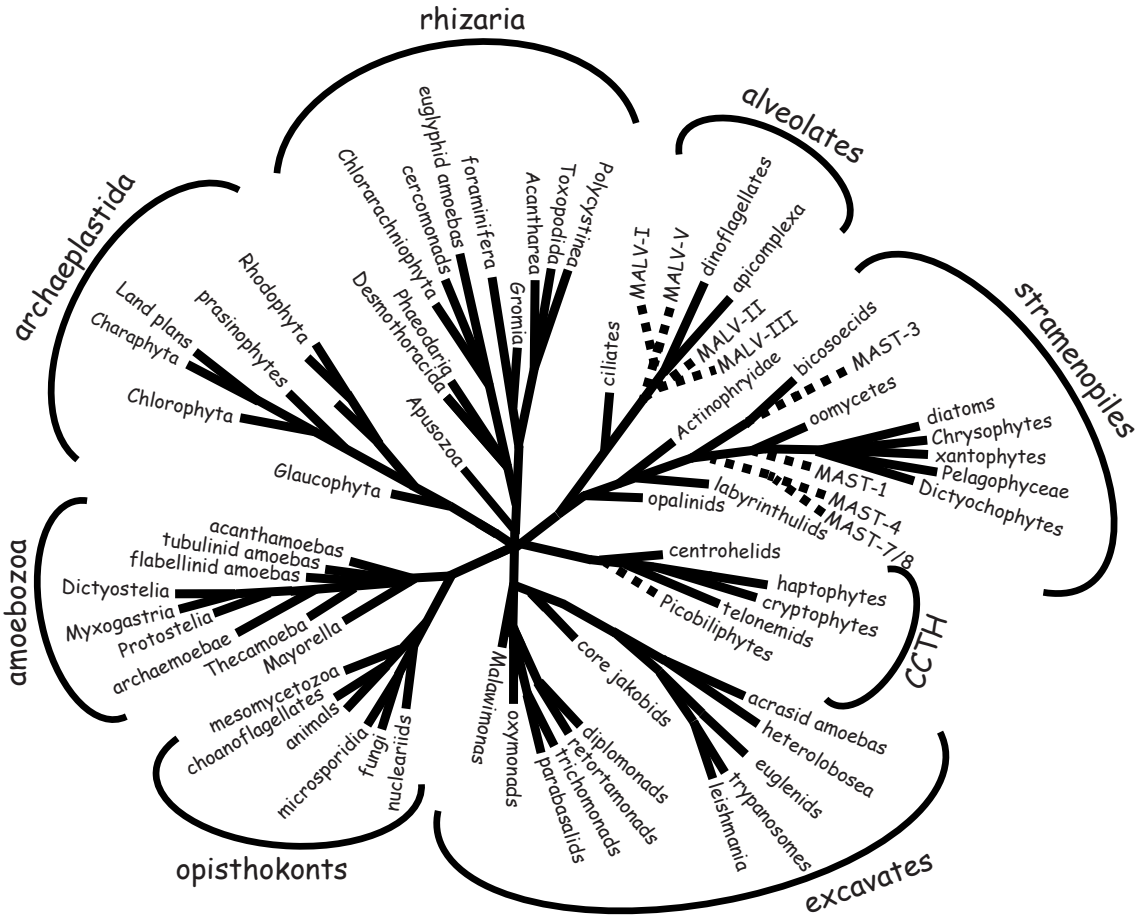


Figure I.1. Eukaryotic tree of life. A consensus phylogeny of the major eukaryotic groups based on published molecular phylogenetic and ultrastructural data (adapted from (Baldauf 2003)). Dotted lines indicate positions of major lineages known primarily from culture-independent molecular surveys. MALV (marine alveolates), MAST (marine stramenopiles) and CCTH (Cryptophytes, Centrohelids, Telonemids plus Haptophytes).

A body plan very common in the eukaryotic tree of life is that of unicellular colorless microorganisms with one or a few flagella. This type of organization, generally referred as protozoan (or heterotrophic) flagellates, can be seen in 27 of the 60 protists lineages among eukaryotes (Patterson and Larsen 1991). Thus, the flagellates are a grade of organization and not a consistent monophyletic assemblage. They are those organisms that spend most of their existence moving or feeding with a small number of flagella. Their size range between 1-2 μm up to 20 μm. Flagella arose early in eukaryote evolution, and we are not able to identify any groups of protists that are primitively without flagella. It is supposed that the last eukaryotic common ancestor was also a kind of flagellate originated by a symbiogenetic fusion between eubacteria and archaebacteria (Margulis et al 2006). And obviously, this primitive eukaryote was colorless and heterotrophic.

Microbial molecular surveys raise the eukaryotic diversity

The use of molecular biological approaches on microbial ecology, developed during the turn of the 20th century, has transformed the field of protistan diversity. In general, the identity of most small fragile protist was very difficult to assess by direct inspection of natural samples. So, a classical way of identification is to obtain these organisms in culture for a proper classification. In the case of autotrophic protists, the cultured strains are more or less representatives of the natural communities. This is probably because it is easier to simulate the natural conditions in a culturing bottle, since these cells require mostly inorganic nutrients and light. Nevertheless, it is likely that cultures do not cover the full in situ diversity of autotrophic protists (Vaulot et al 2008). In the case of heterotrophic bacterivorous protists, the long list of formally described species (Lee and Patterson 1998) derives mainly from cultures or enrichments started by adding a substrate for bacterial growth that, in turn, are the food for the protists. Cell cultures yield fundamental ecophysiological information but while it is obvious that these easily enriched strains live in the sea, it is doubtful that they are dominant members of natural assemblages. A now classic study demonstrated that the bacterivorous protists dominating several enrichments were rare in the original samples (Lim et al 1999). More recent studies have confirmed these results and have provided the mechanistic explanation of this culture bias in heterotrophic flagellates (HF) (del Campo 2011)

Environmental (culture-independent) studies of sequencing the 18S rRNA genes have resulted in an increased appreciation of the diversity of protists in nature. Thus, molecular surveys have revealed numerous sequences of unknown protists, indicating unanticipated levels of protistan diversity in many environments and retrieving very few sequences related to cultured protists (Amaral-Zettler et al 2009, Brown et al 2009, Countway et al 2007, Díez et al 2001, Head et al 1998, Lim 1996, López-García et al 2001, Moon-van der Staay et al 2001, Richards et al 2005, Stoeck et al 2006, Vigil et al 2009). In marine systems, these surveys have revealed a large number of uncultured lineages, such as the marine alveolates (MALV) and the marine stramenopiles (MAST) (Massana et al 2004a) that appear in virtually all studies. Still, most of this diversity remains poorly known. It is thus clear that culture isolations and molecular surveys are providing different views on the species composition of marine protists in general and of HF in particular. While often these environmental studies were ecologically-driven in order to identify the dominant members of natural assemblages, it is obvious that they have also provided fundamental new insights into eukaryotic phylogeny and new branches in the tree of life are exclusively composed by these so-far uncultured lineages.

Stramenopiles, an important supergroup in marine systems

The supergroup Stramenopiles (Adl et al 2005) are formed by many heterogenous lineages, some of them of crucial importance in marine systems. One of the few characteristics shared by most stramenopile motile cells is the presence of a flagellum with two opposite rows of mastigonemes, tripartite hairs ("stramenopiles"), which reverse the flow around the flagellum so that the cell is dragged forward rather than pushed along. Most also possess a second, shorter smooth flagellum (hence the alternative name "heterokont"). This extraordinarily diverse group includes numerous lineages of single-celled heterotrophs (bicosoecids) and phototrophs (diatoms), slime nets (labyrinthulids), plasmodial parasites (oomycetes), and large to giant multicellular algae (phaeophytes).

There are at least five known lineages of non-photosynthetic stramenopiles (figure I.2) (Baldauf 2008). Oomycetes (water molds and downy mildews) were previously classified as fungi and include numerous extremely destructive plant parasites such as *Phytophthora infestans*, the cause of potato blight, and *Plasmopara viticola*, the cause of grapevine downy mildew. The bicosoecids are small heterotrophic biflagellates, such as the well-known *Cafeteria* (Fenchel 1988). The *Blastocystis* spp. are commensals in the guts of animals (Stechmann et al 2008) and some species, like *Blastocystis hominis*, can infect humans. Labyrinthulids (slime nets) form filamentous "railway-like" networks patrolled by amoeboid-like cells. They were placed with the Thraustochytrids (Cavalier-Smith et al 1994), which also have the tendency to form cell aggregates. The fine taxonomy of both groups requires 18S ribosomal gene comparisons (Honda et al 1999).

Photosynthetic stramenopiles (figure I.3) are formed by at least eleven distinct lineages, including some of the most important and abundant algae (Baldauf 2008). Diatoms have intricately patterned bipartite silica tests that fit together like lidded boxes. They are ubiquitous and often abundant in marine and freshwaters, with ~11,000 described and possibly as much as 10^7 non-described species (Fehling et al 2007). Chrysophytes (golden algae) are generally free-swimming and unicellular, but there are also filamentous and colonial forms. Pigmented chrysophytes contain chlorophyll as well as a carotenoid called fucoxanthin that gives them a yellow-brown color. They were considered to be mostly freshwater, but recent studies suggest they could be rather abundant in the marine plankton (Fuller et al 2006, Lepère et al 2009, Shi et al 2011). Phaeophytes (brown algae) are particularly widespread in temperate intertidal and subtidal zones. They have true parenchyma and build "forests" in near-shore waters, as the giant sea kelp forests, supporting complex ecosystems including fish and marine mammals. Xanthophytes (yellow-green algae) are the dominant producers in some salt marshes and also form multicellular organisms. The remaining groups are formed by very small algae, such as Dictyochophytes, Eustigmatophytes,

Phaeothamniophyta, Pelagophytes and Pinguiophytes (Vaulot et al 2008). The Pelagophyceae is a recently described class (Andersen et al 1993), previously classified within the Chrysophyceae, that can be important in the oceanic picoplankton.

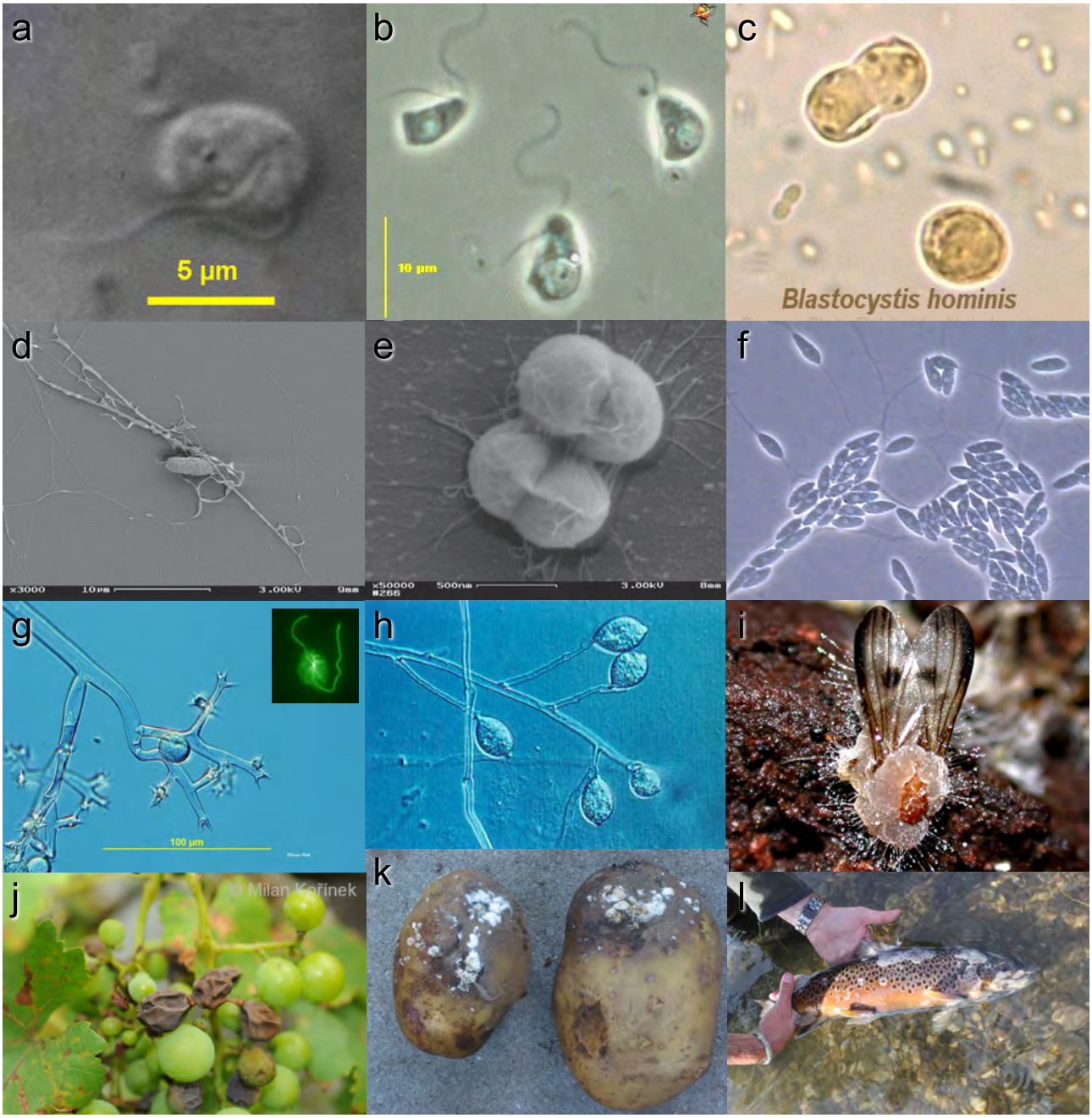


Figure I.2. Pictures of examples of heterotrophic stramenopiles. a) *Developayella elegans*; b) the bicosoecid *Cafeteria roenbergensis*; c) *Blastocystis hominis*; Labyrinthulids: d) *Aplanochytrium,* e) *Thraustochytrium* and f) *Labyrinthula terrestris*; Oomycetes: g) *Pasmopara viticola*, h) *Phytophthora infestans*, i) *Saprolegnia*. The bottom pictures are the infested hosts by the upper oomycetes, j) forming the mildew of the grapevines, k) potatoes and l) trouts. Pictures courtesy of WJ. Lee, D. Patterson, L.A. Zettler, V. Edgcomb, C. Leander, D. Porter, J. Harper, S. Lew and E. Haugen.

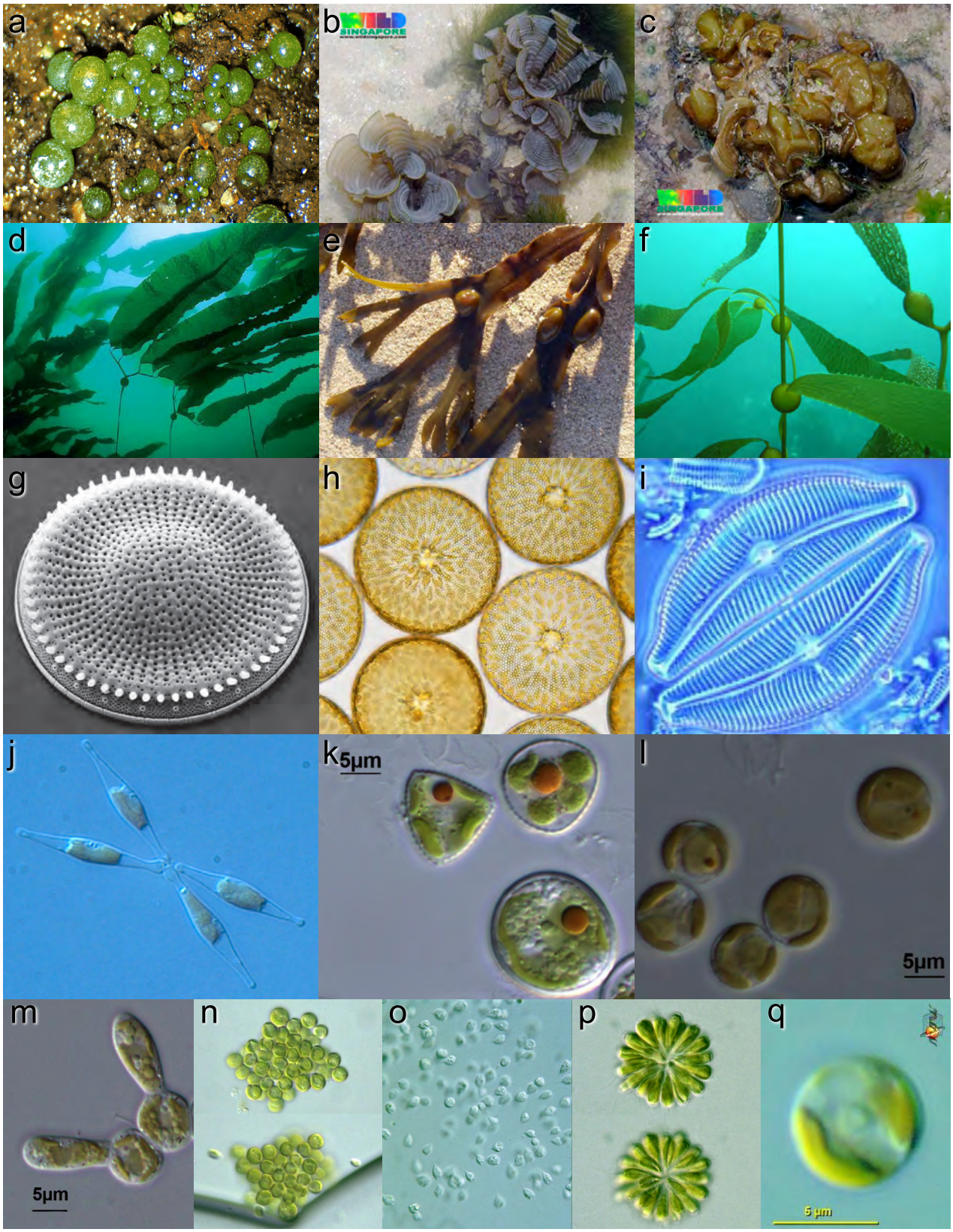
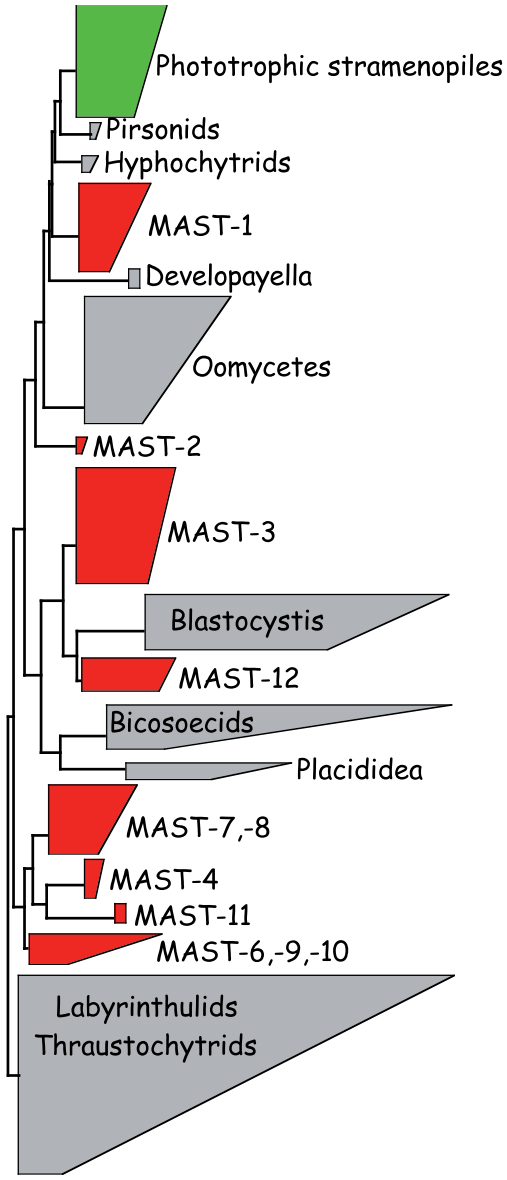


Figure I.3. Pictures of examples of photosintetic stramenopiles. a) The xantophyte *Botrydium*; Phaeophytes: b) *Padina,* c) *Colpomenia*, d) *Pelagophycus porra*, e) *Fucus vesiculosus* and f) *Macrocystis integrifolia*; Diatoms: g) *Stephanodiscus,* h) *Coscinodiscus*, i) *Cymbella tumida* and j) *Phaeodactylum tricornutum*; k) the eustigmatophyte strain 29.96; Phaeothamniophytes: l) *Stichogloea doederleinii* and m) *Phaeothamnion confervicola*; Chrysophytes: n) *Chrysocapsa epiphytica,* o) *Spumella* sp. and p) *Synura* and q) the pinguiophyta *Pinguiococcus pyrenoidosus*. Pictures courtesy of I. Inouye, R. Tan, E. Bierman, D. Mann, A. de Martino, C. Bowler, J.C. Baley, Y. Tsukii, D. Patterson, B. Andersen and U.S. Geological Survey.

MAST, uncultured marine stramenopile lineages

Marine stramenopiles (MAST) were first detected as 18S rDNA sequences retrieved from the marine environment and without a clear phylogenetic placement. They form more than 10 clades at the basal part of the stramenopiles (Massana et al 2004b), where all protists are heterotrophic, including free-living phagotrophic flagellates (bicosoecids), parasites (blastocystis), or osmotrophs (oomycetes and labyrinthulids) (figure I.4). MAST are widely recurrent in molecular surveys, occurring in the five world oceans, and most sequences affiliate with a few clades (MAST-1, MAST-3, MAST-4 and MAST-7). The heterotrophic nature of MAST, first suspected by their phylogenetic placement, was confirmed by FISH (Fluorescent In Situ Hybridization) for clade-1, clade-2 and clade-4 (Massana et al 2006b), and clade-6 (Piwosz and Pernthaler 2010). MAST cells from these



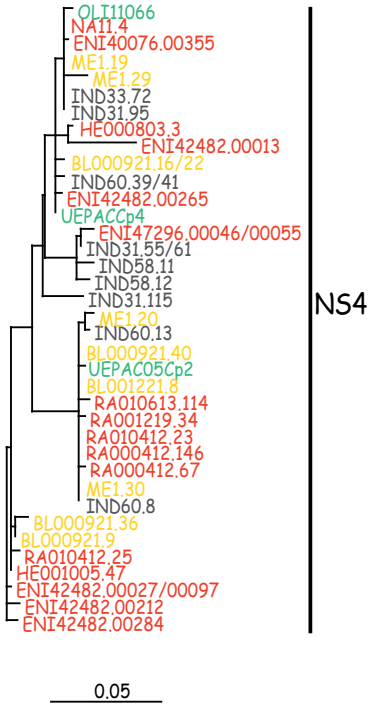


Figure I.5. Phylogenetic tree with partial 18S rDNA sequences of MAST-4. Each color identifies sequences from a different region (Atlantic: red; Pacific: green; Indian: grey; Mediterranean: yellow). The black vertical line shows the coverage of the FISH probe NS4. The scale bar indicates 0.05 substitutions per position. Figure taken from (Massana et al 2006b).

Figure I.4. Phylogenetic position of marine stramenopiles (MAST) within the stramenopile supergroup. Tree with complete 18S rDNA sequences [modified from (Massana et al 2004b)] showing the positions of MAST lineages (red boxes) among cultured phototrophic (green box) and heterotrophic (grey boxes) groups.

groups are small protists (2-8 µm in size), able to grow in the dark and to ingest bacteria. In addition, they are quite abundant in the marine plankton and account for a significant fraction of HF globally (up to 35%).

One group in particular, MAST-4, is found in all samples (except the polar ones) (figure I.5 and I.6). It is a very small protist (2-3 µm in size), so it qualifies as picoeukaryote. Its abundance averages 130 cells ml^{-1} and accounts for 9% of heterotrophic protists in a wide range of marine systems (Massana et al 2006b). This group shows a consistent size in all samples analyzed for a wide range of temperatures (from 5 to 28ºC), not following the "temperature size rule" of decreasing body size with increasing temperature (Atkinson et al 2003). They were determined as HF because of their fast growth in the dark, the absence of chloroplast and the observation of food vacuoles containing bacteria (Massana et al 2006a) (figure I.7). Moreover, a single flagellum was observed. Due to their widespread distribution and global abundance, it is likely that MAST-4 cells contribute substantially to marine food webs in vast areas of the oceans. Concerted efforts without success until now are being made to obtain a representative in culture. It is remarkable that still-uncultured groups can be dominant in the oceans, highlighting the ecological relevance of the novel diversity detected by the molecular approach.

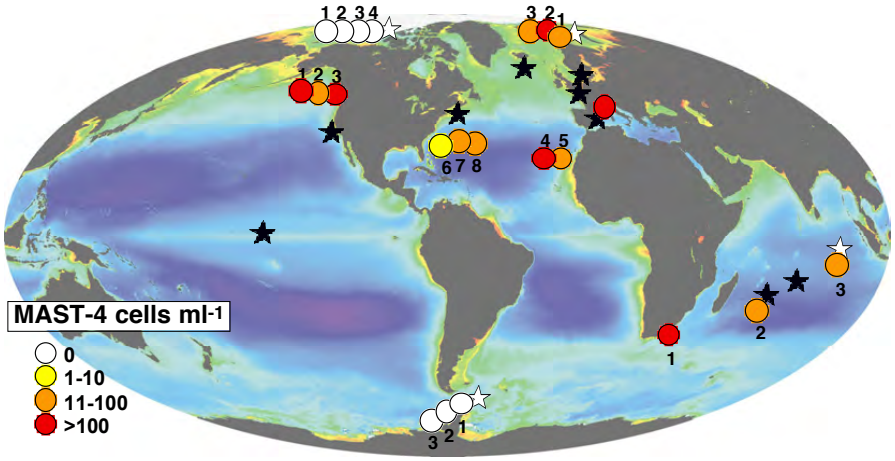


Figure I.6. Global distribution and abundance of MAST-4 cells in the world oceans. Stars indicate sites where 18S rDNA clone libraries have been constructed, black if the library contains MAST sequences and white if it does not. Dots indicate sites where FISH counts have been performed, in different color depending on the recorded cell abundance. Figure taken from (Massana et al 2006b).

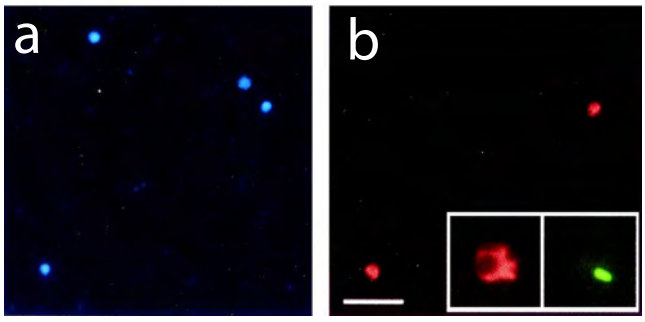


Figure I.7. Epifluorescence micrographs of MAST-4 cells. (a) DAPI-stained cells and the corresponding microscopic field (b) showing that two of the four eukaryotes were MAST-4 cells after FISH (compare a and b). The scale bar is 10 µm. The nuclear region, brightest by DAPI staining, is dimmer by FISH fluorescence, consistent with the cytoplasmic localization of ribosomes. The inset in panel b shows a MAST-4 cell (magnified 3 times) with one ingested FLB. Taken from (Massana et al 2002).

Summary of MAST-4 characteristics:

- Unicellular eukaryote (protist)
- Stramenopile
- Picoeukaryote: cell size below 3 μm
- Marine heterotrophic flagellate
- Bacteria grazer
- Globally distributed, except in polar systems
- Abundant in marine ecosystems (~10% of heterotrophic flagellates)
- Still uncultured

The importance of marine picoeukaryotes

The existence of microbial life suspended in the marine water column has been known for a long time, but only the last few decades have brought an appreciation of the extent of their biological and ecological significance. Most knowledge of marine protists has been confined to larger microscopically recognizable taxa, such as microalgae, ciliates, and larger flagellates. In contrast, marine picoeukaryotes are largely indistinguishable by light microscopy. They form an assemblage of small inconspicuous cells, just slightly larger than marine bacteria (Massana 2011). Phototrophic picoeukaryotes (pigmented cells) are important primary producers that are at the base of food webs. Heterotrophic picoeukaryotes (colorless cells) are mostly bacterivores and play key roles in channeling bacteria to higher trophic levels as well as in nutrient recycling. Mixotrophy and parasitism are relevant but less investigated trophic paths. Yet, so far only a few picoeukaryotes have been isolated and characterized and they remain largely undescribed.

The eukaryotic picoplankton is a heterogeneous collection of small protists with a diameter ranging from 0.8 μm in the case of *Ostreococcus tauri*, the smallest known eukaryote, to an upper range of 2-3 μm. In 1978 a scheme for classification of marine organisms according to size was delineated largely based on sieving technology. Microorganisms were operationally split into three categories: picoplankton (0.2-2 μm in cell diameter), nanoplankton (2-20 μm) and microplankton (20-200 μm). Initially, the picoplankton was thought to be almost exclusively made up of prokaryotes and the nanoplankton mostly of small single-celled eukaryotes. However, the existence and abundance of protists within the picoplankton size class was soon recognized. Today, the term picoeukaryotes is often used a bit loosely to include protists with a size up to 3 μm (Vaulot et al 2008). Direct

inspections of protist assemblages indicate that the 2 μm limit often falls in the middle of the size spectra and that a more coherent group is delimited using a 3 μm upper boundary (Massana 2011). Many picoeukaryotes, both phototrophic (such as *Micromonas pusilla*) and heterotrophic (such as the MAST-4), are flagellated cells (Patterson and Larsen 1991). Thus, a large proportion of the assemblages known as phototrophic and heterotrophic flagellates, would qualify as picoeukaryotes. So, picoeukaryotes would be very important both as primary producers and as bacterial grazers in aquatic ecosystems. We know now that picoeukaryotes are ubiquitous throughout the marine environment, populating surface oceans at abundances around 1000 cells ml^{-1}. Picoeukaryotes are convincingly integral members of marine ecosystems in terms of cell abundance, biomass, activity and diversity and play crucial roles in food webs and biogeochemical cycles (Massana 2011).

Picoeukaryotes have a typical eukaryotic cell structure in a miniaturized state. Due to their small size, they are desirable models for genome sequencing projects, particularly to search for the genetic bases of cell miniaturization and ecological success. Their small size has implications for metabolic rates and the trophic transfer through microbial food webs. Although some examples of cultured picoeukaryotes exist, most, as the MAST-4 group, have only been detected in the last years using culture-independent techniques and therefore are poorly characterized. Molecular surveys of picoeukaryotes have unveiled a large phylogenetic diversity and new lineages, and it is critical to understand the ecological an evolutionary significance of this large and novel diversity. A main goal is to assess how individuals are organized in taxonomic units and how do they participate in ecological processes.

In short, picoeukaryotes defined as protists smaller than 3 μm, are abundant and ecologically important in planktonic marine ecosystems. They include diverse phototrophic and heterotrophic cells, and they play crucial roles as primary producers, bacterial grazers, and parasites. In recent years their phylogenetic and functional diversity, abundance, and widespread distribution has begun to be recognized and they are attracting more attention. This thesis is about one specific uncultured lineage, the MAST-4 lineage, which can be viewed as a model heterotrophic flagellate.

Approaches for the autoecology of microbial specific lineages

The study of the ecology of a single species (or related taxa) is a discipline generally named autoecology. This includes studying the distribution, function and interactions of this species with the environment, and it is a common practice for animals and plants. This can be exemplified by the study of the community diversity and abundance of bird populations (Poirazidis et al 2011). A typical survey can start by establishing different plots of observation for monitoring, and can be done with a pair of binoculars and a good knowledge of the bird's morphology (figure I.8). If a

better register of their movements and activity is required, it is even possible to capture and mark the birds with a ring or a wing tag to distinguish this individual in future observations. Further, it is also possible to attach a radio-transmitter as a backpack to track their movements (Schindler et al 2006), as has been done with the Eurasian black vulture *(Aegypius monachus*) to study their breeding season range (Vasilakis et al 2005) and spatial distribution (Vasilakis et al 2008) in the Dadia National Park, Greece.

Figure I.8. Monitoring of the Eurasian Black Vulture *(Aegypius monachus*) in the Dadia National Park, Greece. Vultures in a feeding place (top-left). Bird observation by ornithologists (top-right). Antenna to search the signal of the transmitter by telemetry (bottom-left). Vultures with ring and wing tag (Pictures courtesy of J. Elorriaga) (bottom-right).

This gets much more complicated, even impossible, in the microbial world. The basic limitation is that the characters observed with classical microscopy for the target organisms are not enough to identify them. This happens with prokaryotes as well as with picoeukaryotes. Using epifluorescence microscopy with a DAPI (4,6-diamidino-2-phenylindole) stain we can obtain

quantitative data of phototrophic (pigmented) or heterotrophic (colorless) cells, but we cannot classify them. Within the assemblage of heterotrophic cells (figure I.9) only two groups can be clearly distinguished: the choanoflagellates by the conspicuous collar and the single flagella, and the dinoflagellates by the granulated appearance of the nucleus. One tool that facilitates the study is to obtain them in pure culture, but nowadays we know that a large number of the dominant microorganisms in natural communities are not yet cultured. Examining them is very important to have a better knowledge of nature (they are the most abundant) but this is difficult because of the absence of cultures. Particularly, the MAST-4 (aim of study in this thesis) is one of these uncultured groups. Therefore, we need original approaches, independent of cultures and direct microscopy, to study them in environmental samples. Next we list three different approaches for the environmental study of microbial specific lineages.

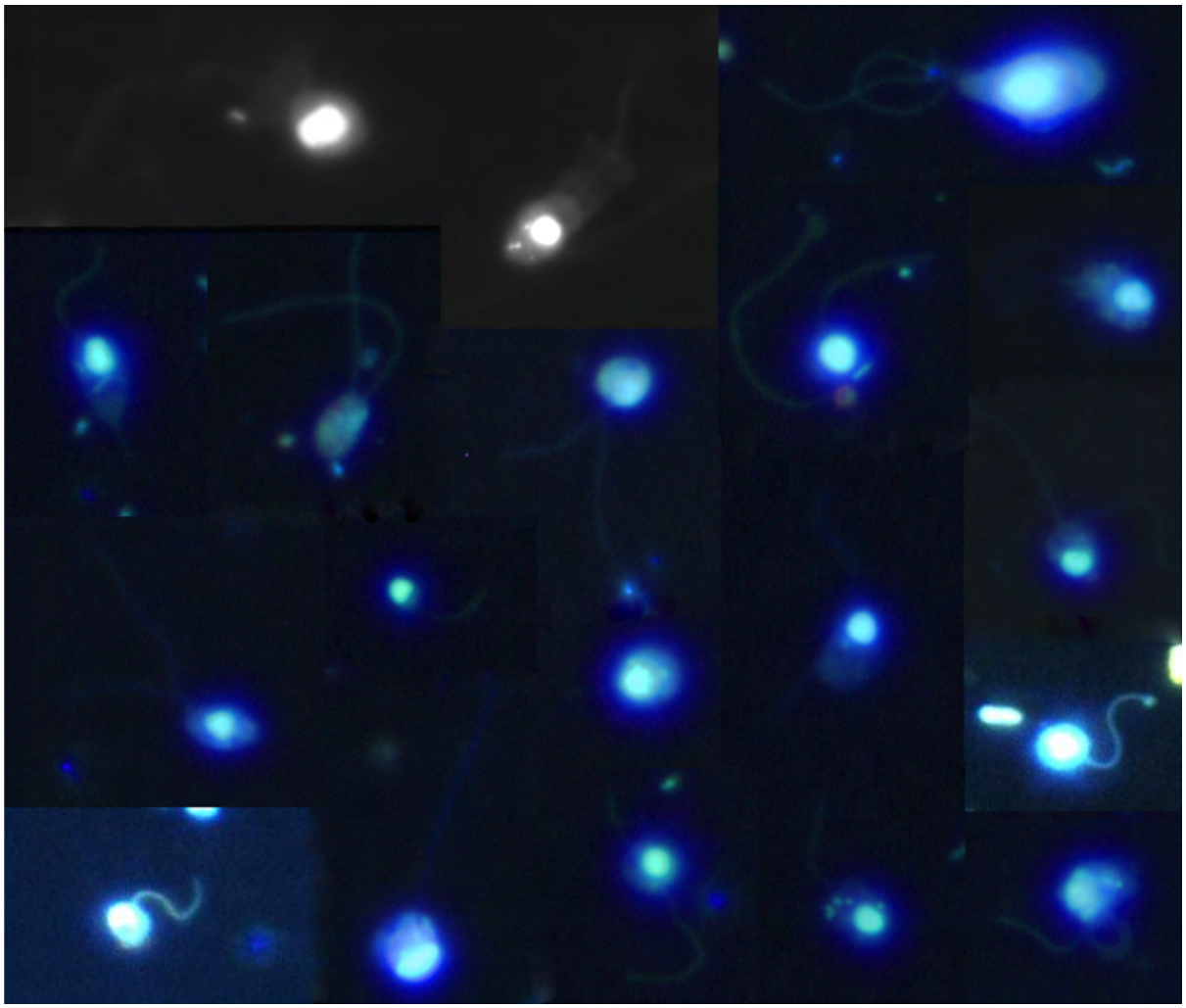


Figure I.9. Epifluorescence micrographs of heterotrophic flagellates DNA-stained with DAPI and observed by UV excitation. Pictures courtesy of R. Massana and I. Forn.

Specific targeting of the ribosomal small subunit genes

During the last years a systematic way to sequence the conserved genes of the ribosomal subunits of all organisms has been developed, opening different routes for the molecular characterization of natural microbial communities (Amann et al 1995). A general flow (figure I.10) consists in a nucleic acid extraction of the total natural community, directly or after cell enrichment. With the advent of the polymerase chain reaction (PCR) (White et al 1989), rRNA gene fragments can be selectively amplified from the mixed environmental DNA and cloned. Clone libraries containing defined fragments can be then sequenced. These sequences are archived in public

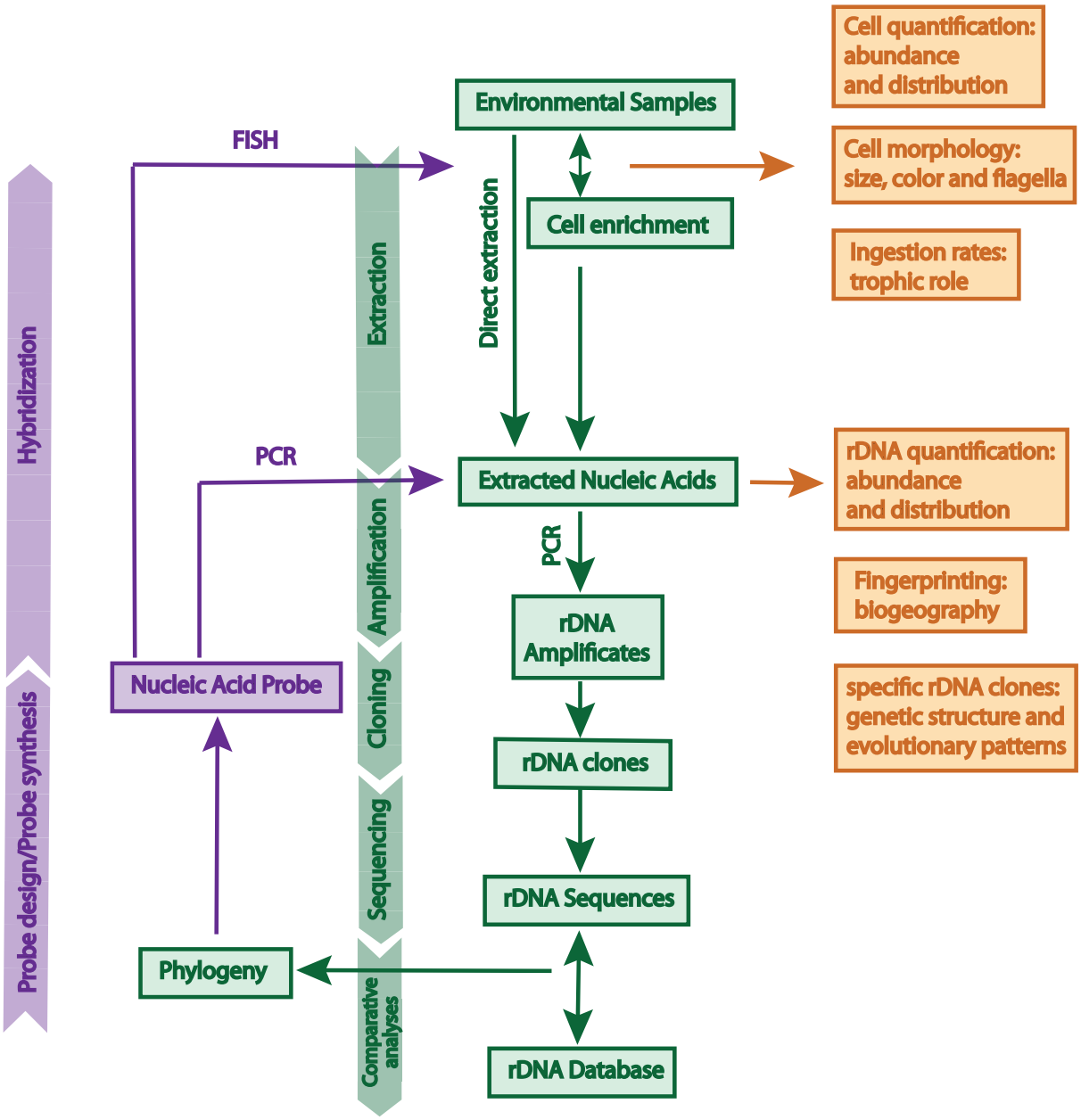


Figure I.10. Flow chart showing the way used in this thesis to characterize environmental samples by comparative rDNA sequence analysis (green), the different options used to target the rDNA of specific lineages (purple) and the results obtained (orange). Figure adapted from (Amann et al 1995).

databases like Genbank at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genbank/), or in sites specifically created for rDNA sequences, such as the Silva database of the software ARB (http://www.arb-silva.de/). Public databases are readily accessible for comparative analysis. Then, we can display environmental sequences in phylogenetic trees and design specific probes or primer sets to a chosen discrete group. Specific microbial taxa can then be detected with an epifluorescence microscope by the FISH technique, or by amplifying its rDNA with the PCR step required for several molecular techniques. The critic requirement is to be sure that the probes or primers are specific for the target microorganisms and that there are no false positives from environmental samples. Furthermore, it is important to keep in mind the number of steps involved in the retrieval of rDNA sequences and be careful not to regard them as unbiased reflections of natural communities. For instance, the unequal rDNA operon copy number, or the uneven amplification efficiency of particular groups, can severely bias the PCR products. Moreover, PCR can retrieve unrealistic sequences derived from for contamination or chimeras.

For the study of the specific lineages, the FISH technique (Pernthaler et al 2001) allows us to detect the target microbial taxa, measure and quantify it in natural samples, and even observe the ingestion of different preys. Using PCR, we can sequence its genes to assess its diversity and genetic structure or we can compare the taxa intradiversity in different samples via fingerprinting. When using quantitative PCR, the absence of cultures is a challenge for preparing a standard for quantification. Typically, when a culture is available, the standard is prepared by relating the added DNA and the number of cells. In this thesis, this problem was solved cloning plasmids with the 18S rDNA insert of the target taxa. Thus, standards with serial dilutions of 18S rDNA molecules can be done (figure I.11). Values thus obtained are not directly translated to cell numbers because eukaryotes have multiple copies of the rDNA operon but they are very useful for quantification.

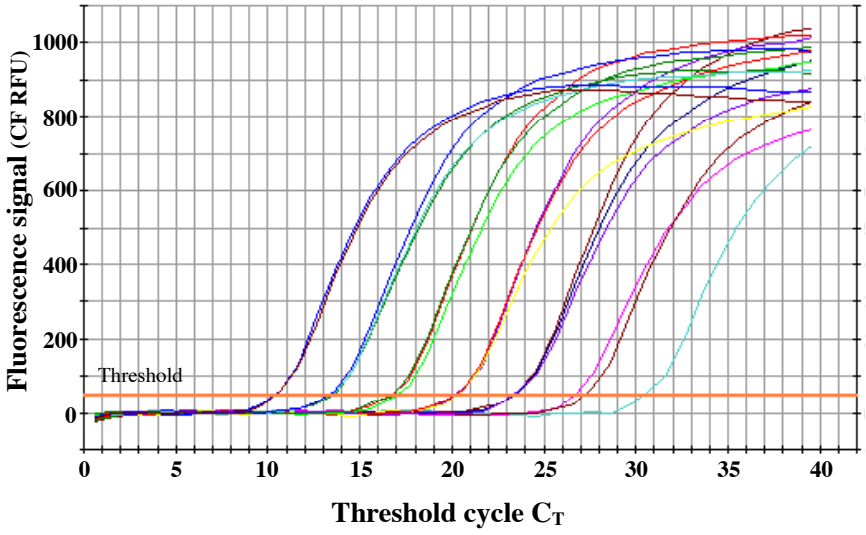


Figure I.11. Serial dilutions of the plasmid (from 10^8 to 10^2 copies) used as standard in quantitativePCR. The good performance of the standard is shown by the high correlation coefficient (0.999) and a PCR efficiency close to 100% (99.2%).

Unamended seawater incubations

The advantages of having pure cultures are that morphological and functional parameters can be estimated, such as ultrastructure by electron microscopy, functional and numerical responses, food size spectra, growth efficiency, temperature optima, and survival responses. In the absence of pure cultures for many heterotrophic flagellates, a compromise to obtain some of these parameters is the use of unamended seawater incubations (Massana et al 2006a), known to promote the growth of uncultured HF in mixed assemblages. The setup includes a prefiltration through a 3 µm filter to remove larger predators, and a dark incubation that prevents the growth of phototrophs. During the incubation, bacteria initially increase a few times (figure I.12) followed by increasing numbers of HF a few days later coincident with the decrease in bacterial numbers. Photosynthetic flagellates and *Synechococcus* decrease continuously in numbers during the incubations, owing to dark conditions and perhaps to predation by heterotrophic flagellates. The protistan assemblage change from one dominated by phototrophic cells to one dominated by heterotrophic cells. In most samples, this simple setup results in the growth of several MAST groups (Massana et al 2006a). The growth of these abundant, uncultured HF was probably because in the unamended incubations bacteria are kept at realistic abundances and sizes (del Campo 2011). The increase of flagellates was moderate (10-100 fold), sufficient to measure growth rates of uncultured groups and to provide excellent material for activity measurements. Although short-lived, these events provide interesting phenotypic and functional information on uncultured protists. Thus, unamended seawater incubations can select for HF abundant *in situ* but not yet isolated in pure culture. This simple approach was used in several chapters of this thesis.

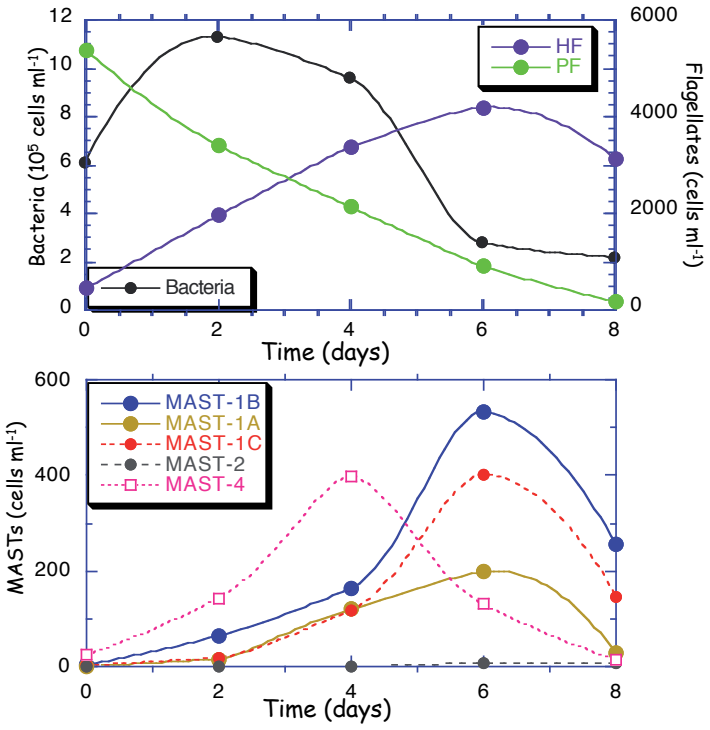
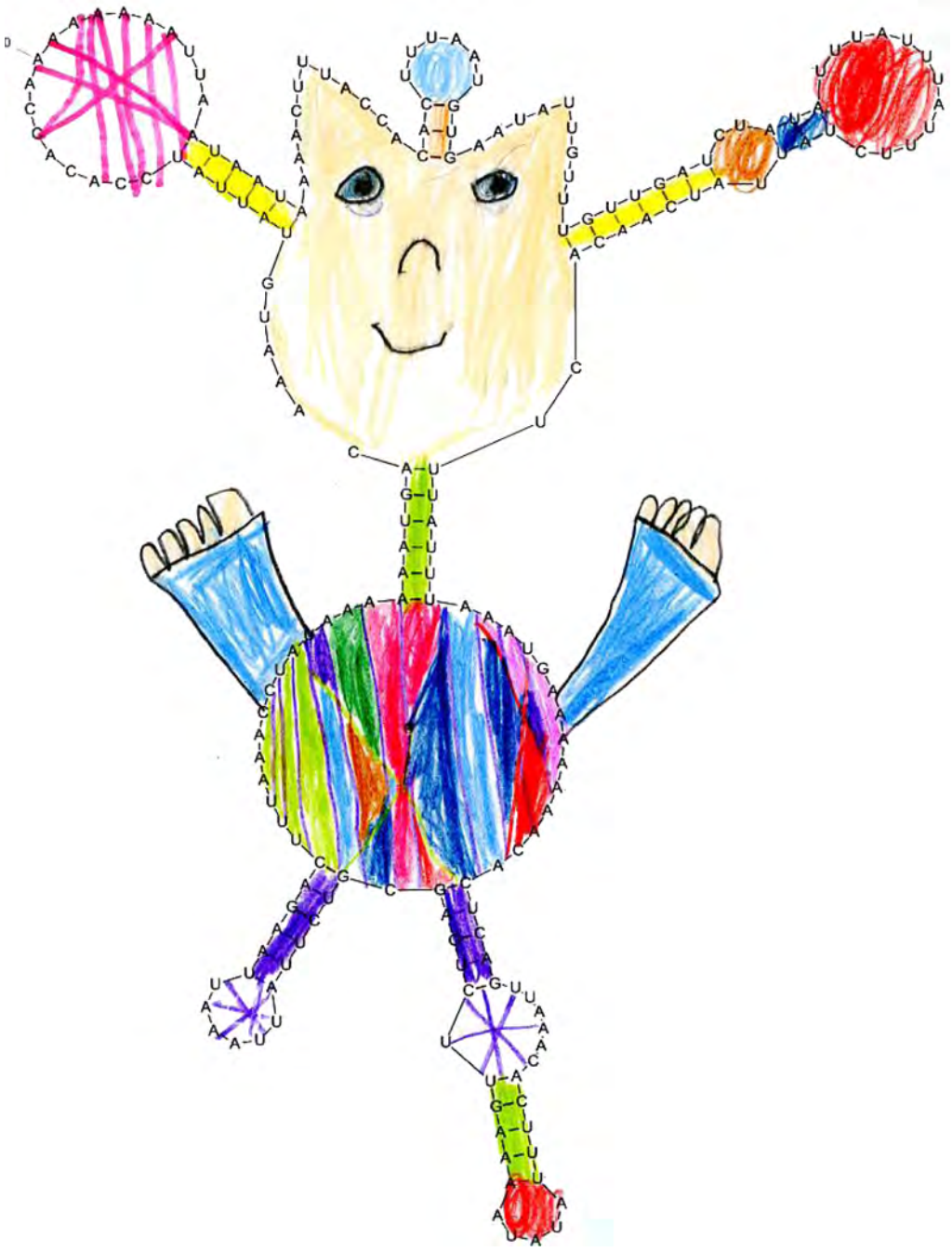


Figure I.12. Dynamics of microbial components in an unamended seawater incubation performed in the Norwegian Sea (from (Massana et al 2006a). Upper panel: Cell abundance estimated by DAPI counts of bacteria (black), heterotrophic (purple) and phototrophic flagellates (green). Lower panel: Cell abundance estimated by FISH of five MAST groups.

Single amplified genomes (SAGs)

Recently, there have been exciting advances in single-cell analyses. The sorting capacities of modern flow cytometers, combined with the use of lysotraker, a green fluorescing probe that stains food vacuoles (Rose et al 2004), are opening new avenues in microbial ecology. Single microbial cells can then be used as inoculums to start pure cultures, or as template for whole genomic amplification prior to genome sequencing (Yoon et al 2011). This approach has been recently applied to sort HF from marine assemblages (Heywood et al 2011, Rose et al 2004). Assessing the diversity of heterotrophic protists based on single cell sorting, whole genome amplification and rDNA sequencing is better than that given by community surveys, since the number of rDNA copies per cell is not a problem anymore. Moreover, SAGs are the only way to access to genomic information of uncultured cells, and might enable to analyze ecological interactions (grazing, symbiosis) between protists and prokaryotes (Martínez-García et al 2012). Preparing and analyzing SAGs from MAST-4 seems to be a promising approach to complement the study presented in this thesis.

Aims and Outline of the Thesis



The general goal of the thesis was to study the ecology of a relevant uncultured heterotrophic flagellate taxon, the MAST-4 lineage. This protist is an important and widespread picoeukaryote in marine systems, and represents a measurable fraction of the heterotrophic flagellate assemblage. Moreover, it has the advantage to be readily enriched in unamended incubations and easily detected with molecular tools.

Heterotrophic flagellates (HF) are routinely quantified by epifluorescence microscopy after DAPI staining (Porter and Feig 1980), but this reveals few morphological features, so they remain generally unidentified. With the appearance of molecular surveys, oligonucleotide probes against several MAST lineages have been designed (Massana et al 2002, Massana et al 2006b) and used by FISH to identify them. Further studies of their distribution and abundance revealed that they were globally distributed, and a single group, MAST-4, contributed to ~9% of HF in surface marine systems (Massana et al 2006b). So far, MAST-4 has been quantified by FISH, which is very reliable but time-consuming to process the large number of samples generated during oceanographic cruises. The first aim of this thesis **(chapter 1)** was to develop a fast and sensitive technique to assess the abundance and distribution of the uncultured heterotrophic flagellate MAST-4 based on the real-time quantitative polymerase chain reaction (Q-PCR) detection of its 18S rRNA genes.

Bacterial grazing is of fundamental importance in aquatic ecosystems and is carried out mostly by small flagellated protists up to 5 μm in diameter (Sherr and Sherr 2002). It controls bacterial abundances in a wide range of ecosystem conditions, channels organic carbon to higher trophic levels, and releases inorganic nutrients that often are controlling primary production (Jürgens and Massana 2008, Pernthaler 2005). The second aim of the thesis was to study the grazing rates and prey preferences **(chapter 2)** and the functional responses **(chapter 3)** of uncultured HF living in natural assemblages, including the MAST-4. This part is based on the estimation of the feeding activity of specific grazers detected by FISH after short-term ingestion experiments with tracer preys, which are then counted inside the protist food vacuoles.

Microbes have vital roles for the functioning of the biosphere (Falkowski et al 2008), but currently we are far from having acceptable estimates of their diversity. Furthermore, it is unclear how microbial diversity is distributed in space and time, and how diversity ranks are translated into ecologically meaningful interactions or processes. The marine protists of very small size, the picoeukaryotes, are among the underexplored microbes with large ecological importance (Massana 2011). The third aim of this thesis **(Chapter 4)** was to understand the genetic structure and evolutionary patterns of the MAST-4 picoeukaryote. It was based on sequencing a large fragment of the rDNA operon and investigating ITS (Internal Transcribed Spacer) secondary structures to explore possible sexual boundaries among related types.

Biogeography is the study of the distribution of biodiversity over space and time. The current evidence confirms that environmental selection is fundamental for the spatial variation in microbial diversity (Martiny et al 2006). The next frontier is to figure out whether these patterns are also influenced by geographical barriers that facilitate evolution and diversification. Contradictory results have been obtained in the last years, with no universal picture emerging, partly because the answer may depend on the particular situation analyzed. The last aim of this thesis **(chapter 5)** was to study marine protist biogeography using the MAST-4 as a model. Its community structure and distribution was assessed by combining automated ribosomal intergenic spacer analysis (ARISA) (Fisher and Triplett 1999) and 18S-ITS1 gene libraries.

The outline of the different topics studied is:

• *Objective 1: Abundance and distribution*

Chapter 1 "Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR"

We developed a Q-PCR protocol to determine rapidly the abundance of this group using environmental DNA. We designed a primer set targeting the 18S rRNA genes of MAST-4 and optimized and calibrated the Q-PCR protocol using a plasmid with the target sequence as insert. The Q-PCR was applied to quantify MAST-4 along three transects, longitudinal in the Indian Ocean, latitudinal in the Drake Passage and coastal–offshore in the Mediterranean Sea, and to a temporal study in a Mediterranean Sea coastal station.

• *Objective 2: Trophic role*

Chapter 2 "Grazing rates and functional diversity of uncultured heterotrophic flagellates"

Here we measured grazing rates of uncultured protists in natural assemblages (detected by FISH), and investigated their prey preference over several bacterial preys in short-term ingestion experiments. These included fluorescently labeled bacteria (FLB) and two strains of the Rhodobacteraceae and Flavobacteriaceae families, of various cell sizes, which were offered alive and detected by catalyzed reporter deposition-FISH after the ingestion. We obtained grazing rates of MAST-4 and MAST-1C flagellates.

Chapter 3 "Functional responses of three heterotrophic flagellates taxa in mixed natural assemblages"

Here we determined the functional response (maximum ingestion rate and half-saturation constant) of three heterotrophic flagellates taxa (MAST-4, *Minorisa minuta* candidatus and *Paraphysomonas* sp.) and of the total community from a mixed natural assemblage. We used fluorescence labeled bacteria added at different final abundance (from 10^5 to 10^7 cells ml^{-1}) and counted them inside the

protist food vacuoles.

• *Objective 3: Genetic structure and evolutionary patterns*

Chapter 4 "Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4)"

In this study, we investigated the diversity of MAST-4, aiming to assess its limits and structure. We used rDNA sequences obtained here (both pyrosequencing reads and clones with large rDNA operon coverage), complemented with GenBank sequences. Conserved regions of the ITS1 and ITS2 secondary structures were evaluated for delineating different biological species.

• *Objective 4: Biogeography*

Chapter 5 "Biogeography of the uncultured marine picoeukaryote MAST-4: temperature driven distribution patterns"

We studied the biogeography of MAST-4 by combining ARISA fingerprints and gene libraries of the ITS1 region. This study addresses this question by examining both spatial and temporal trends in MAST-4 assemblages and associated environmental factors.

“And so it was indeed: she was now only ten inches high, and her face brightened up at the thought that she was now the right size for going through the little door into that lovely garden.”

Lewis Carroll (1865)

Chapter 1

Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR

Rodríguez-Martínez R, Labrenz M, Del Campo J, Forn I, Jürgens K, Massana R (2009). Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environmental Microbiology* **11:** 397-408.

Environmental Microbiology (2009) **11**(2), 397–408 doi:10.1111/j.1462-2920.2008.01779.x

Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR

Raquel Rodríguez-Martínez,[1]* Matthias Labrenz,[2] Javier del Campo,[1] Irene Forn,[1] Klaus Jürgens[2] and Ramon Massana[1]**

[1]*Institut de Ciències del Mar, CMIMA (CSIC). Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain.*

[2]*Leibniz Institute for Baltic Sea Research, Seestrasse 15, 18119 Rostock-Warnemünde, Germany.*

Summary

Molecular surveys of marine picoeukaryotes have revealed a large number of sequences unrelated to cultured organisms, such as those forming the marine stramenopile (MAST)-4 clade. Recent FISH (fluorescent *in situ* hybridization) data have shown that MAST-4 cells are uncultured heterotrophic flagellates of 2–3 μm in size that have a global distribution in non-polar marine waters. However, FISH is time-consuming and hard to apply to the many samples generated during oceanographic cruises, so we developed a real-time quantitative polymerase chain reaction (Q-PCR) protocol to determine rapidly the abundance of this group using environmental DNA. We designed a primer set targeting the 18S rRNA genes (rDNA) of MAST-4 and optimized and calibrated the Q-PCR protocol using a plasmid with the target sequence as insert. The Q-PCR was then applied to quantify MAST-4 rDNA molecules along three marine transects, longitudinal in the Indian Ocean, latitudinal in the Drake Passage and coastal–offshore in the Mediterranean Sea, and to a temporal study in a Mediterranean Sea coastal station. MAST-4 was detected in all samples processed (averaged abundances between 500 and 1000 rDNA molecules ml^{-1}) except in mesopelagic and Antarctic samples, where it was virtually absent. In general, it was more abundant in the coast than offshore and in the deep chlorophyll maximum than at surface. A comparison of Q-PCR and FISH signals in well-controlled microbial incubations indicated that MAST-4 cells have around 30 copies of the rDNA operon. This Q-PCR assay quickly yielded quantitative data of uncultured MAST-4 cells and confirmed their wide distribution and putative ecological importance.

Received 20 May, 2008; accepted 17 August, 2008. For correspondence. E-mails *raquelr@icm.csic.es; **ramonm@icm.csic.es; Tel. (+34) 93 2309500; Fax (+34) 93 2309555.

Introduction

Marine heterotrophic flagellates are small unpigmented protists that are the main picoplankton grazers and nutrient remineralizers in aquatic ecosystems (Fenchel, 1986). Together with viruses (Suttle, 2005), they maintain picoplankton populations at relatively stable abundances in seawater (Pernthaler, 2005). Heterotrophic flagellates are routinely quantified by epifluorescence microscopy after DAPI (4,6-diamidino-2-phenylindole) staining (Porter and Feig, 1980), but this reveals few morphological features, so they remain generally unidentified. Recent molecular surveys are providing new insights into the phylogenetic affiliation of these minute protists. Clone libraries of 18S rDNA from marine eukaryotic picoplankton have revealed a large *in situ* diversity and the existence of novel groups unrelated to cultured protists (Díez *et al*., 2001; López-García *et al*., 2001; Moon-van der Staay *et al*., 2001). Among them, marine stramenopiles (MAST) appear recurrently and include more than 10 different phylogenetic lineages affiliating among the basal stramenopile groups (Massana *et al*., 2004). Oligonucleotide probes against several MAST lineages have been designed (Massana *et al*., 2002; 2006a) and used by FISH (fluorescence *in situ* hybridization) to identify their morphology and infer their trophic mode. The MAST cells resulted to be heterotrophic flagellates of various sizes (from 2 to 8 μm), able to grow in the dark and graze bacteria. Further studies of their distribution and abundance revealed that they were globally distributed, and a single group, MAST-4, contributed to 9% of heterotrophic flagellates in all surface marine systems except polar waters (Massana *et al*., 2006a).

Thus, MAST-4 is an abundant protistan group in marine ecosystems and likely contributes to picoplankton grazing and nutrient remineralization. Concerted efforts without success until now are being made to obtain a represen-

tative in culture. So far, MAST-4 has been quantified by FISH, which is very reliable but too time-consuming to process the large number of samples generated during oceanographic cruises. For this reason, we developed a quantitative polymerase chain reaction (Q-PCR) assay, based on the 18S rDNA, to quantify this protistan group in marine samples. This technique allows an accurate estimation of the copy number of a target gene in a given DNA extract by measuring the increase of its PCR product in real time. Three different detection systems are generally used: the SYBR Green I stain that binds unspecifically to double-stranded DNA (Schneeberger *et al*., 1995) and the more specific 'Taqman' and 'Molecular Beacon' technologies (Walker, 2002). The Q-PCR was initially developed for clinical applications (Heid *et al*., 1996), and recently has been applied to marine research. Several studies have dealt with planktonic prokaryotes, targeting broad bacterial groups (Suzuki *et al*., 2000), an uncultured bacteria (Labrenz *et al*., 2004), and several ecotypes of marine picocyanobacteria (Becker *et al*., 2002; Ahlgren *et al*., 2006; Johnson *et al*., 2006). Other Q-PCR studies quantified marine protists, always targeting cultured organisms: the prasinophytes *Bathycoccus*, *Micromonas* and *Ostreococcus* (Zhu *et al*., 2005; Marie *et al*., 2006; Countway and Caron, 2006), the dinoflagellates *Pfiesteria piscicida* (Bowers *et al*., 2000) and *Alexandrium minutum* (Galluzzi *et al*., 2004) and the alveolate *Perkinsus marinus* (Audemard *et al*., 2004).

The aim of this study was to develop a fast and sensitive technique to assess the abundance of the uncultured heterotrophic flagellate MAST-4 in large sets of environmental samples, based on the Q-PCR detection of its 18S rRNA gene. This technique takes advantage of the DNA extracts stored in our lab and collected in oceanographic cruises done during the last 10 years. One problem when targeting uncultured protists is the variable copy number of the rDNA operon in different species (Prokopowich *et al*., 2003; Zhu *et al*., 2005). To estimate this value, we prepared unamended seawater incubations (Massana *et al*., 2006b), where we expected to find varying amounts of MAST-4 cells, and sampled them carefully to compare Q-PCR and FISH results in the best conditions. To our knowledge, this is the first application of Q-PCR to assess the distribution and abundance of an uncultured protist and provides a general view of its broad distribution.

Results

Optimization of Q-PCR for MAST-4

DNA cleaning for optimal amplification. The different DNA extracts (from oceanic cruises, incubations or plasmids) were used as templates for Q-PCR with the specific 18S rDNA MAST-4 primer set. Amplifications were positive with the plasmid but initially negative with environmental samples. In order to eliminate potential PCR inhibitors, DNA extracts were further purified with an ethanol precipitation step. Virtually all environmental samples yielded positive amplifications after this cleaning step. To check for potential DNA loss during precipitation, the plasmid was processed before and after precipitation (two replicates from seven dilutions), and no significant differences in the quantification were seen (data not shown).

Q-PCR efficiency and melting curve. After processing the DNA extracts with the cleaning step, the efficiencies of each Q-PCR run were calculated by comparing the signal in the different dilutions of the standard (plasmid) and the relative standard (one environmental sample). Both efficiencies were always close to 100% (98% on average; range of 93–101% in the standard and of 94–102% in the relative standard). Moreover, in the same run the efficiencies of both standards were highly similar (0.8% difference on average, 3.2% being the maximal difference). Thus, the plasmid standard could be used to quantify MAST-4 sequences in environmental samples. To confirm that the measured PCR product derives from the target sequence, a melting curve was done after the last cycle in all tubes. A single peak at 84°C appears in all samples, implying that the fluorescence readings derive integrally from the specific PCR product. Primer-dimer signals only appeared in some tubes with milliQ water or with environmental samples that did not amplify. Non-specific PCR products [melting temperature (Tm) = 89°C] were only observed in the coldest samples from the Drake passage and were considered negative. In fact, the size of this PCR product was around 700 bp (as measured in an agarose gel), confirming that they were clearly non-target amplicons.

DNA storage and thawing tests. Some of the DNA extracts processed by Q-PCR were kept at –80°C during almost 10 years and defrost several times for particular analysis, with the risk of an unknown level of DNA degradation. The possible effect of freeze–thaw cycles on DNA degradation and Q-PCR results was studied on three samples, the plasmid and two environmental samples. These were thawed, an aliquot taken for Q-PCR, frozen again at –80°C and the process repeated several times. All samples were then processed in the same Q-PCR run (Fig. 1). Data were arctan-transformed, tested for normality (Kolmogorov–Smirnov test) and heteroscedasticity (Bartlett test), and then processed using a repeated measures ANOVA with the STATISTICA software (version 5.1; StatSoft). The two environmental samples did not show any sign of DNA degradation, as the number of MAST-4 molecules was always similar even when defrosted

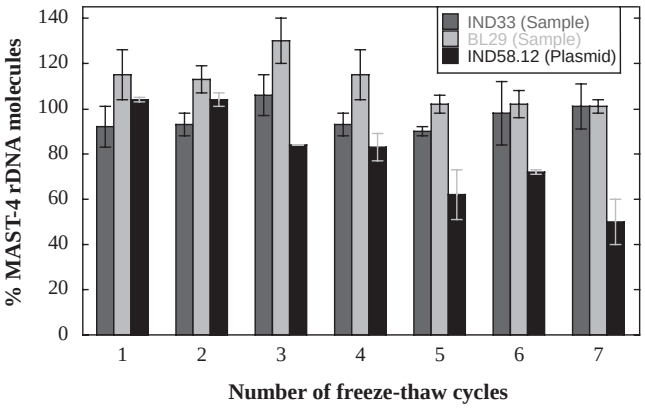


Fig. 1. Effect of freeze/thaw cycles on DNA recovery of the standard (plasmid IND58.12) and environmental samples from the Indian Ocean (IND33) and Blanes Bay (BL29). Shown are the mean (and SE; $n = 3$) of the percentage of MAST-4 rDNA molecules assessed by Q-PCR after several cycles relative to the initial quantification.

seven times (IND33, $F = 0.45$, $p = 0.89$; BL29, $F = 1.77$, $p = 0.16$). On the other hand, DNA degradation occurred in the plasmid in the last cycle (IND58.12, $F = 7.36$, $p = 0.006$; the *post hoc* test Scheffe detected significant differences between the initial cycle and the last cycle). To prevent a biased calibration the plasmid was never used as standard after being defrosted twice.

Specificity of the primer set. The specificity of the MAST-4 primer set was assessed with 18S rDNA clones from the Indian Ocean libraries, including two MAST-4 (with one mismatch with primer M41f) and seven non-target clones representative of the groups with highest clonal abundance (dinoflagellates: 14%; marine alveolate-I: 25%; marine alveolate-II: 11%; MAST: 11%; radiolaria: 15%) (Not *et al*., 2008). The pelagophyte clone was included because it presented the lowest number of mismatches with a given primer (two with the reverse primer). These clones were compared with the standard (perfect match with both primers), by adding the same number of rDNA molecules (10^8 copies) and quantifying the MAST-4 molecules after the Q-PCR run. The positive clones with one mismatch amplified close to 100% whereas the negative clones amplified always less than 0.01% of the molecules added and often did not amplify at all (Fig. 2), showing that the primer set was highly specific for MAST-4 sequences.

Spiking experiments. These were done to address two questions: the putative presence of PCR inhibitors in samples with negative amplification, and the putative inhibition of the PCR when increasing non-target DNA. For the first question, a known amount of target molecules (standard plasmid) was added to three environmental samples that did not amplify. In these mixtures, the number of target molecules quantified by Q-PCR was similar to the molecules added (data not shown), indicating that the negative amplification in the original samples was not due to the presence of inhibitors but to the absence of target sequences. For the second question, four different mixtures with known and variable amounts of target and non-target molecules were prepared. Again, the target molecules quantified by Q-PCR in these mixtures were similar to the molecules added (data not shown). So, the TaqDNA polymerase did not modify its efficacy depending on the proportion of target DNA.

Overall success of the method. A total of 214 samples were processed in 22 PCR runs. Most of them gave a good quantification the first time (144 samples), whereas others had to be repeated to obtain a reliable estimate (26 samples). Some gave a clear negative signal, indicating the absence of target sequences (34 samples, 22 with no amplification and 12 with an unspecific PCR product). Finally, a few samples (10 samples) appeared positive but the signal was too low to be properly quantified. This quantification success was related to the amount of target molecules in the sample. We assumed a detection limit of our protocol of 100 molecules per PCR tube (negative milliQ samples yielded between 0 and 90 molecules, 10 on average). This translates that the natural sample needs to have at least 20 molecules ml^{-1} for a proper quantification, assuming the volumes typically processed (10 l of seawater and 200 µl of DNA extract).

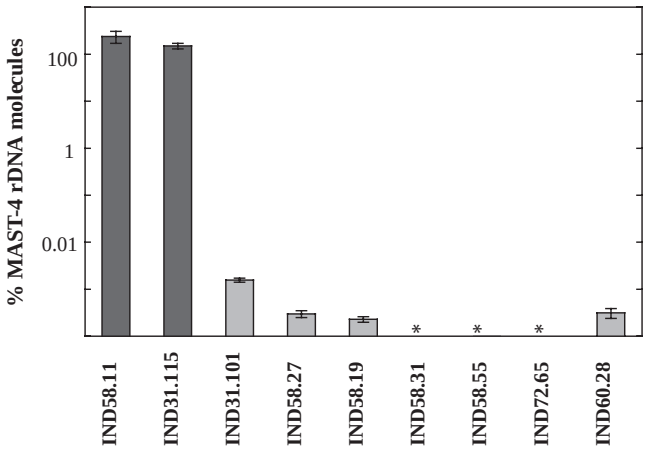


Fig. 2. Percentage (mean and SE; $n = 4$) of MAST-4 rDNA molecules (with respect to rDNA molecules added) recovered from target and non-target clones: IND58.11 (MAST-4; 1 mismatch with M41f and 0 mismatch with NS4), IND31.115 (MAST-4; 1 and 0), IND31.101 (dinoflagellate; 8 and 5), IND58.27 (marine alveolate-I; 8 and 6), IND58.19 (MAST-3; 6 and 5), IND58.31 (MAST-1C; 12 and 3), IND58.55 (marine alveolate-II; 14 and 9), IND72.65 (radiolaria; 15 and 5), IND60.28 (pelagophyte; 10 and 2). Samples with an asterisk present negative amplification.

Quantification of MAST-4 molecules in environmental samples

The MAST-4 rDNA molecules were quantified by Q-PCR using DNA extracts from three oceanographic cruises (Fig. 3A). The first cruise was a transect from South Africa to Australia crossing the Indian Ocean Gyre, where 15 stations were sampled at five to six depths (from 5 to 1000 m). The second cruise was a coastal–offshore transect crossing the northern part of the Western Alboran Gyre, where three stations (shelf, slope and deep) were sampled at five depths (from 5 to 500 m). The third cruise was a transect across the Polar Front in the Drake Passage, from South Atlantic to Antarctic waters. This transect was formed by nine stations sampled at three depths (from 5 to 60 m).

The abundance of MAST-4 rDNA molecules in the Indian Ocean was higher in the coastal samples and lower in the centre of the transect, in the most oligotrophic central gyre (Fig. 3B). Regarding the vertical gradient, MAST-4 was present all along the photic zone, being more abundant around the DCM (deep chlorophyll maximum) and with low values in the upper aphotic zone, represented here by the 200 m depth. The figure only shows data from surface to 200 m, as deeper samples (between 650 and 1000 m) generally gave negative amplification. The abundance of MAST-4 in the Alboran Sea shows a similar tendency, being more abundant at the coastal station than the offshore station (Fig. 3C). Similarly, MAST-4 also occupies the whole photic zone, being more abundant at 50–100 m depth than at surface, and being very low (250 m) or virtually absent (500 m) in the aphotic zone. Considering both cruises, 11 of the 15 deeper samples (500 to 1000 m) gave clear negative signal, and the remaining four gave unclear signals. The pattern found in the Drake Passage was remarkably different from the other two cruises (Fig. 3D). The MAST-4 molecules were only detected in the two northernmost stations, those with warmer temperatures (5.5°C and 5.8°C at the surface of stations 30 and 32 respectively). In the other stations, with temperatures below 5°C, target molecules were undetected. Here, only the upper water column was quantified, so no clear vertical pattern could be described.

Besides the three oceanographic cruises, samples from a coastal Microbial Observatory were also analysed. These were collected monthly at the surface of Blanes Bay (North-western Mediterranean), with a total of 62 samples during 6 years (Fig. 4). A clear seasonal pattern was not observed, and abrupt changes occurred between consecutive dates. What was apparent, though, was an important interannual variation, with some years (2002 and 2005) with low mean abundance (350 molecules ml^{-1}) and other years (2003 and 2004) with high mean abundance (1450 molecules ml^{-1}). The other 2 years (2001 and 2006) had an intermediate mean abundance (800 molecules ml^{-1}).

An overview of the abundances of MAST-4 in the four marine systems investigated is shown in Table 1. The average abundances of MAST-4 in the Indian Ocean, Alboran Sea and Blanes Bay, systems with comparable mean temperatures (16–24°C), were similar (between 500 and 1000 molecules ml^{-1}). MAST-4 seems to be more abundant near the coast than at the open sea (see Fig. 3B and C). Also, Table 1 shows that the abundance of MAST-4 is higher (almost twice) in the DCM as compared with the surface (5 m). The upper photic region (200–250 m) generally gives clear signals, although very low (40–50 molecules ml^{-1}), whereas the lower mesope

Table 1. Molecule and cell numbers quantified by Q-PCR and FISH respectively from the different places analysed in this study and their relationship with the temperature.

	18S rDNA molecules ml^{-1}			Cells ml^{-1}			Temperature (°C)	
	Mean	Range	*n*	Mean	Range	*n*	Mean	Range
Indian Ocean								
Above DCM	504	52–2484	28	98	82–118	5	24	21–24
DCM	864	91–2845	14	155	79–262	3	22	19–25
Upper aphotic (200 m)	48	0–285	13	0	0	3	16	8–20
Alboran Sea								
Surface (5 m)	706	469–826	3				16.5	16.5
Subsurface (50–100 m)	954	320–2205	6				15	14–16.5
Upper aphotic (250 m)	38	21–48	3					
Drake Passage								
Above SAF (5–100 m)	150	31–351	6				5.7	5.5–5.8
Below SAF (5–60 m)	0	0	20				1.3	(–1)–4.4
Blanes Bay								
Surface (5 m)	926	34–4500	62	91	18–244	26	17.6	11–26

SAF, Sub Antarctic Front.

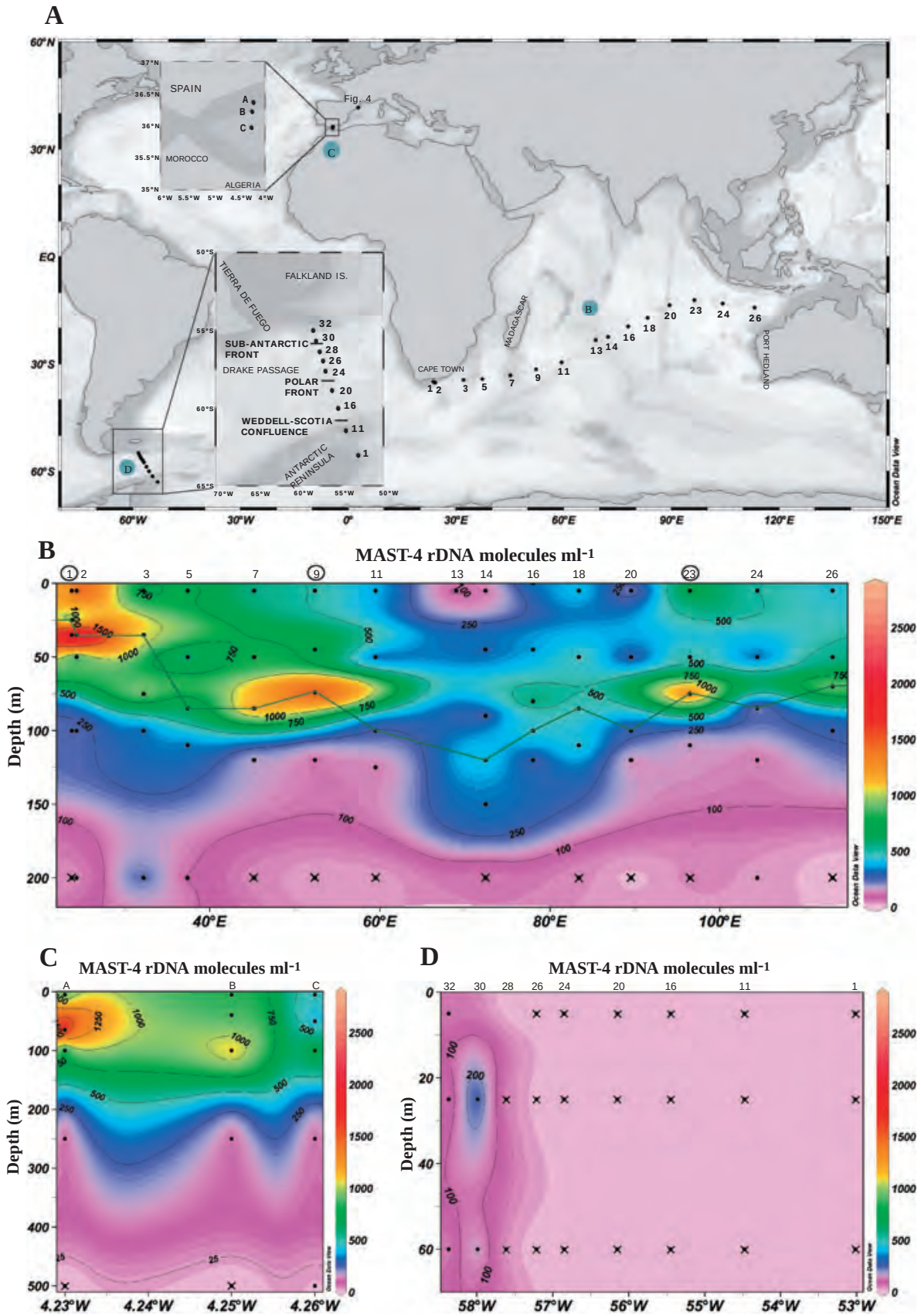


Fig. 3. Overview of the marine systems investigated (A) and abundance of MAST-4 (18S rDNA molecules ml^{-1}) at several depths in three of them: (B) Indian Ocean transect. The green line marks the DCM and stations analysed by FISH are encircled. (C) Alboran Sea transect. (D) Drake Passage transect. Samples with a cross in (B), (C) and (D) indicate negative amplification.

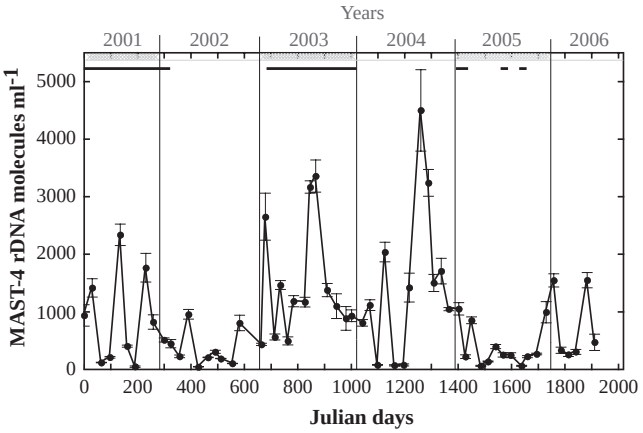


Fig. 4. Abundance (mean and SE; $n = 4$) of MAST-4 molecules in Blanes Bay during a 6-year study sampled monthly. Black lines at the top of the figure mark the periods where FISH data were also obtained.

lagic region generally gives no or very low amplification. Finally, MAST-4 was absent from the cold Antarctic waters below 5°C and present at low abundances (150 molecules ml^{-1}) north of the South Atlantic Front.

Comparison of Q-PCR and FISH quantifications

The FISH counts were done in a subset of samples from Blanes Bay (monthly during years 2001 and 2003 and some more during 2005, Fig. 4) and from the Indian Ocean (vertical profiles at stations 1, 9 and 23, Fig. 3). Except the deepest Indian samples investigated (200 m), all samples yielded a significant MAST-4 count by FISH, with concentrations ranging from 18 to 244 cells ml^{-1} (Table 1). Correlating Q-PCR and FISH estimates for these two data sets was moderate, with a R^2 of 0.47 for Blanes samples (Fig. 5A) and of 0.65 for Indian samples (Fig. 5B). The slope of these correlations, representing the rDNA copy number per MAST-4 cell, was estimated to be 11 and 6 respectively.

In the unamended seawater incubations, samples to be processed by Q-PCR were collected to maximize DNA recovery, in contrast with environmental samples that were generally collected to optimize DNA quality (see *Discussion*). In the Indian Ocean incubations, MAST-4 peaked after 2 days (2×10^4 molecules ml^{-1}), whereas in the Blanes Bay incubations the peak occurred the fourth day (3×10^4 molecules ml^{-1}) (data not shown). As expected in these incubations, the abundance of MAST-4 molecules decreased after the peak to very low levels in both cases. The comparison of Q-PCR signal and FISH counts was much better with these samples, with a R^2 of 0.95 for Blanes samples (Fig. 5C) and of 0.99 for Indian samples (Fig. 5D). The number of the rDNA copies per cell was estimated to be 29 in the Blanes and 37 in the Indian incubation data sets. Interestingly, in an additional incubation performed in Blanes Bay on July 2005, the comparison of Q-PCR and FISH data yielded an rDNA copy number of 22 ($R^2 = 0.99$; $n = 4$) (data not shown).

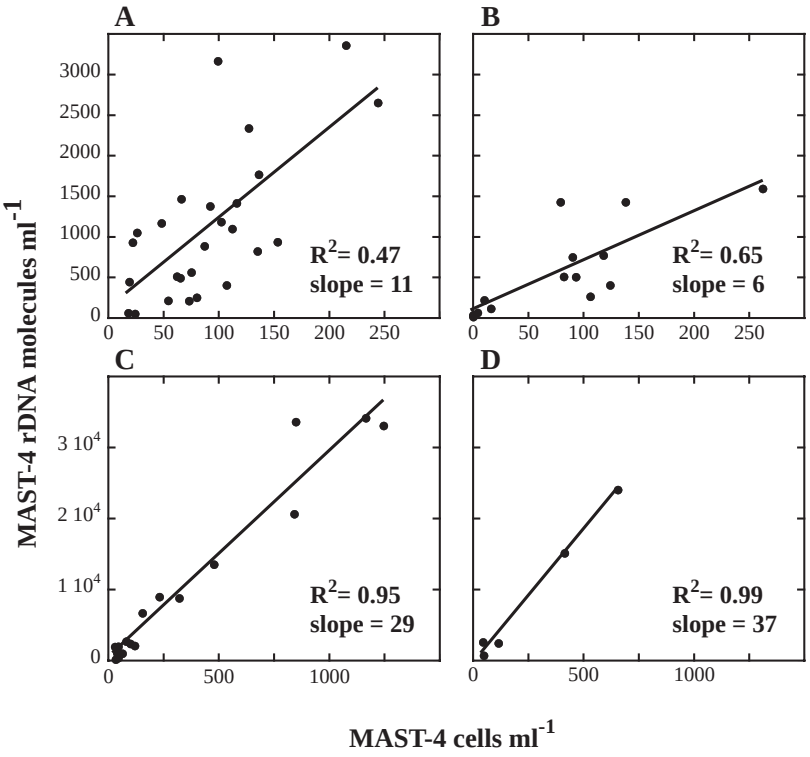


Fig. 5. Relationship between the MAST-4 signal estimated by Q-PCR (rDNA molecules ml^{-1}) and by FISH (cells ml^{-1}) in environmental samples from the Mediterranean Sea (A) and the Indian Ocean (B), and samples from unamended seawater incubations from the Mediterranean Sea (C) and the Indian Ocean (D).

Discussion

Quantification of an uncultured protist taxa by Q-PCR

Here we have optimized, calibrated and validated a Q-PCR protocol to assess the abundance in marine waters of the uncultured protist MAST-4. Whereas some previous studies have applied Q-PCR for uncultured marine prokaryotes, such as *Sulphurimonas denitrificans*-like (Labrenz *et al*., 2004), our study represents the first application of Q-PCR to uncultured protists, whose presence has been only inferred by molecular tools (environmental sequencing and FISH probing). The main difference when dealing with uncultured organisms is that the standard to calibrate the Q-PCR signal cannot be cultured cells. Here we have used a plasmid carrying the target sequence and, therefore, target molecules are quantified instead of target cells. To convert the abundance of rDNA molecules to cells, the rDNA operon copy number in the genome must be known, and this can be derived by comparing the Q-PCR and the FISH signals from the same samples (see later).

The reliability of the Q-PCR data depends on several critical aspects, such as the quality of the DNA extract, the specificity of the amplified PCR product and obtaining optimal amplification efficiencies (Cankar *et al*., 2006). In our study, it has been essential to further purify the DNA extracts from environmental samples, which did not amplify initially, by a DNA precipitation step with ethanol. The plasmid DNA samples, extracted with a different system, did not need the precipitation step to amplify properly. This confirms that different extraction methods can influence the purity of the DNA and have a great impact on the results obtained by Q-PCR (Peano *et al*., 2004). After this cleaning step, the regular Q-PCR checks gave very satisfactory results. Thus, the primer set used was highly specific for MAST-4, the efficiencies of the standard and the environmental samples were similar and close to 100%, and the melting curve analysis indicate that a single PCR product ($Tm = 84°C$) was always generated in the samples with positive signal.

Moreover, once the Q-PCR protocol was stabilized, we performed additional tests to assess its applicability on environmental samples. First, we studied the influence of freeze/thaw cycles on the percentage of recovered target molecules, as it is known that this process can compromise the integrity of DNA (Bellete *et al*., 2003). Here we show that up to seven cycles do not affect the MAST-4 quantified in environmental samples. This result gives confidence that the numbers obtained are realistic despite years of storage of the DNA extracts, including several thawing events. This might be due to the use of a very short amplicon (only 188 bp), which would still be amplified properly even if some DNA breakage occurs. Other studies have shown the stability of DNA extracts during long-term storage (Jerome *et al*., 2002). Second, with the spiking experiments we demonstrated the absence of PCR inhibitors in the negative samples (so they really lacked target molecules) and that the same number of target molecules amplify roughly equally even when diluted with a lot of non-target DNA (Bellete *et al*., 2003). Thus, the Q-PCR protocol presented here appears robust and adequate to quantify the rDNA molecules of MAST-4 in marine samples.

Q-PCR and FISH comparison and estimated rDNA copy number

There is a very good correlation between Q-PCR and FISH signals in the incubation samples, particularly robust in the Blanes incubation (Fig. 5C). With these results we can estimate that MAST-4 cells have around 30 copies of the rDNA operon. This value is fundamental to interpret the Q-PCR results targeting the rDNA genes, specially in the light that the rDNA copy number can vary orders of magnitude in protists, from 1–4 (some green algae) to more than 10 000 (some dinoflagellates). Comparing different eukaryotic species, a strong correlation has been found between the rDNA copy number and genome size (Prokopowich *et al*., 2003) and cell length (Zhu *et al*., 2005). Having around 30 copies of the rDNA operon (a comparatively low number) in MAST-4 cells is consistent with its small size, 2–3 μm in diameter (Massana *et al*., 2006a), and fits well within the relationship described for 18 phytoplankton strains (Zhu *et al*., 2005).

The correlation between Q-PCR and FISH signals in environmental samples is considerably less robust (Fig. 5A and B) and the copy number is lower (around 10). A possible explanation for this noisier signal is that the distinct MAST-4 lineages in the environmental samples (the probes target a phylogenetic group with up to 3% divergence in the complete 18S rDNA) also vary in their rDNA copy number, whereas the incubations are likely selecting a single genotype. Nevertheless, we consider that the main cause of the larger variability and the lower rDNA copy number in the environmental samples was that they were not collected to be quantitatively processed (as, for instance, detailed in Boström *et al*., 2004). First, in some samples the volume of seawater filtered and the volume of DNA extract obtained were only approximate. Second, samples were collected on encapsulated Sterivex filters that are known to be less efficient in DNA recovery. We have seen that Sterivex units might recover only half of the DNA quantity as compared with regular filters (R. Massana, unpubl. data). Third, during the DNA extraction, emphasis was done on the quality of DNA and this surely caused significant DNA losses, particularly during the phenolization step. These factors could cause a less efficient and inexact DNA recovery, therefore

yielding a lower and noisier Q-PCR signal. In fact, when these issues were properly addressed in the incubation samples, both techniques correlated very well and the rDNA copy number was higher. So, the Q-PCR is very suitable to quantify MAST-4 as long as it is combined with a careful sample collection and DNA extraction. Taking into account the constrains of applying Q-PCR with our environmental samples, it is clear that the data obtained provides interesting insights into the distribution and abundance of MAST-4 (Fig. 3). Samples were processed similarly, so they can be compared to provide useful global views of MAST-4 distributional patterns (presence and predominance). In addition, the data generated can be regarded as minimal estimates of the abundance of MAST-4 molecules.

Distribution of MAST-4 in the oceans

In a previous study, 24 surface samples from different world oceans were processed by FISH to estimate the abundance of MAST-4 cells (Massana *et al*., 2006a). This protist appeared in all samples investigated, except the polar ones, with averaged abundances of 131 cells ml^{-1}. The Q-PCR protocol allows a faster sample processing, so we have increased the number of samples analysed by one order of magnitude (214 samples), addressing new aspects such as the depth distribution or the interannual variation. Our data confirm and expand the FISH data indicating that MAST-4 is a widespread protist and, interestingly, also finds some environmental constrains that limit this broad distribution.

The vertical profiles in the Alboran Sea and the Indian Ocean show that MAST-4 is found in the upper ocean, including the photic zone and the upper aphotic zone. MAST-4 seems to be more abundant at subsurface (near the DCM) than at surface (5 m). Perhaps there is a negative effect of UV light at surface (Moran and Zepp, 2000) or simply the DCM is a more active and convenient habitat for MAST-4. Conversely, MAST-4 was hardly found in the deeper aphotic region, which is in accordance to the fact that heterotrophic flagellates and their bacterial food are becoming more scarce in mesopelagic waters (Tanaka and Rassoulzadegan, 2002; Fukuda *et al*., 2007). So, if present, MAST-4 cells were likely below the detection limit of Q-PCR. In fact, the three samples quantified by FISH at 200 m in the Indian Ocean did not reveal any target cell (abundance < 2 cells ml^{-1}). Moreover, cloning and sequencing of mesopelagic and bathypelagic waters has never retrieved a MAST-4 sequence (López-García *et al*., 2001; Countway *et al*., 2007; Not *et al*., 2007).

Seawater temperature seems to be a second constrain in the distribution of MAST-4. In a previous study, this protist could not be detected by FISH in polar samples (Massana *et al*., 2006a) and here we had a unique opportunity to identify its real boundary by analysing a transect in the Drake Passage from South Atlantic to Antarctic waters. Indeed, the coldest samples from this transect, with temperatures up to 4°C, gave negative signal for MAST-4. Only the northernmost stations of the transect, with warmer temperatures (5–6°C), show MAST-4 signal, although rather low as compared with that from the other systems. Thus, MAST-4 seems to be excluded in polar waters below 5°C, an intriguing feature shared by other microorganisms such as marine picocyanobacteria (Partensky *et al*., 1999).

So, combining the FISH data with the Q-PCR data it appears that MAST-4 is a structural component of protist assemblages in marine temperate photic waters. Virtually all samples from epipelagic waters (surface to 120 m) and with temperatures above 5°C have MAST-4 molecules. This broad and systematic presence is shared with several marine bacteria like SAR11 (Morris *et al*., 2002), and *Roseobacter* (Selje *et al*., 2004) and could be a common trait of smallest marine protists, as clone libraries of picoeukaryotes retrieve similar groups in distant oceans (Epstein and López-García, 2008). There are few reports on the abundance and distribution of picoeukaryotes at large oceanographic scales. Recently, small prasinophytes have been studied (Not *et al*., 2005; Not *et al*., 2008), and they change orders of magnitude from the coast, where they are more abundant, to the open sea. For larger protists, those that can be identified by microscopy, it is well known that they are not always present in marine samples, and temporality seems to be extremely important. Nevertheless, we have to keep in mind the significant phylogenetic diversity of MAST-4. So, although they look the same by FISH (Massana *et al*., 2006a), they can include distinct lineages with different and complementary ecological adaptations that might explain this broad distribution, as has been proposed for other picoeukaryotes (Rodríguez *et al*., 2005).

Sampling a coastal station allows a detailed temporal assessment of microbial dynamics. This cannot be easily done in the open sea, where each cruise represents a single temporal snapshot. We could not identify a seasonal pattern for the abundance of MAST-4 in the Mediterranean coastal station. There was a large variation (up to 3000 molecules ml^{-1}) between consecutive dates, and averaging different periods did not yield systematic trends. Probably MAST-4 varies on a shorter time scale and sampling only once a month does not properly describe its temporal variation. Data from the incubation experiments indicate that MAST-4 might be a typical r-strategist that responds with high growth rates to increase in prey and declines rapidly when prey diminishes. This could explain the irregular peaks during the season. Surprisingly, we detected important interannual variations, and we could not explain these by the environmental parameters currently taken,

such as temperature, salinity, inorganic nutrients, chlorophyll or microbial counts.

In summary, we present a very robust Q-PCR protocol for a fast quantification of rDNA molecules of the uncultured protist MAST-4. The extent that this protocol gives absolute abundance on environmental samples depends on the care with which sample collection and DNA extraction were done. The application of this protocol to a large sample collection from different oceanographic cruises, including some where only DNA extracts were available, yields a global vision of the distribution of this taxon. MAST-4 appears as a constitutive member (always found) of most marine systems and also identifies some habitats where it is excluded, such as mesopelagic and polar waters.

Experimental procedures

Environmental DNA from marine assemblages of small protists

Samples from three oceanographic cruises were taken at different depths with Niskin bottles attached to a CTD rosette. Alboran Sea samples were collected on 2–4 May 1998 during cruise MTP-II-MATER/HESP/04-98 on board the Spanish RV *Hespérides*. Drake Passage samples were collected on 6–14 December 1998 during cruise DHARMA on board RV *Hespérides*. Indian Ocean samples were collected on 16 May to 11 June 2003 on board RV *Melville* (Scripps Institution of Oceanography, US). Some physico-chemical (temperature, salinity, inorganic nutrients, chlorophyll) and biological data have already been published for the Alboran Sea (Arin *et al*., 2002), Drake Passage (Díez *et al*., 2004) and Indian Ocean (Not *et al*., 2008) cruises. We also collected surface samples from the Blanes Bay Microbial Observatory in the Mediterranean Sea (41°40′N, 2°48′E), monthly from March 2001 to June 2006. Samples (5–20 l) were first prefiltrated through a 200 μm nylon mesh and then collected in Sterivex filter units of 0.2 μm pore size (Durapore; Millipore) after being prefiltrated through 3 μm (Indian Ocean and Blanes Bay) or 5 μm (Alboran Sea and Drake Passage). Sterivex units were filled up with lysis buffer (40 mM EDTA, 50 mM Tris-HCl and 0.75 M sucrose) and kept frozen (–20°C during the cruises and –80°C afterwards) until DNA extraction.

Unamended incubations were prepared by filtering surface seawater by gravity first through a nylon mesh of 200 μm and second through 3-μm-pore-size polycarbonate filters (Massana *et al*., 2006b). Subsequently, the filtered seawater was dispensed into Nalgene polycarbonate bottles and incubated at near *in situ* temperature. Two bottles were prepared in the Indian Ocean (this is the Coastal incubation in Massana *et al*., 2006b), one incubated at ambient light and the other in the dark. Another two bottles were prepared in the Blanes Bay (sampled on 7 March 2006) and incubated in the dark. Mean temperatures for the Indian Ocean and Blanes Bay incubations were 21°C and 13°C respectively. Bottles were sampled every 1–2 days: 100–200 ml of seawater was filtered onto 25 mm Durapore filters of 0.2 μm pore size, submerged in lysis buffer and kept as before.

DNA extraction was done as described before (Massana *et al*., 2000). Cell lysis was performed by digestion with lysozyme followed by proteinase K and SDS treatments. DNA was purified twice with phenol : chloroform : isoamyl alcohol (25:24:1, pH 8) and once with chloroform : isoamyl alcohol (24:1), desalted and concentrated with a Centricon-100 (Millipore). Special effort was done with the samples from the unamended incubations to maximize the quantity of DNA recovered during the phenolization step. The integrity of the DNA was checked by agarose gel electrophoresis. Nucleic acid extracts were stored at –80°C until they were analysed. To obtain a positive amplification, all the environmental samples needed a DNA precipitation step. The DNA extract was mixed with 2.8 vols of precipitation mix (absolute ethanol, 2 M NaAc and 1 M $MgCl_2$) and kept more than half an hour at –80°C. Samples were then centrifuged at 14.000 r.p.m. for 15 min, washed twice with ethanol 70%, and re-suspended in milliQ water.

Development and optimization of the Q-PCR protocol

Design of specific primers for 18S rDNA of MAST-4. The forward primer M41f (5′-GTC TGC ACT GGA GTC GG-3′) was designed in base of all MAST-4 sequences available so far (34 clones from nine different marine sites). It matches perfectly all these clones, except two from the Indian Ocean (IND 31.115 and IND 58.11) that have one mismatch. It has more than five mismatches to all non-target sequences in GenBank. The reverse primer has the sequence of the probe NS4 designed for FISH (5′-TAC TTC GGT CTG CAA ACC-3′), which matches all target sequences and has at least two to three internal mismatches with all non-target sequences in GenBank (Massana *et al*., 2002). Primers were optimized using the PerlPrimer software (Marshall, 2004) in order to check the no formation of primer-dimers, the GC content and the theoretical melting temperature. The amplicon was 188 bp of length. The specificity of the primer set was checked by standard PCR. It gave negative amplification for 15 non-target clones and positive amplification for six target clones (including the two clones with one mismatch).

Preparation of the standard plasmid. 18S rRNA genes from an Indian Ocean sample (Not *et al*., 2008) were amplified with the universal eukaryotic primers EukA and EukB (Medlin *et al*., 1988) and cloned with the TOPO TA cloning kit (Invitrogen). A clone (IND58.12) having a MAST-4 insert was used as standard for the Q-PCR. Its plasmid was extracted with the Plasmid DNA Purification kit (QIAGEN) and linearized by digesting the supercoiled plasmid with the restriction endonuclease NotI (Sigma). Similar to other studies (Suzuki *et al*., 2000), we have seen that the Q-PCR signal of the linearized plasmid is 10 times higher than the supercoiled plasmid (data not shown). The plasmid extract was purified with the precipitation step (described before) to process it equally to the environmental samples. The DNA concentration and purity of the plasmid extract were assessed with a NanoDrop (ND-1000 Spectrophotometer). The number of rDNA molecules in the plasmid extract was calculated using the following formula:

$$\text{molecules } \mu l^{-1} = [a/(5736 \times 660)] \times 6.022 \times 10^{23}$$

where a is the plasmid DNA concentration (g μl^{-1}), 5736 is the plasmid length (3931 bp of the vector plus 1805 bp of the 18S

rDNA insert), 660 is the average molecular weight of one base pair and 6.022×10^{23} is the molar constant (Avogadro constant).

Optimization of the Q-PCR conditions. The Q-PCR reactions (final volume of 15 μl) were done with 0.45 μl of forward and reverse primers (both 10 μM), 7.5 μl of iQ SYBR Green Supermix (Bio-Rad), 5.6 μl of sterile milliQ water and 1 μl of DNA template. The reaction mixtures were prepared in thin-wall tubes and cap strips (Bio-Rad) and filter Safeseal-Tips (Biozym), DNase, RNase, Pyrogen-free, inside a UV-sterilized chamber. Reactions were performed in a iCycler iQ Multi-Color Detection System (Bio-Rad) programmed with an enzyme activation step (95°C, 3 min) and 40 cycles of 10 s of denaturation at 94°C, 30 s of annealing extension at 59°C and 30 s of data collection at 72°C. Data were analysed using the Multicolor Real-Time PCR Detection System v 3.1 software (Bio-Rad). These conditions were decided after different tests done with the positive plasmid. The optimal primer concentration (0.3 μM each) was the one that gave the earliest target amplification and the lowest amount of primer-dimer (nine combinations tested). The optimal annealing-extension temperature (58.9°C) was found by testing a gradient from 55°C to 65°C.

For each Q-PCR run (96 tubes) we prepared three replicates of seven serial dilutions (from 10^8 to 10^2 rDNA molecules) from the standard (IND58.12 plasmid), three replicates of four serial dilutions (10^{-1}–10^{-5} μl of DNA extract) from the environmental sample used as relative standard and PCR efficiency control, three negative samples with milliQ water and two replicates of two dilutions (10^{-1} and 10^{-2} μl of DNA extract; undiluted samples generally did not amplify) from each environmental sample to assay (15 in total). After the Q-PCR run, the number of rDNA molecules in the tubes with environmental samples was obtained. These values were converted to true concentration (molecules ml^{-1} of seawater) in three steps: (i) considering the dilution factor to find out the number of molecules μl^{-1} of DNA extract, (ii) multiplying the later number by the volume of DNA extract (100–300 μl) to obtain the total number of molecules in the extract and (iii) dividing the later number by the volume of seawater collected (5–20 l).

Melting curve analysis. The SYBR Green I binds all double-stranded DNA, including specific and unspecific PCR products and primer-dimers. These can be distinguished by their different melting temperatures, which depend on their base composition and length. In the melting curve analysis the temperature is raised by a fraction of a degree and the change in fluorescence is measured. At Tm, the two DNA strands separate and the fluorescence rapidly decreases. The software plots the rate of change of the relative fluorescence units (RFU) with temperature (T) (-d(RFU)/dT) on the *y-axis* versus the temperature on the x-axis, and this will peak at the Tm. A dissociation curve from 55°C to 94°C was measured after the last Q-PCR cycle in all samples.

Q-PCR efficiency with plasmid and environmental samples. The Q-PCR efficiency was assessed by comparing the number of molecules estimated in the dilutions from the same sample. With an optimal efficiency of 100%, a 1–10 dilution should yield 10% of the molecules (or a Ct difference between dilutions of 3.3). The efficiency of the standard IND58.12 plasmid (three replicates in seven serial dilutions) was compared each time with the efficiency of one environmental sample used as relative standard (three replicates in four serial dilutions). Samples with different PCR efficiencies (or too far from 100%) were excluded from further analyses.

FISH

Samples for FISH were collected during the Indian Ocean cruise and the seasonal sampling in Blanes Bay. Seawater (100–200 ml) was fixed with filtered formaldehyde (3.7% final concentration) and filtered through 0.6 μm pore diameter polycarbonate filters.The FISH samples (80–150 ml) during the unamended seawater incubations were similarly collected. Filters were kept at –80°C until processed. For FISH we used the probe NS4, specific for MAST-4 (Massana *et al*., 2002), supplied with a CY3 fluorophore at the 5′ end. Thin pieces of filters were hybridized with the CY3-NS4 probe following the protocol described before (Pernthaler *et al*., 2001; Massana *et al*., 2002) and counter-stained with DAPI. Positive cells were then observed by epifluorescence with green light excitation (CY3-specific signal) and checked with UV radiation (DAPI staining).

Acknowledgements

We thank the captains and crews of Research Vessels *Hespérides* and *Melville* and the chief scientists (T. Calafat, C. Pedrós-Alió and D. Blackman) for providing an optimal environment for sampling. This study was supported by projects TRANSINDICO (REN2000-1471-CO2-01/MAR), ESTRAMAR (CTM2004-12631/MAR, MEC) and PROTAL (HA2004-088, MCyT) to R.M. and by an F.P.I. fellowship from the Spanish Ministry of Education and Science to R.R.M. We thank R. Martínez-Gutiérrez, F. Not, H Brockmöller and I. Ontoria for laboratory assistance, V. Balagué and C. Cardelús for maintaining the Blanes Bay sampling station, M. Pastor for ODV advice, M. Ribes for statistic help and F. Unrein for useful comments.

References

Ahlgren, N.A., Rocap, G., and Chisholm, S.W. (2006) Measurement of Prochlorococcus ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol* **8:** 441–454.

Arin, L., Morán, X.A.G., and Estrada, M. (2002) Phytoplankton size distribution and growth rates in the Alboran Sea (SW Mediterranean): short term variability related to mesoscale hydrodynamics. *J Plankton Res* **24:** 1019–1033.

Audemard, C., Reece, K.S., and Burreson, E.M. (2004) Real-time PCR for detection and quantification of the protistan parasite *Perkinsus marinus* in environmental waters. *Appl Environ Microb* **70:** 6611–6618.

Becker, S., Fahrbach, M., Böger, P., and Ernst, A. (2002) Quantitative tracing, by *Taq* nuclease assays, of a *Synechococcus* ecotype in a highly diversified natural population. *Appl Environ Microb* **68:** 4486–4494.

Bellete, B., Flori, P., Hafid, J., Raberin, H., and Sung, R.T.M. (2003) Influence of the quantity of nonspecific DNA and repeated freezing and thawing of samples on the quantification of DNA by the Light Cycler. *J Microbiol Methods* **55:** 213–219.

Boström, K.H., Simu, K., Hagstrom, A., and Riemann, L. (2004) Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol Oceanogr Meth* **2:** 365–373.

Bowers, H.A., Tengs, T., Glasgow, H.B., Burkholder, J.M., Rublee, P.A., and Oldach, D.W. (2000) Development of real-time PCR assays for rapid detection of *Pfiesteria piscicida* and related dinoflagellates. *Appl Environ Microb* **66:** 4641–4648.

Cankar, K., Štebih, D., Dreo, T., Žel, J., and Gruden, K. (2006) Critical points of DNA quantification by real-time PCR – effects of DNA extraction method and sample matrix on quantification of genetically modified organisms. *BMC Biotechnol* **6:** 37.

Countway, P.D., and Caron, D.A. (2006) Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microb* **72:** 2496–2506.

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.

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

Díez, B., Massana, R., Estrada, M., and Pedrós-Alió, C. (2004) Distribution of eukaryotic picoplankton assemblages across hydrographic fronts in the Southern Ocean, studied by denaturing gradient gel electrophoresis. *Limnol Oceanogr* **49:** 1022–1034.

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

Fenchel, T. (1986) The ecology of heterotrophic microflagellates. *Adv Microb Ecol* **9:** 57–97.

Fukuda, H., Sohrin, R., Nagata, T., and Koike, I. (2007) Size distribution and biomass of nanoflagellates in meso- and bathypelagic layers of the subarctic Pacific. *Aquat Microb Ecol* **46:** 203–207.

Galluzzi, L., Penna, A., Bertozzini, E., Vila, M., Garcés, E., and Magnani, M. (2004) Development of a real-time PCR assay for rapid detection and quantification of *Alexandrium minutum* (a dinoflagellate). *Appl Environ Microb* **70:** 1199–1206.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996) Real time quantitative PCR. *Genome Res* **6:** 986–994.

Jerome, K.R., Huang, M.L., Wald, A., Selke, S., and Corey, L. (2002) Quantitative stability of DNA after extended storage of clinical specimens as determined by real-time PCR. *J Clin Microbiol* **40:** 2609–2611.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. *Science* **311:** 1737–1740.

Labrenz, M., Brettar, I., Christen, R., Flavier, S., Bötel, J., and Höfle, M.G. (2004) Development and application of a real-time PCR approach for quantification of uncultured bacteria in the central Baltic Sea. *Appl Environ Microb* **70:** 4971–4979.

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

Marie, D., Zhu, F., Balague, V., Ras, J., and Vaulot, D. (2006) Eukaryotic picoplankton communities of the Mediterranean Sea in summer assessed by molecular approaches (DGGE, TTGE, QPCR). *FEMS Microbiol Ecol* **55:** 403–415.

Marshall, O.J. (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **20:** 2471–2472.

Massana, R., DeLong, E.F., and Pedrós-Alió, C. (2000) A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microb* **66:** 1777–1787.

Massana, R., Guillou, L., Díez, B., and Pedrós-Alió, C. (2002) Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Appl Environ Microb* **68:** 4554–4558.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., *et al.* (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microb* **70:** 3528–3534.

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

Massana, R., Guillou, L., Terrado, R., Forn, I., and Pedrós-Alió, C. (2006b) Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol* **45:** 171–180.

Medlin, L., Elwood, H.J., Stickel, S., and Sogin, M.L. (1988) The characterization of enzymatically amplified eukaryotic 16s-like rRNA-coding regions. *Gene* **71:** 491–499.

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

Moran, M.A., and Zepp, R.G. (2000) UV radiation effects on microbes and microbial processes. In *Microbial Ecology of the Oceans*. Kirchman, D.L. (ed.) New York, NY, USA: Wiley-Liss, pp. 201–228.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Not, F., Massana, R., Latasa, M., Marie, D., Colson, C., Eikrem, W., *et al.* (2005) Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50:** 1677–1686.

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.

Not, F., Latasa, M., Scharek, M., Viprey, P., Karleskind, V., Balagué, V., *et al.* (2008) Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep-Sea Res I* (in press): 10.1016/j.dsr.2008.1006.1007.

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

Peano, C., Samson, M.C., Palmieri, L., Gulli, M., and Marmiroli, N. (2004) Qualitative and quantitative evaluation of the genomic DNA extracted from GMO and Non-GMO foodstuffs with four different extraction methods. *J Agric Food Chem* **52:** 6962–6968.

Pernthaler, J. (2005) Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* **3:** 537–546.

Pernthaler, J., Glöckner, F.O., Schönhuber, W., and Amann, R. (2001) Fluorescence *in situ* hybridization (FISH) with rRNA-targeted oligonucleotide probes. *Methods Microbiol* **30:** 207–226.

Porter, K.G., and Feig, Y.S. (1980) The use of DAPI for identifying and counting aquatic microflora. *Limnol Oceanogr* **25:** 943–948.

Prokopowich, C.D., Gregory, T.R., and Crease, T.J. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46:** 48–50.

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

Schneeberger, C., Speiser, P., Kury, F., and Zeillinger, R. (1995) Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain. *PCR Methods Appl* **4:** 234–238.

Selje, N., Simon, M., and Brinkhoff, T. (2004) A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature* **427:** 445–448.

Suttle, C.A. (2005) Viruses in the sea. *Nature* **437:** 356–361.

Suzuki, M.T., Taylor, L.T., and DeLong, E.F. (2000) Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. *Appl Environ Microb* **66:** 4605–4614.

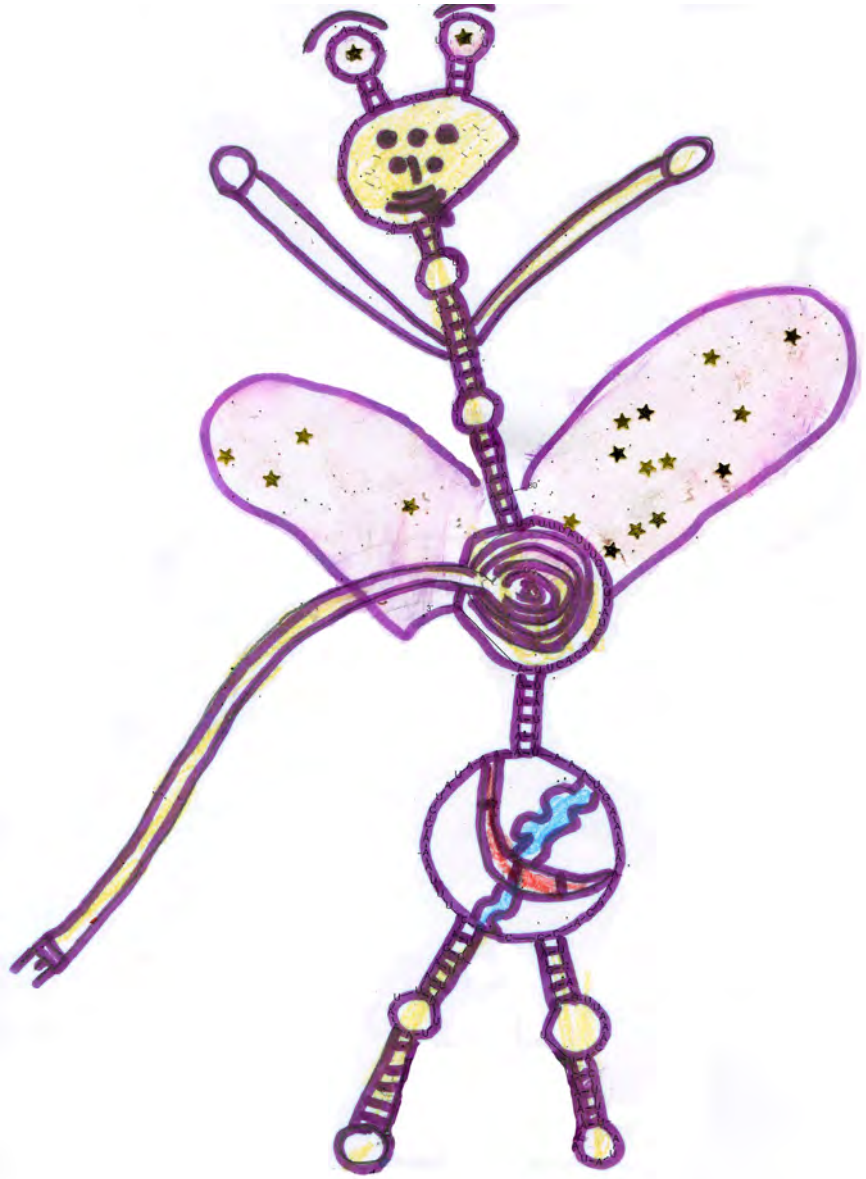
Tanaka, T., and Rassoulzadegan, F. (2002) Full-depth profile (0–2000 m) of bacteria, heterotrophic nanoflagellates and ciliates in the NW Mediterranean Sea: vertical partitioning of microbial trophic structures. *Deep-Sea Res II* **49:** 2093–2107.

Walker, N.J. (2002) A technique whose time has come. *Science* **296:** 557–559.

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

Chapter 2

Grazing rates and functional diversity of uncultured heterotrophic flagellates



Massana R, Unrein F, **Rodríguez-Martínez R**, Forn I, Lefort T, Pinhassi J *et al* (2009). Grazing rates and functional diversity of uncultured heterotrophic flagellates. The *ISME Journal* **3:** 588-596.

The ISME Journal (2009) 3, 588–596

www.nature.com/ismej

ORIGINAL ARTICLE

Grazing rates and functional diversity of uncultured heterotrophic flagellates

Ramon Massana[1], Fernando Unrein[1,3], Raquel Rodríguez-Martínez[1], Irene Forn[1], Thomas Lefort[1], Jarone Pinhassi[2] and Fabrice Not[1,4]

[1]*Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Barcelona, Catalonia, Spain and* [2]*Marine Microbiology, Department of Pure and Applied Natural Sciences, University of Kalmar, Kalmar, Sweden*

Aquatic assemblages of heterotrophic protists are very diverse and formed primarily by organisms that remain uncultured. Thus, a critical issue is assigning a functional role to this unknown biota. Here we measured grazing rates of uncultured protists in natural assemblages (detected by fluorescent *in situ* hybridization (FISH)), and investigated their prey preference over several bacterial tracers in short-term ingestion experiments. These included fluorescently labeled bacteria (FLB) and two strains of the *Roseobacter* lineage and the family *Flavobacteriaceae*, of various cell sizes, which were offered alive and detected by catalyzed reporter deposition-FISH after the ingestion. We obtained grazing rates of the globally distributed and uncultured marine stramenopiles groups 4 and 1 (MAST-4 and MAST-1C) flagellates. Using FLB, the grazing rate of MAST-4 was somewhat lower than whole community rates, consistent with its small size. MAST-4 preferred live bacteria, and clearance rates with these tracers were up to 2 nl per predator per h. On the other hand, grazing rates of MAST-1C differed strongly depending on the tracer prey used, and these differences could not be explained by cell viability. Highest rates were obtained using FLB whereas the flavobacteria strain was hardly ingested. Possible explanations would be that the small flavobacteria cells were outside the effective size range of edible prey, or that MAST-1C selects against this particular strain. Our original dual FISH protocol applied to grazing experiments reveals important functional differences between distinct uncultured protists and offers the possibility to disentangle the complexity of microbial food webs.

The ISME Journal (2009) **3,** 588–596; doi:10.1038/ismej.2008.130; published online 8 January 2009
Subject Category: microbial ecology and functional diversity of natural habitats
Keywords: functional diversity; grazing rates; MAST-4; MAST-1C; prey preference; uncultured flagellates

Introduction

Bacterial grazing is of fundamental importance in aquatic ecosystems and is carried out mostly by small flagellated protists up to 5 μm in diameter (Sherr and Sherr, 2002). It controls bacterial abundances in a wide range of ecosystem conditions, channels organic carbon to higher trophic levels, and releases inorganic nutrients that often are limiting primary production (Pernthaler, 2005; Jürgens and Massana, 2008). There are two main approaches to estimate community bacterivory rates: tracer techniques that follow the fate of an added bacterial surrogate and manipulation techniques that uncouple predator and prey populations (Strom, 2000). At the community level, sound information is available on how bacterial grazing relates to system productivity, temperature or other environmental variables in a wide range of oceanographic conditions (Sanders *et al.*, 1992; Vaqué *et al.*, 1994). Recently, a large phylogenetic diversity of small marine protists, mostly uncultured, has been unveiled (Moon-van der Staay *et al.*, 2001; Epstein and López-García, 2008), which likely implies a large functional diversity that needs to be considered for a better understanding of microbial food webs. Some studies have tried to tell apart the black box of bacterial grazers. For instance, the most popular tracer technique to estimate bacterivory, which inspects fluorescently labeled bacteria (FLB) inside protistan food vacuoles after short-term incubations (Sherr *et al.*, 1987), allows categorizing the grazers according to cell size, pigmentation or conspicuous morphologies (Simek *et al.*, 2004;

Correspondence: R Massana, Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta 37-49, Barcelona, Catalonia 08003, Spain.
E-mail: ramonm@icm.csic.es
[3]Current address: Instituto de Investigaciones Biotecnológicas—Instituto Tecnológico de Chascomús (CONICET), Argentina.
[4]Current address: Groupe Plancton Océanique, Station Biologique de Roscoff, France.
Received 20 October 2008; revised 1 December 2008; accepted 1 December 2008; published online 8 January 2009

Unrein *et al.*, 2007). Nowadays molecular techniques offer new tools to address simultaneously the phylogenetic and functional diversity of bacterial grazers.

There are several ways to explain how different grazers apparently using the same resource actually coexist and occupy separate ecological niches. First, each species might have different environmental optimum, being better adapted to a given range of physicochemical or biotic parameters. When this applies to strains from the same species the term ecotype is used (Rodríguez *et al.*, 2005; Boenigk *et al.*, 2007). Second, each species might have different prey preferences, being adapted to consume a specific part of the bacterial assemblage. The most critical parameter to define the grazing vulnerability of a given bacteria is its cell size (González *et al.*, 1990), but other factors such as cell viability (Landry *et al.*, 1991), surface properties (Matz and Jürgens, 2001), motility (Matz and Jürgens, 2005), phylogenetic affiliation (Jezbera *et al.*, 2005) or food quality (Shannon *et al.*, 2007) have been also demonstrated. Finally, intrinsic physiological parameters like the functional response (relationship of grazing rates with prey concentration) or the growth efficiency (conversion of ingested food to biomass) might explain adaptations to specific environmental settings. These functional features have been studied on model organisms grown in cultures (Fenchel, 1982; Eccleston-Parry and Leadbeater, 1994; Mohapatra and Fukami, 2004), but it has been suggested that these do not represent the dominant grazers in the sea (Massana *et al.*, 2006a).

Here we present an approach to study the grazing rates and prey preferences of uncultured heterotrophic flagellates (HFs) living in natural assemblages. It is based on the estimation of the feeding activity of specific grazers detected by fluorescent *in situ* hybridization (FISH) after short-term ingestion experiments with tracer preys. As grazers, we targeted two marine stramenopiles (MAST) lineages, each one including a significant and similar phylogenetic diversity (up to 3–4% in the 18S rDNA gene). These protists represent a noticeable fraction of *in situ* HFs, are globally distributed, and are bacterivorous (Massana *et al.*, 2006a), but little is known with respect to their feeding behavior. Besides the commonly used FLB as tracer, we also used live bacteria that were stained after the ingestion by a secondary FISH step. Our combined use of short-term ingestion experiments and double FISH procedure targeting both prey and predators shows that phylogenetic diversity is indeed contributing to functional diversity within marine bacterivorous assemblages.

Materials and methods

Natural microbial assemblages

Surface water from the Blanes Bay Microbial Observatory was taken on 13 June 2006 and carried to the laboratory in less than 2 h. This sample was prefiltered through a 100 μm mesh by an inverse filtration and used to perform a first grazing experiment with the *in situ* assemblage of protistan predators, mostly HFs and mixotrophic algae smaller than 5 μm. Simultaneously, 12 l of seawater was gently filtered by gravity through 3 μm polycarbonate filters and incubated in the dark at *in situ* temperature (20 °C) as explained before (Massana *et al.*, 2006b). After 2 days of unamended incubation, this sample was used to carry out a second grazing experiment with the incubated protistan assemblage.

Counts of bacteria (including heterotrophic bacteria and archaea), *Synechococcus*, HFs and phototrophic flagellates were carried out by epifluorescence microscopy (Porter and Feig, 1980). Glutaraldehyde-fixed aliquots (1% final concentration) were stained with 4′,6-diamidino-2-phenylindole (DAPI; 5 μg μl^{-1}) and filtered on 0.2 (for bacteria) or 0.6 μm (for flagellates) pore size polycarbonate filters. The filters were kept frozen until observed by ultraviolet irradiance and blue light in an Olympus BX61 microscope. Pictures of DAPI-stained bacteria were taken with a digital camera (Spot RT Slider; Diagnostic Instruments Inc., Sterling Heights, MI, USA) and processed with the Image Pro Plus software analyzer (Media Cybernetics Inc., Bethesda, MD, USA) to calculate the biovolume of 100–500 cells after the measured area and perimeter (Massana *et al.*, 1997). Bacterial viability was assessed with the nucleic acid double-staining (NADS) protocol (Grégori *et al.*, 2001) that uses SYBR Green to stain all cells and propidium iodide to stain cells with compromised membranes. Both populations were counted by flow cytometry and cells with intact membranes were considered 'alive' (Falcioni *et al.*, 2008).

Specific protist taxa were detected by FISH. Aliquots were fixed with formaldehyde (3.7% final concentration), filtered on 0.6 μm pore size polycarbonate filters and kept frozen until processed. Oligonucleotide probes for five MAST groups were used, NS4 for MAST-4 (Massana *et al.*, 2002), NS1A for MAST-1A, NS1B for MAST-1B, NS1C for MAST-1C and NS2 for MAST-2 (Massana *et al.*, 2006a), together with the general eukaryotic probe Euk502 (Lim *et al.*, 1999). Probes labeled with the fluorescent dye CY3 at the 5′ end were supplied by Thermo Electron Corporation (Waltham, MA, USA). For FISH we followed the protocol and conditions detailed previously (Pernthaler *et al.*, 2001; Massana *et al.*, 2006a). Briefly, filter portions with protist cells were hybridized for 3 h at 46 °C in the appropriate buffer (with 30% formamide), washed at 48 °C in a second buffer, counter-stained with DAPI and mounted in a slide. Cells were then observed by epifluorescence microscopy under green light excitation.

Bacterial strains as prey

Brevundimonas diminuta (syn. *Pseudomonas diminuta*; *Caulobacteraceae*, α-*Proteobacteria*) was

obtained from Colección Española de Cultivos Tipo (Valencia, Spain), grown in Luria-Bertrani agar plates and used to prepare FLB (Sherr *et al.*, 1987). Two-week-old colonies were scraped, diluted in carbonate–bicarbonate buffer, stained with 5-[4,6-dichlorotriazinyl]aminofluorescein for 2 h at 60 °C, kept at −20 °C, and thawed and sonicated before use as explained before (Unrein *et al.*, 2007). Strains MED479 (*Nereida* sp., *Roseobacter* lineage, *Rhodobacteraceae*, α-*Proteobacteria*) and MED134 (*Dokdonia* sp., *Flavobacteriaceae*, *Bacteroidetes*) were isolated in 2003 from the Blanes Bay Microbial Observatory in Zobell agar plates and kept since then in glycerol frozen stocks (Lekunberri *et al.*, submitted). Before the experiments, cells were grown on agar plates with Marine Broth 2216 (Difco, Lawrence, KS, USA) and then regrown in diluted (1:10) liquid Marine Broth (MED479) or in filtered autoclaved seawater (MED134).

Complete 16S rDNA sequences of MED479 (FJ482233) and MED134 (DQ481462) were imported to ARB (http://www.arb-home.de) to design specific oligonucleotide probes: NER380 (5′-GCATCGCTA GATCAGGGTTT-3′; *Escherichia coli* positions 381–400), and DOK196 (5′-TCTTATACCGCCGAAACT-3′; *E. coli* positions 197–228). Besides MED479, probe NER380 targets 22 GenBank entries of uncultured marine bacteria (including one clone from Blanes Bay) and five cultured strains (*Nereida ignava*, an isolate from the surface microlayer, and three other Blanes Bay strains). Besides MED134, probe DOK196 targets 27 uncultured marine clones and 28 cultured strains (5 *Dokdonia* sp., 5 *Krokinobacter* sp., 1 *Flexibacter* sp., 2 isolates from marine sponges and 15 from coastal bacterioplankton). Probes were supplied by Thermo Electron Corporation with an aminolink (C6) at the 5′ end, ligated with a horseradish peroxidase enzyme (Urdea *et al.*, 1988), and then optimized for catalyzed reporter deposition (CARD)-FISH (Pernthaler *et al.*, 2002) with bacterial cells (fixed and filtered as before) from the target culture. Filter pieces were permeabilized with lysozyme and achromopeptidase before hybridizing overnight at 35 °C in a buffer containing 30% formamide (which gave the best results after trying a range from 10% to 60%). After hybridization the signal was amplified with Alexa 488-labeled tyramide and counter-stained with DAPI. Filter pieces were mounted on a slide and observed by epifluorescence microscopy under blue light excitation.

Grazing experiments

Samples with the natural assemblages of bacteria and protists were acclimated in a large container (>10 l) for 2–4 h at *in situ* temperature (20 °C) and mean light intensity in the water column (200 μmol m^{-2} s^{-1}). Several 2-l bottles were filled-up, inoculated with a different bacterial suspension added at tracer concentrations (ca 15% of total bacteria), and dispensed into three 0.5-l bottles (triplicates). *In situ* abundance of strains MED479 and MED134 was checked. At time 0 and after 40 min of incubation, aliquots for DAPI-stained microbial counts and FISH analysis were taken as before, with the exception that fixation was carried out with an equal volume of diluted fixative to reduce cell egestion (Sieracki *et al.*, 1987) and that the final glutaraldehyde concentration was 2%. The incubation time (40 min) was chosen based on a previous time series that showed a plateau in the number of ingested bacteria at 45 min (Unrein *et al.*, 2007). In one case, grazing was estimated by counting FLB inside HFs in the DAPI-stained samples. In all other cases, counting ingestion involved a FISH step targeting specific predators. A single FISH step sufficed when using FLB as tracer, but an additional CARD-FISH hybridization was required to assess the ingestion of MED479 and MED134 cells that were offered alive and unstained. Optimal signals were obtained when CARD-FISH for bacteria was carried out first, followed by FISH for protists, according to the protocols explained above.

After the single or dual FISH procedures, the filter was inspected by epifluorescence at 1000 × under green light excitation to detect positive predator cells. When one was detected, the excitation was changed to blue light to count the tracer items ingested. On average, 125 predator cells were observed for each data point. The average number of tracer bacteria per predator was estimated for the initial sample (I_0) and the sample at 40 min (I_{40}), and used to calculate clearance rates (CRs: nl per predator per h) according to

$$CR = (I_{40} - I_0) \times (60/40) \times 1/(T)$$

where (T) represents the tracer prey concentration (in cells per nl). CRs were converted to ingestion rates (bacteria per predator per h) by multiplying by the bacterial concentration (native plus tracer cells, in cells per nl), assuming that both bacterial types were ingested at similar rates. Each grazing rate estimate (three replicate bottles × two times × two separate hybridizations) represents approximately 12 h of microscopy.

Results

To infer grazing rates of uncultured protist taxa, ingestion experiments were carried out with two starting microbial assemblages, the *in situ* assemblage from Blanes Bay and the assemblage resulting after 2 days of unamended dark incubation (Table 1). During this incubation, bacteria initially increased (Figure 1a) and then decreased concomitantly with the growth of HFs (Figure 1b). The protistan assemblage changed from one dominated by phototrophic cells (78%) to one dominated by heterotrophic cells (84%). Specific HF taxa belonging to different MAST lineages were quantified by FISH along the incubation. Two MAST groups (-4 and -1C) had abundances high enough (Table 1) to allow

Table 1 Microbial counts and bacterial parameters of the *in situ* (Blanes Bay, 13 June 2006) and incubated (unamended dark incubation for 2 days) samples used in the short-term ingestion experiments

	In situ *sample*	*Incubated sample*
Counts by DAPI staining (cells per ml)		
Bacteria	1203 582	1714 615
Synechococcus	25 598	14 195
Heterotrophic flagellates <5 μm	973	11 868
Phototrophic flagellates <5 μm	3421	2327
Counts by FISH (cells per ml)		
MAST-4	70	890
MAST-1C	11	217
MED479 (*Roseobacter* strain)	367	5350
MED134 (flavobacteria strain)	1749	0
Bacterial parameters		
Biovolume in μm^3 (s.e.)	0.099 (0.007)	0.069 (0.001)
Cells alive (%)	86	ND

Abbreviations: DAPI, 4′,6-diamidino-2-phenylindole; FISH, fluorescent *in situ* hybridization; MAST, marine stramenopiles; ND, not determined.

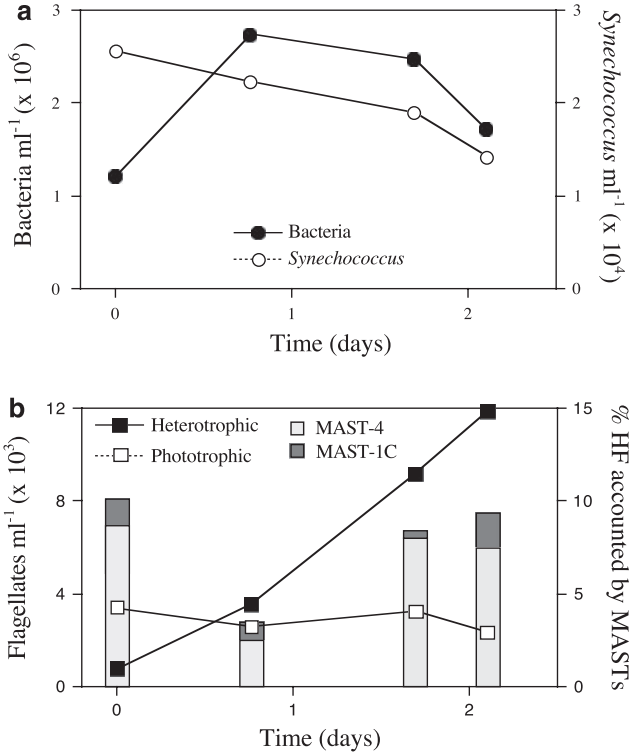


Figure 1 Abundance of bacteria and *Synechococcus* (**a**) and heterotrophic and phototrophic flagellates (**b**) during an unamended dark incubation started with Blanes Bay surface seawater. The percentage of heterotrophic flagellates accounted for marine stramenopile group 4 (MAST-4) and MAST-1C during the incubation is shown in (**b**).

determining their grazing rates, whereas the other three groups examined (-1A, -1B and -2) were too rare for this purpose (less than 20 cells per ml). Note that even though the number of HF increased during incubation, the relative contribution of MAST-4 and MAST-1C to the HF assemblage was maintained (Figure 1b).

The *Roseobacter* (*Nereida* sp. MED479) and the flavobacteria (*Dokdonia* sp. MED134) strains to be used as tracers in ingestion experiments were found at low *in situ* abundance (less than 0.3% of total bacteria; Table 1). The mean cell volume of MED479 and MED134 was 0.21 and 0.26 μm^3 when grown in standard rich media, and after starving for 1 or 2 weeks they reached a volume similar to that of natural bacteria (ca 0.1 μm^3; Tables 1 and 2). Besides getting smaller, the starvation compromised the viability of MED479 cells (only 5–25% of cells kept intact membranes), but not of MED134 cells. The double FISH hybridization we performed to estimate specific protistan grazing gave optimal signal, allowing easy detection of specific ingested bacteria (MED134 or MED479 cells) within specific predators (MAST-4 or MAST-1 cells; Figure 2).

Short-term ingestion experiments using FLB and live MED479 and MED134 (of different cell size) were carried out with the *in situ* sample (experiments 1–5; Table 2), and the incubated sample (experiments 6–8; Table 2). All the experiments yielded grazing rates for the uncultured MAST-4 protists (Figure 3). To compare rates obtained at different prey abundance (especially between *in situ* and incubated samples) we calculated both CRs (volume cleared) and ingestion rates (bacteria ingested), and both estimates gave a similar picture. CRs of MAST-4 were virtually identical in experiments 1 and 6 (Figure 3a; see FLB estimates), resulting in somewhat higher ingestion rates in the incubated sample because of the higher prey abundance (Figure 3c). This suggests that the specific activity of MAST-4 was not artificially stimulated by the incubation. Moreover, important differences in grazing rates were seen when using different tracers (analysis of variance (ANOVA), $F(7, 14) = 4.69$, $P = 0.0068$), with FLB giving lowest clearance and ingestion rates (0.7 nl per predator per h and 1.0 bacteria per predator per h) and MED134 highest rates (1.9 nl per predator per h and 2.9 bacteria per predator per h). These differences could not be explained by the cell size of tracers because rates can vary highly using tracers of similar biovolume (Figures 3a and c). A clear and statistically significant pattern ($R > 0.70$; $P < 0.05$) emerged when relating grazing rates with the percentage of live cells (as determined by NADS) in the tracer (Figures 3b and d). MAST-4 appears to prefer bacteria that are in good physiological condition. Clearance and ingestion rates of the *in situ* HF assemblage (cells <5 μm) measured with FLB in experiment 1 (2.1 nl per predator per h and 3.1 bacteria per predator per h) were higher than that of MAST-4, which therefore seemed to be less active than the average HF cell.

In experiments 6–8, another uncultured flagellate (MAST-1C) was relatively abundant and targeting

Table 2 Summary of grazing experiments performed, showing the sample used, the bacterial strains employed as tracer prey (biovolume in μm^3 and s.e. in brackets) and the predators assessed in each experiment

Experiment	*Sample*	*Prey used as tracer*			*Predators*
		Name	*Biovolume*	*% alive*	
1	*In situ*	FLB	0.179 (0.006)	0.0	MAST-4, heterotrophic flagellates
2	*In situ*	MED479	0.080 (0.001)	5.0	MAST-4
3	*In situ*	MED479	0.121 (0.004)	25.2	MAST-4
4	*In situ*	MED134	0.090 (0.005)	79.4	MAST-4
5	*In situ*	MED134	0.119 (0.003)	99.3	MAST-4
6	Incubated	FLB	0.179 (0.006)	0.0	MAST-4, MAST-1C, eukaryotes
7	Incubated	MED479	0.109 (0.003)	6.8	MAST-4, MAST-1C, eukaryotes
8	Incubated	MED134	0.096 (0.004)	94.2	MAST-4, MAST-1C, eukaryotes

Abbreviation: MAST, marine stramenopiles.
FLB refers to fluorescently labeled bacteria, MED479 to the *Roseobacter* strain and MED134 to the flavobacteria strain.

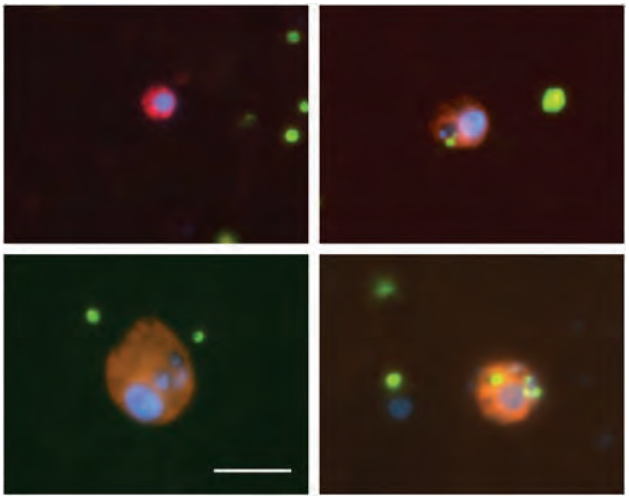


Figure 2 Epifluorescence micrographs of marine stramenopile group 4 (MAST-4; upper panels) and MAST-1C (lower panels) cells at the beginning (left panels) and at the end (right panels) of the ingestion experiment. Each image is an overlay of three pictures of the same cell observed under ultraviolet (UV) radiation (showing the blue nucleus after 4′,6-diamidino-2-phenylindole (DAPI) staining), green light (red cytoplasm after fluorescent *in situ* hybridization (FISH)) and blue light excitation (green MED134 or MED479 cells after catalyzed reporter deposition (CARD)-FISH). Scale bar is 5 μm and applies to all figures.

the assemblage with an eukaryotic probe provided rates mostly attributable to HF cells (Table 1). Grazing rates of these three predators (MAST-4, MAST-1C and eukaryotes) were estimated using different tracers (Figure 4). As these experiments were carried out at similar prey abundance, ingestion rates paralleled CRs and are not shown here. MAST-4 CRs are a subset of those presented in Figure 3, but under another display format, and showed again that live bacteria were preferred over FLB. The pattern observed for the whole eukaryotic community was comparable to that of MAST-4, with MED134 giving highest rates, although community rates were generally higher than MAST-4 rates, as observed in the *in situ* sample. On the other hand, MAST-1C deviated clearly from this picture, with FLB yielding the highest CR (with a value 2.5 times higher than that of MAST-4), whereas live bacteria were ingested at much lower rates, especially MED134 that was almost not ingested. All predator groups showed a significant difference between rates obtained with FLB and MED134 (ANOVA, *post hoc* Fisher's least significant difference test, $P<0.05$). Clearly, the food preferences of MAST-4 and MAST-1C were distinct and the latter did not behave as the average HF cell in the assemblage.

Finally, we did additional experiments with the incubated sample using other cell suspensions as tracers (data not shown). We employed MED134 cells that were damaged by heating the culture at 60 °C for 1 h. Cell viability was compromised (70% of cells lost membrane integrity), but cells preserved their shape and size and were readily detected by CARD-FISH. To our surprise, we did not detect any ingestion by any of the predators investigated (that is, MAST-4, MAST-1C or eukaryotes) when using these dead MED134 cells as tracers. We also used cultures of *Micromonas pusilla* and *Ostreococcus* sp. in other experiments, applying CARD-FISH with probes MICRO01 and OSTREO01 (Not *et al*., 2004) to detect ingestion of these picoprasinophytes. The results are not formally presented because these cells were added at unrealistically high abundance (10^4–10^5 cells per ml, whereas *in situ* abundance was 10^2–10^3 cells per ml) but they were still too scarce to obtain robust ingestion data (not enough ingestion cases were seen). Nevertheless, the experiments suggest that the two picoalgae were readily ingested by eukaryotes, MAST-4 and MAST-1C at CRs similar to the highest rates measured with bacterial tracers.

Discussion

Experimental settings

The aim of this study was to estimate *in situ* grazing rates of specific taxa of uncultured HFs and to study the putative prey preference (that is, functional diversity) among these taxa. To address our first objective we performed short-term ingestion

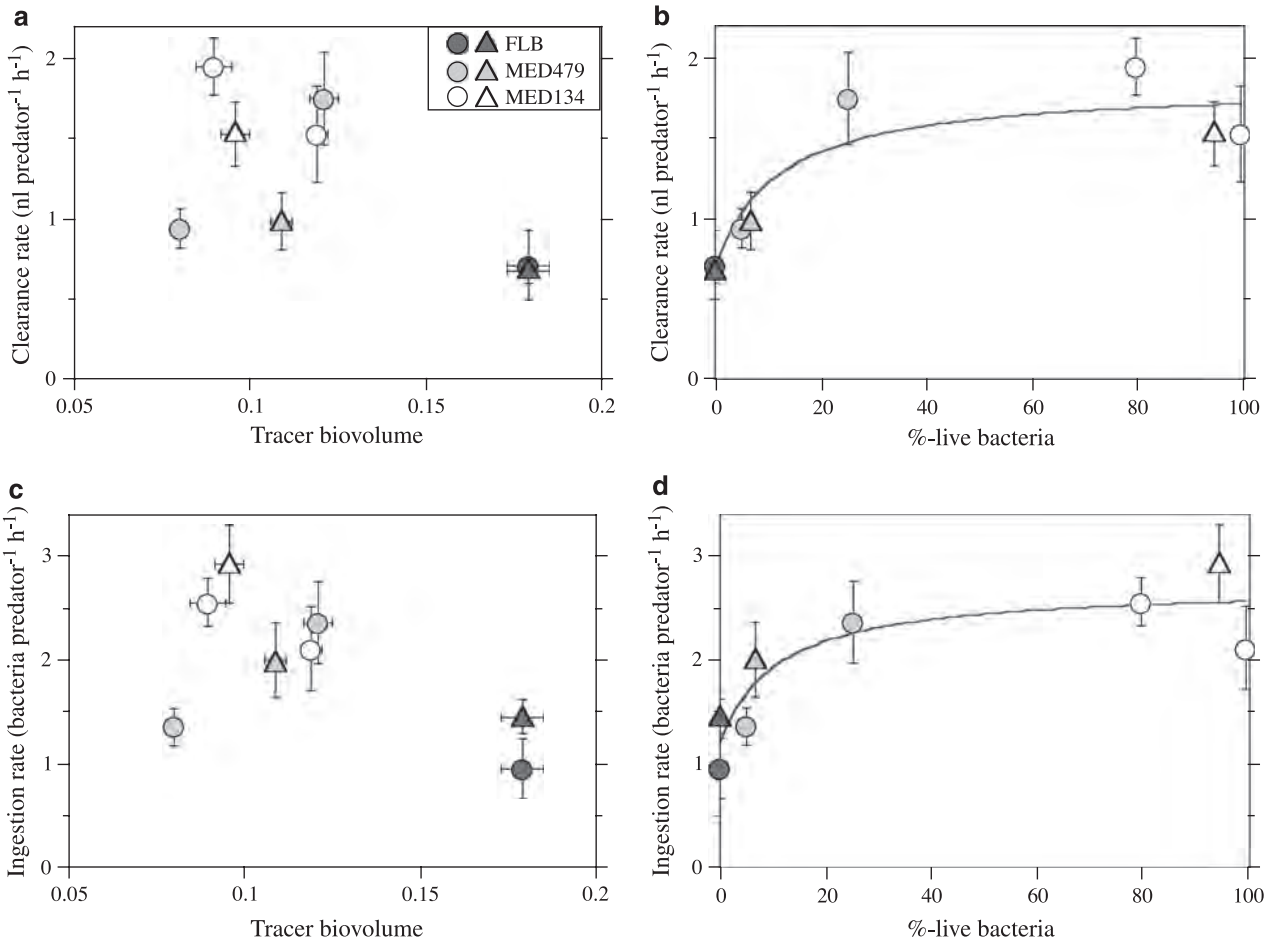


Figure 3 Clearance rates (**a**, **b**) and ingestion rates (**c**, **d**) of marine stramenopile group 4 (MAST-4) cells plotted against the biovolume (**a**, **c**) and the percentage of nucleic acid double-staining (NADS)-positive 'live' cells (**b**, **d**) of the bacterial tracer used in each of the eight independent experiments (circles when carried out with the *in situ* sample; triangles with the incubated sample). FLB refers to fluorescently labeled bacteria, MED479 to the *Roseobacter* strain and MED134 to the flavobacteria strain. A hyperbolic fit (Michaelis–Menten equation with an initial constant) was applied in (**b**) and (**d**). Bars represent standard errors.

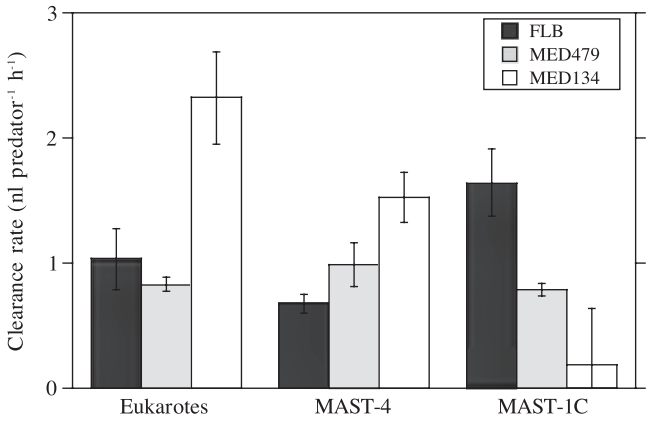


Figure 4 Clearance rates of the eukaryotic assemblage (mostly heterotrophic flagellate (HF) cells) and the specific marine stramenopile group 4 (MAST-4) and MAST-1C taxa in the incubated sample estimated with three different bacterial tracer preys: fluorescently labeled bacteria (FLB), MED479 (the *Roseobacter* strain) and MED134 (the flavobacteria strain). Bars represent standard errors.

experiments with natural microbial assemblages and applied FISH to measure the feeding rates of specific taxa. In addition to the *in situ* sample we analyzed a sample from an unamended incubation (Massana *et al.*, 2006b), which promote the growth of uncultured HF and therefore increase the chances of finding specific predators. For the second objective, besides the standard FLB, we used live bacteria that affiliate to the well-represented marine groups *Roseobacter* and *Flavobacteriaceae* (Kirchman, 2002; Buchan *et al.*, 2005; Alonso-Sáez *et al.*, 2007). The strains we used as tracers were isolated from the sampling point (Blanes Bay) and were too scarce in the original sample to interfere with the grazing experiments. We starved the bacterial cultures to reach a cell size comparable to that of natural bacteria, and assessed if this was accompanied by changes in other cell properties, such as membrane integrity. Also, we developed a dual FISH method to target simultaneously the predators and the ingested prey. This methodological setup led to a reliable protocol to estimate grazing rates and prey preferences of uncultured HF taxa on specific live bacteria. Until now, grazing rates and prey preferences were only known for cultured HF taxa under controlled laboratory conditions (Eccleston-Parry and Leadbeater, 1994; Mohapatra

and Fukami, 2004; Shannon *et al.*, 2007). Prey preference has been recently studied using FISH against specific bacterial prey within food vacuoles (Jezbera *et al.*, 2005), but this approach could not provide concurrently grazing rates nor specific activity for particular grazers.

Grazing rates of uncultured flagellates

We obtained *in situ* grazing rates of the uncultured HF taxa MAST-4 and MAST-1C. These have been detected only in molecular surveys (18S rDNA sequences and FISH-targeted cells) and are important members of marine assemblages, accounting for 9.2% and 2.7% of HF cells globally (Massana *et al.*, 2006a). Using the classical FLB procedure, clearance and ingestion rates for MAST-4 were 0.7 nl per predator per h and 1.0–1.5 bacteria per predator per h, and rates for MAST-1C were 1.6 nl per predator per h and 3.6 bacteria per predator per h. As their functional responses are still unknown, these values likely underestimate maximal CRs (if half-saturation constant (km) is low relative to actual bacterial abundances), or maximal ingestion rates (if km is high). Comparing these rates with those from the whole HF assemblage, it appears that MAST-4 is less active and MAST-1C more active than the average HF cell. This is consistent with a larger cell size of MAST-1C than MAST-4 (5.6 and 3.3 μm diameter on average in these samples). The specific grazing rates of these MAST taxa are comparable to *in situ* rates measured in Blanes Bay (Unrein *et al.*, 2007) and worldwide (Vaqué *et al.*, 1994), but much lower than most estimates derived from cultured strains. Indeed, maximal CRs from cultured HF strains range from 1 to 58 nl per predator per h, and maximal ingestion rates range from 5 to 259 bacteria per predator per h (Eccleston-Parry and Leadbeater, 1994). So, these cultured HF could be poor models of natural and dominant HF taxa.

Comparing FLB and live bacteria in ingestion experiments with MAST-4

Several studies have analyzed the effect of using dead bacteria as tracers in ingestion experiments. These generally show that using live bacteria result in grazing rates significantly higher than FLB (Landry *et al.*, 1991; Boenigk *et al.*, 2001). Differences can also be seen when comparing growing versus starving bacteria, the first being preferentially consumed (González *et al.*, 1993). Our results fit well with this general picture, and higher grazing rates of MAST-4 were obtained when using live bacteria over FLB (2–3 times higher). Moreover, MAST-4 grazing rates could be plotted to respect cell viability of the tracer suspensions, with maximal values being reached from ca 20% of live cells. An extreme case of negative selection is shown by the experiments using heat-killed MED134 cells. Those were not ingested at all, and the underlying mechanism for this prey avoidance is unknown. Finally, besides differences related to cell viability, no other differences in measured grazing rates were seen when using the two bacterial strains, even though they belong to distant phylogenetic groups with different life strategies. Members of the *Roseobacter* lineage tend to be free-living bacteria and typical of somewhat rich conditions (Buchan *et al.*, 2005), whereas flavobacteria tend to be particle-associated bacteria with high exoenzymatic activity (Kirchman, 2002). In our study, *Roseobacter* and flavobacteria cells were ingested equally by MAST-4, so these important differences in phylogeny and life strategy did not determine prey preference.

Functional differences between MAST-4 and MAST-1C

There were clear functional differences between these two taxa. MAST-4 cells preferred live prey and somehow represented the average *in situ* HF. MAST-1C cells, on the other hand, behaved very differently and the dead FLB yielded the highest rates. Most strikingly, the flavobacteria MED134 were almost not ingested by MAST-1C, which was unexpected because these bacterial cells showed the best physiological state and, in the same bottle (experiment 8), gave highest rates for MAST-4 and eukaryotes. A possible explanation would be that MAST-1C does not like this particular strain as food, opening an interesting and complex scenario of specific trophic interactions for some flagellates (but not for others). However, a more plausible explanation would be that the boundary of optimal prey size for MAST-1C falls within the size range of the tested bacteria. MED134, being the smallest of the bacteria tested, could be outside the size range of edible bacteria and escape predation. This could also explain the moderate rates measured with MED479 and the highest rates with the largest FLB. The fact that MAST-1C could be adapted to feed on larger bacteria than MAST-4 is consistent with its larger size, following the established relationship between predator and prey size (Fenchel, 1987). Our results clearly show functional differences between MAST-4 and MAST-1C, but the underlying mechanisms for such differences remain to be elucidated. The functional diversity we observed gives an ecological meaning to the high phylogenetic diversity of marine heterotrophic protists (Vaulot *et al.*, 2002).

Our study opens up the black box of bacterivory in marine ecosystems by showing different specific activity and prey preferences in distinct uncultured taxa. If our interpretation is correct, cell size was the main factor in prey vulnerability, as commonly accepted (González *et al.*, 1990), and there were sharp size boundaries outside which preys could not be ingested (Fenchel, 1987; Jürgens and Matz, 2002). Secondarily, when all preys fell within the edible size range then other factors interplayed with a less dramatic impact. For instance, MAST-4

preferred live bacteria in good physiological state, but still fed on dead FLB at one third of the maximal rate. Our data did not reveal differential feeding behavior related to the phylogenetic affiliation of the tested bacteria. This study assigned differential functional roles to distinct uncultured HF taxa, effectively linking phylogenetic and functional diversity within a natural assemblage. Our combination of FISH for specific predators with the use of live bacteria as prey surrogates allows addressing the huge complexity of microbial food webs.

Acknowledgements

This study was supported by the project ESTRAMAR (CTM2004-12631/MAR, MEC) to RM. FN was supported by the Marie-Curie fellowship ESUMAST (MEIF-CT-2005-025000) and TL by the project METAOCEANS (MEST-CT-2005-019678). We thank Josep M Gasol for help in flow cytometry, Matthias Engel for microscopic counts and Marta Ribes for statistical advice.

References

Alonso-Sáez L, Balagué V, Sà EL, Sánchez O, González JM, Pinhassi J *et al.* (2007). Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* **60**: 98–112.

Boenigk J, Arndt H, Cleven EJ. (2001). The problematic nature of fluorescently labeled bacteria (FLB) in *Spumella* feeding experiments—an explanation by using video microscopy. *Arch Hydrobiol* **152**: 329–338.

Boenigk J, Jost S, Stoeck T, Garstecki T. (2007). Differential thermal adaptation of clonal strains of a protist morphospecies originating from different climatic zones. *Environ Microbiol* **9**: 593–602.

Buchan A, González JM, Moran MA. (2005). Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol* **71**: 5665–5677.

Eccleston-Parry JD, Leadbeater BSC. (1994). A comparison of the growth kinetics of six marine heterotrophic nanoflagellates fed with one bacterial species. *Mar Ecol Prog Ser* **105**: 167–177.

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

Falcioni T, Papa S, Gasol JM. (2008). Evaluating the flow-cytometric nucleic acid double-staining protocol in realistic situations of planktonic bacterial death. *Appl Environ Microbiol* **74**: 1767–1779.

Fenchel T. (1982). Ecology of heterotrophic microflagellates. II. Bioenergetics and growth. *Mar Ecol Prog Ser* **8**: 225–231.

Fenchel T. (1987). *Ecology of Protozoa: The Biology of Free-Living Phagotrophic Protists*. Brock/Springer Series in Contemporary Bioscience, Springer: Berlin.

González JM, Sherr EB, Sherr BF. (1990). Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. *Appl Environ Microbiol* **56**: 583–589.

González JM, Sherr EB, Sherr BF. (1993). Differential feeding by marine flagellates on growing versus starving, and on motile versus nonmotile, bacterial prey. *Mar Ecol Prog Ser* **102**: 257–267.

Grégori G, Citterio S, Ghiani A, Labra M, Sgorbati S, Brown S *et al.* (2001). Resolution of viable and membrane-compromised bacteria in freshwater and marine waters based on analytical flow cytometry and nucleic acid double staining. *Appl Environ Microbiol* **67**: 4662–4670.

Jezbera J, Hornak K, Simek K. (2005). Food selection by bacterivorous protists: insight from the analysis of the food vacuole content by means of fluorescence *in situ* hybridization. *FEMS Microbiol Ecol* **52**: 351–363.

Jürgens K, Matz C. (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie van Leeuwenhoek* **81**: 413–434.

Jürgens K, Massana R. (2008). Protistan grazing on marine bacterioplankton. In: Kirchman DL (ed). *Microbial Ecology of the Oceans*, 2nd edn. Wiley: New York, pp 383–441.

Kirchman DL. (2002). The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.

Landry MR, Lehner-Fournier JM, Sundstrom JA, Fagerness VL, Selph KE. (1991). Discrimination between living and heat-killed prey by a marine zooflagellate *Paraphysomonas vestita* (Stokes). *J Exp Mar Biol Ecol* **146**: 139–152.

Lim EL, Dennet MR, Caron DA. (1999). The ecology of *Paraphysomonas imperforata* based on studies employing oligonucleotide probe identification in coastal water samples and enrichment cultures. *Limnol Oceanogr* **44**: 37–51.

Massana R, Gasol JM, Bjørnsen PK, Blackburn N, Hagström Å, Hietanen S *et al.* (1997). Measurement of bacterial size via image analysis of epifluorescence preparations: description of an inexpensive system and solutions to some of the most common problems. *Sci Mar* **61**: 397–407.

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

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

Massana R, Guillou L, Terrado R, Forn I, Pedrós-Alió C. (2006b). Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol* **45**: 171–180.

Matz C, Jürgens K. (2001). Effects of hydrophobic and electrostatic cell surface properties of bacteria on feeding rates of heterotrophic nanoflagellates. *Appl Environ Microbiol* **67**: 814–820.

Matz C, Jürgens K. (2005). High motility reduces grazing mortality of planktonic bacteria. *Appl Environ Microbiol* **71**: 921–929.

Mohapatra BR, Fukami K. (2004). Comparison of the numerical grazing response of two marine heterotrophic nanoflagellates fed with different bacteria. *J Sea Res* **52**: 99–107.

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, Latasa M, Marie D, Cariou T, Vaulot D, Simon N. (2004). A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in

the Western English Channel. *Appl Environ Microbiol* **70**: 4064–4072.

Pernthaler A, Pernthaler J, Amann R. (2002). Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Env Microbiol* **68**: 3094–3101.

Pernthaler J, Glöckner FO, Schönhuber W, Amann R. (2001). Fluorescence *in situ* hybridization (FISH) with rRNA-targeted oligonucleotide probes. In: Paul JH (ed). *Marine Microbiology*. Academic Press: London, pp 207–226.

Pernthaler J. (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* **3**: 537–546.

Porter KG, Feig YS. (1980). The use of DAPI for identifying and counting aquatic microflora. *Limnol Oceanogr* **25**: 943–948.

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

Sanders RW, Caron DA, Berninger UG. (1992). Relationships between bacteria and heterotrophic nanoplankton in marine and fresh waters: an inter-ecosystem comparison. *Mar Ecol Prog Ser* **86**: 1–14.

Shannon SP, Chrzanowski TH, Grover JP. (2007). Prey food quality affects flagellate ingestion rates. *Microb Ecol* **53**: 66–73.

Sherr BF, Sherr EB, Fallon RD. (1987). Use of monodispersed, fluorescently labeled bacteria to estimate *in situ* protozoan bacterivory. *Appl Environ Microbiol* **53**: 958–965.

Sherr EB, Sherr BF. (2002). Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* **81**: 293–308.

Sieracki ME, Haas LW, Caron DA, Lessard EJ. (1987). Effect of fixation on particle retention by microflagellates: underestimation of grazing rates. *Mar Ecol Prog Ser* **38**: 251–258.

Simek K, Jezbera J, Hornák K, Vrba J, Sed′a J. (2004). Role of diatom-attached choanoflagellate of the genus *Salpingoeca* as pelagic bacterivores. *Aquat Microb Ecol* **36**: 257–269.

Strom SL. (2000). Bacterivory: interactions between bacteria and their grazers. In: Kirchman DL (ed). *Microbial Ecology of the Oceans*. Wiley-Liss Inc: New York, pp 351–386.

Unrein F, Massana R, Alonso-Sáez L, Gasol JM. (2007). Significant year-round effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal system. *Limnol Oceanogr* **52**: 456–469.

Urdea MS, Warner BD, Running JA, Stempien M, Clyne J, Horn T. (1988). A comparison of non-radioisotopic hybridization assay methods using fluorescent, chemiluminescent and enzyme labeled synthetic oligodeoxyribonucleotide probes. *Nucleic Acids Res* **16**: 4937–4956.

Vaqué D, Gasol JM, Marrasé C. (1994). Grazing rates on bacteria: the significance of methodology and ecological factors. *Mar Ecol Prog Ser* **109**: 263–274.

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

Chapter 3

Functional responses of three heterotrophic flagellates taxa in mixed natural assemblages

Rodríguez-Martínez R, Vaqué D, Forn I and Massana R. Functional responses of three heterotrophic flagellates taxa in mixed natural assemblages.
In preparation to *Protist*

Functional responses of three heterotrophic flagellates taxa in mixed natural assemblages

Raquel Rodríguez-Martínez*, Dolors Vaqué, Irene Forn and Ramon Massana*

Institut de Ciències del Mar (CSIC). Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain

Summary

Grazing controls bacterial abundances and composition in a wide range of ecosystems and in aquatic systems, heterotrophic flagellates seem to be the main bacterial predators. Natural assemblages of marine heterotrophic flagellates are primarily formed by taxa that remain uncultured or that are cultured mimicking oligotrophic conditions. Thus, many aspects of their trophic behavior, including their functional response, are poorly known. Here we assessed by the first time the functional response (maximum ingestion rates and half-saturation constant) of three heterotrophic flagellates (MAST-4, *Minorisa minuta* candidatus and *Paraphysomonas* sp.) and of the total natural heterotrophic flagellates assemblage. We used fluorescently labeled bacteria as tracers and counted them inside protist food vacuoles. Natural heterotrophic flagellates had a Ks of 6.7 10^5 prey ml^{-1} lower than that of the traditional cultured flagellates, thus being well adapted to typical bacterial abundances of marine planktonic environments. Moreover, inside the heterotrophic flagellate assemblage, there were different taxa adapted to different ecological niches. *M. minuta* candidatus, recently cultured by mimicking natural conditions, was well adapted to low prey abundances being very efficient in ingesting FLBs. Uncultured MAST-4 cells were also well adapted to the typical abundances of marine bacteria but were less voracious. In contrast, *Paraphysomonas* sp., a typical cultured flagellate, did not achieve saturation of the ingestion rate even at the highest prey concentrations assayed (near 10^7 prey ml^{-1}). Our study sets the basis for the fundamental differences between cultured and uncultured bacterial grazers.

Keywords: Functional response, grazing, heterotrophic flagellates, MAST-4, *Minorisa minuta* candidatus, *Paraphysomonas imperforata*

Introduction

Heterotrophic flagellates (HF) are small predators considered to be the main consumers of aquatic bacteria (Sherr and Sherr 2002). Grazing by small flagellates controls bacterial abundance and composition in a wide range of ecosystem conditions, channels organic carbon to higher trophic levels, and releases inorganic nutrients that often are controlling primary production (Jürgens and Massana 2008, Pernthaler 2005). In fact, HF are central in the microbial loop concept (Azam et al 1983). Grazing rates of natural HF assemblages are estimated using tracer techniques that follow the fate of an added bacterial surrogate, or by manipulation techniques that uncouple predators and preys (Strom 2000, Vaqué et al 1994). Grazing rates may be then used to calculate community growth rates (Fenchel 1987). These rates average the activities of all populations in the community, each one perhaps having different rates and preferences. Indeed, recent molecular surveys have unveiled a large diversity of protists assemblages in the marine environment (Díez et al 2001, López-García et al 2001, Moon-van der Staay et al 2001), including HF (Jürgens and Massana 2008, Massana et al 2006b, Vaulot et al 2002). So, for a better understanding

of bacterial grazing in the sea, it would help to investigate particular physiological parameters of the building HF populations.

Prey abundance is one of the most important factors influencing grazing rates (Holling 1959). Its effect on grazing rates, named functional response, has been estimated in a wide variety of marine predators, including copepods (Henriksen et al 2007, Isari and Saiz 2011), dinoflagellates (Jeong et al 2005, Kim and Jeong 2004) or ciliates (Jonsson 1986, Jürgens and Simek 2000, Massana et al 1994). Functional responses can be adjusted to different mathematical models (Holling 1959), being the most popular among ecologists the equivalent to the enzyme kinetic model developed in 1913 by Leonor Michaelis and Maude Menten. This model is described by two parameters, the maximum ingestion rate (IR_{max}), determined by the mechanistic capacity of the predator to capture, handle and digest the prey, and the half-saturation constant (Ks: prey concentration that allows half the maximal rate), which is a proxy of the food concentrations at which the predator is adapted to live (Fenchel 1980). Only a few studies have measured directly the functional response of small flagellates (Jürgens and Matz 2002), since it is much easier to measure the numerical response, the relationship of growth rates with prey abundance. For these small predators, the growth efficiency is considered constant regardless prey abundance, so growth rates are directly proportional to ingestion rates and the numerical and functional responses have the same form (Fenchel 1987). Numerical responses have been studied on model flagellates grown in cultures (Anderson et al 2011, Eccleston-Parry and Leadbeater 1994, Fenchel 1982, Mohapatra and Fukami 2004), but it has been suggested that these do not represent the dominant grazers in the sea (Massana et al 2006b), which in many cases remain uncultured. A few trophic experiments (i.e. grazing rates and prey preferences) have been conducted with some of these uncultured flagellates (Massana et al 2009, Piwosz and Pernthaler 2010), but it is still unknown if they have a fundamentally different functional response than model cultured flagellates.

The aim of this study was to determine the functional response of uncultured flagellates living in mixed natural assemblages. We combined short-term ingestion experiments based on the use of fluorescently labeled bacteria (FLB) counted inside the protist food vacuoles (Sherr et al 1987), with specific FISH counts of HF within mixed assemblages. This time-consuming approach is so far the only way to provide specific ingestion rates for taxonomic classes of flagellates. Our study could have been done directly in natural assemblages, but low in situ abundances would have complicated the measurements. Instead, we used unamended seawater incubations, which are known to promote the growth of natural assemblages of uncultured HF (Massana et al 2006a). We obtained the functional responses of the total natural community and of three taxa of heterotrophic flagellates (MAST-4, *Minorisa minuta* candidatus and *Paraphysomonas* sp.). This is the first report of the functional response of specific uncultured flagellates, and highlights fundamental differences that might explain why they are not cultured by classical approaches.

Methods

Enrichment of uncultured heterotrophic flagellates by an unamended incubation

Surface seawater from the Blanes Bay Microbial Observatory was taken on October 16, 2007 and carried to the laboratory in less than 2 h. Six liters of seawater were filtered by gravity through a 200 µm nylon mesh and then through 3 µm pore size polycarbonate filters to reduce the presence of large predators in the sample. This water was incubated into Nalgene polycarbonate bottles at in situ temperature (19ºC) in the dark, to prevent the growth of phototrophic cells (Massana et al 2006a), and sampled daily during 5 days. Glutaraldehyde-fixed aliquots (2% final concentration) were stained with 4,6-diamidino-2-phenylindole (DAPI; 5 mg ml^{-1}) and filtered on 0.2 (for bacteria) or 0.6 µm (for flagellates) pore size black polycarbonate filters. Counts of bacteria (heterotrophic bacteria plus archaea) and heterotrophic flagellates (HF) were carried out by epifluorescence microscopy

(Porter and Feig 1980) with UV irradiance and blue light in an Olympus BX61 microscope at 1000X magnification. The grazing experiment reported here was done with the sample after three days of unamended incubation.

Detection of possible predators by clone library and FISH

Before starting the grazing experiment, 100 ml of seawater were collected onto a 0.2 μm pore size Durapore filter and DNA extraction was done as described before (Massana et al 2000). Cell lysis was performed by digestion with lysozyme followed by proteinase K and SDS treatments. DNA was purified with phenol:chloroform:isoamyl alcohol and concentrated with a Centricon-100 (Millipore). The PCR mixture (50 μl) contained 2 ng of template DNA, 0.5 μM of each primer, 200 μM of each dNTP, 1.5 mM $MgCl_2$, 1.25 units of a *Taq* DNA polymerase (ProOmega), and the enzyme buffer. We used the eukaryotic 18S rDNA primers 528F (Elwood et al 1985) and EUKR (Medlin et al 1988). PCR cycling, carried out in a BioRad thermocycler, was: initial denaturation at 94ºC for 3 min; 30 cycles with denaturation at 94ºC for 45 sec, annealing at 55ºC for 1 min and extension at 72ºC for 3 min; and a final extension at 72ºC for 10 min. PCR products were purified with the QIAquick PCR Purification kit (QIAGEN) and cloned using the TOPO-TA cloning kit (Invitrogen) with the vector pCR2.1. Putative positive bacterial colonies were picked to a new LB (Luria-Bertani) plate and finally into LB-glycerol solution for -80°C stocks. Presence of correct insert was checked by PCR reamplification with the same primers using a small aliquot of bacterial culture as template. Amplicons with the right insert size were sequenced at the Macrogen sequencing service (Korea). The phylogenetic affiliation of clones and the detection of putative chimeras were obtained by a basic local alignment search tool (BLAST).

Aliquots for FISH targeting small protists were fixed with formaldehyde (3.7% final concentration), filtered on 0.8 or 1 μm pore size polycarbonate filters and kept at -80ºC until processed. Oligonucleotide probes (Table 1) labeled with the fluorescent dye CY3 at the 5' end were supplied by Thermo Electron Corporation (Waltham, MA, USA). For FISH we followed the protocol and conditions detailed before (Massana et al 2006b, Pernthaler et al 2001). Briefly, filter portions with protist cells were hybridized for 3 h at 46ºC in the appropriate buffer (with 30% formamide), washed at 48ºC in a second buffer, counter-stained with DAPI and mounted in a slide. Cells were then observed by epifluorescence microscopy at 1000X under green light excitation.

Fluorescence labeled bacteria used as prey

Brevundimonas diminuta (syn. *Pseudomonas diminuta*; Caulobacteraceae, alfa-Proteobacteria) was obtained from the Colección Española de Cultivos Tipo (Valencia, Spain), grown in LB agar plates and used to prepare FLB (Sherr et al 1987). *B. diminuta* has already been used to prepare FLB (Vazquez-Dominguez et al 1999) because of their small size comparable to that of natural marine bacteria. Two-week-old colonies were scraped, diluted in carbonate–bicarbonate buffer (CO_3Na_2-HCO_3Na pH 9.5), and stained with 100 pg mL^{-1} of 5-(4,6-dichlorotriazinyl)-aminofluorescein (DTAF) for 2 h in a water bath at 60ºC. Stained cells were rinsed with 0.2 μm-filtered carbonate-bicarbonate buffer, resuspended, and centrifuged 5 times (10 min, 10,000 rpm) to prevent the transfer of leftover dye to experimental samples. Cell suspensions (average cell biovolume 0.073 μm^3) were kept frozen at -20ºC. Before using in the grazing experiments, the FLB solution was thawed and gently sonicated for three 10-s rounds with the microtip at 35% of power output (Dynatech sonic dismembrator, Model 300) to prevent cell clustering as explained before (Unrein et al 2007).

Grazing experiments

Seawater with the natural assemblages of bacteria and protists from the unamended incubation was divided in two sets (Fig. 1). One set was diluted (1 to 5.5) in order to decrease the initial bacterial concentration, whereas the other set remained undiluted. Five hundred ml bottles were filled-up with seawater from both sets (400 and 350 ml, respectively) and acclimated in a large container for

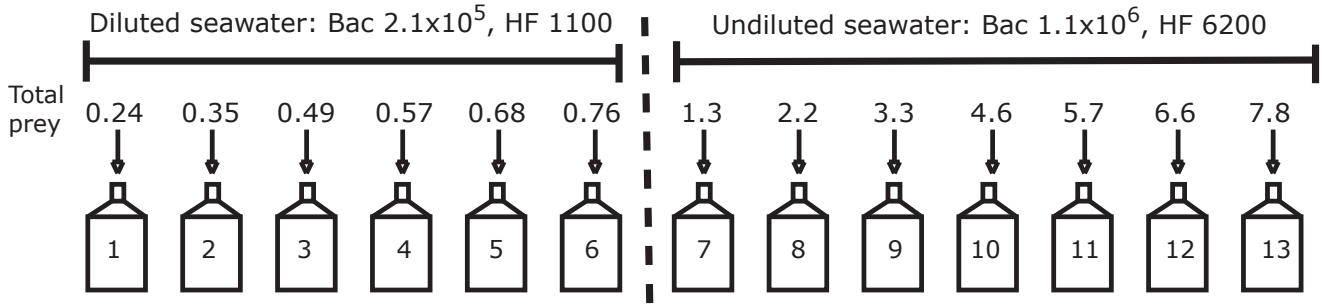


Figure 1. Scheme of the experimental design for assessing the functional response of natural HF. Several bottles with increasing amounts of prey (native bacteria plus FLBs added, in 10^6 cell ml^{-1}) were prepared. Bacterial and HF concentration (cells ml^{-1}) are indicated for the diluted (left) and undiluted (right) sets. A short-term ingestion experiment was performed in each experimental bottle.

2-4 h at in situ temperature (19ºC). Then, increasing amounts of FLBs were added to the bottles, at tracer concentrations (~15% of total bacteria) in the first bottle, and becoming the main bacterial prey in the other bottles (~600% in the last bottle). Instead replicating the same prey concentration, we decided to obtain more points along the prey gradient, a recommended strategy in a regression analysis (Montagnes and Berges 2004). Aliquots for DAPI-stained microbial counts (bacteria and small protists) and for FISH analyses (only small protists) were taken immediately after the addition of FLB and after 40 min of incubation. Fixation was carried out with an equal volume of diluted fixative to reduce cell egestion (Sieracki et al 1987), reaching the same final concentrations detailed above. The incubation time (40 min) was chosen based on a previous time series that showed a plateau in the number of ingested bacteria at 45 min (Unrein et al 2007). Grazing of the natural assemblage of heterotrophic flagellates was estimated by counting FLBs inside colorless flagellates in the DAPI-stained samples. Grazing of specific predators was assessed counting FLBs inside FISH positive cells, by combining green light excitation (FISH signal) and blue light excitation (FLB detection). The mean number of cells counted was 325, 300, 100 and 50 for community HF, *M. minuta*, MAST-4 and *Praphysomonas* sp., respectively. The number of FLBs per predator was multiplied times the ratio of total prey (native bacteria plus FLBs, obtained by separate DAPI counts) to FLB, to obtain the preys ingested at time 0 (I_0) and at 40 min (I_{40}). Ingestion rates (IRs: preys ingested per predator per h) were then calculated according to:

$$IR = (I_{40} - I_0) \times (60/40)$$

Clearance rates (CR: nl filtered per predator per h) were calculated by dividing the ingestion rate by the prey concentration (in nl). Data for IR and prey abundance were fitted by iteration to the hyperbolic Michaelis-Menten equation:

$$IR = x\, IR_{max}/(Ks + x)$$

where IR_{max} is the maximum ingestion rate, x is prey concentration (prey per milliliter) and Ks is the half-saturation constant (prey per milliliter). The maximum clearance rate (CR_{max}) was calculated by IR_{max}/ Ks (Fenchel 1986)

Results

The grazing experiment was done with a natural sample incubated for 3 days, during which in situ bacterial abundance (6.5 10^5 cells ml^{-1}) increased to 1.1 10^6 cells ml^{-1} and in situ HF abundance (840 cells ml^{-1}) increased to 6200 cells ml^{-1}, being in exponential growth at the moment of the experiment (data not shown). For some experimental bottles, this HF-enriched sample was diluted, and short-term ingestion experiments were prepared along a gradient of prey (native bacteria plus FLBs) covering almost two orders of magnitude (10^5 to 10^7 preys ml^{-1}) in 13 bottles (Fig. 1). This gradient covers the natural marine bacterial concentration, typically around 10^6 cells ml^{-1}. This experimental

set-up allowed to measure grazing rates at different prey abundances, and therefore estimate the functional responses of the whole natural assemblage and of specific heterotrophic flagellate taxa.

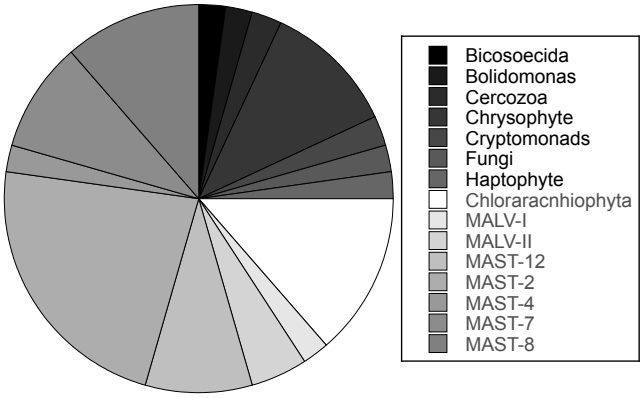


Figure 2. Pie chart representing the relative contribution of different phylogenetic groups in the clone library from the experimental sample (n= 44 clones). Black and dark grey indicates groups with cultured organisms, white indicates a group recently cultured simulating natural conditions, and light grey indicates uncultured groups.

A clone library (Fig. 2) was prepared with the HF-enriched sample to search for possible flagellate predators. The 44 clones with a correct sequence were distributed in 15 phylogenetic groups, and most were highly similar to environmental sequences retrieved in previous marine surveys. More than 50% of the library (24 clones) affiliated to uncultured MAST (Marine Stramenopiles). MAST-2 was the group more represented (10 clones), followed by MAST-8, -7 and -12 (4-5 clones) and MAST-4 with a single clone. Another three clones affiliated to uncultured MALV (Marine Alveolates) I and II. Six clones were almost identical (99.6% similarity) to a recently cultured heterotrophic flagellate of the Chlorarachniophyta group, *Minorisa minuta* candidatus (Del Campo et al submitted). The Chrysophyte group was represented with 5 clones, two of them belonging to uncultured clades, two highly similar to the cultured HF *P. imperforata* and *P. foraminifera*, and one moderately related (95.7% similarity) to *Ochromonas sp.* The remaining groups (Haptophyte, Fungi, Cercozoa, Bolidomonas and Bicosoecida) presented only one sequence. The bicosoecida sequence was 96.4% similar to the cultured HF *Caecitellus parvulus*.

Once we had a view of the heterotrophic flagellate composition in the experimental sample, we chose six FISH probes from uncultured and cultured heterotrophic flagellates (Table 1). No clear positive cell was observed with probes NS2 and CET1, despite they match perfectly with the sequences extracted from the clone library. Probe NS7 gave positive cells only when used in the CARD-FISH technique. So, despite MAST-7 cells were moderately abundant in the sample (300 cell ml^{-1}), they could not be used in the grazing experiments since the CARD-FISH and the FLB signals had the same epifluorescence signal. The remaining three probes, NS4, CRN02 and PIMP663, gave positive cells. *M. minuta* candidatus was the most abundant flagellate of the three (780 cells ml^{-1}), followed by MAST-4 (160 cells ml^{-1}), despite the later was represented by only one clone, and *Paraphysomonas* sp. (70 cells ml^{-1}). The probe PIMP 663 had one different mismatch with each one of the two sequences obtained in the clone library (99.3% similar among them), therefore it is likely that the FISH signal came from both strains.

Table 1. Probes checked in the unamended seawater incubation and the abundance of targeted cells.

Probe	Sequence (5'-3')	Ref.[a]	Group	Species	Cells ml^{-1}
NS4	TACTTCGG TCTGCAAACC	1	MAST-4		160
NS2	ATGGGCCGACCGGTCGCT	2	MAST-2		nd
NS7*	TCATTACCATAGTACGCA	3	MAST-7		300
CRN 02	TACTTAGCTCTCAGAACC	4	Chloraracnhiophyta	*M. minuta*	780
PIMP 663	GGACGCAGAGACCAGGTGCACA	5	Chrysophyte	*P. foraminifera*	70
CET1	CAGCTCAATACGGACACC	6	Bicosoecida	*C. parvulus*	nd

* CARD-FISH

[a]References: 1) Massana et al 2002, 2) Massana et al 2006a, 3) Massana et al In preparation, 4) del Campo et al 2011, 5) Lim et al 1999, 6) Massana et al 2007

nd: Not determined: no positive cells were observed in the hybridization

In the grazing experiments, we first estimated the functional response of the heterotrophic flagellates community (Fig. 3a), just by inspecting DAPI-filters without FISH hybridization. The maximal FLBs found inside a predator was 15 at the highest prey concentration. The estimated maximum ingestion rate was 2.3 prey HF^{-1} h^{-1} and the Ks was 6.7 10^5 prey ml^{-1}. Then, we estimated the functional response of three different heterotrophic flagellate taxa. *M. minuta* candidatus had up to 7 FLBs inside a single cell. The IR_{max} was more than double the community rates (Fig. 3b), and the Ks was only slightly lower. For the MAST-4, the maximum FLB number observed inside a cell was 5. The IR_{max} was rather low, only 1 prey HF^{-1} h^{-1} and the Ks was slighlty higher than that of the HF community (Fig. 3c). Finally, despite its low abundance, grazing rates could be calculated for *Paraphysomonas* sp. (although based in only 50 cells per filter). For this flagellate the maximum FLBs ingested per cell was 7, like in *M. minuta* candidatus. However ingestion rates did not saturate along the prey concentration used in this experiment (Fig. 3d), so we could not calculate IR_{max} and Ks for this species. Using the six bottles with highest prey abundance, we calculated the average percentage of predator cells ingesting different numbers of FLBs (Fig. 4). The HF assemblage and MAST-4 presented a large percentage of cells without ingestion (65 and 73%, respectively). Then, MAST-4 presented a peak of cells with only one FLB ingested, and few cells with more than two FLB. The HF community presented a less marked profile, with many cells having one to four FLB. On the other hand, *M. minuta* candidatus presented a lower percentage of cells without ingestion (28%), and had a peak of two FLB ingested per cell, with a notable percentage of cells with three or four FLB.

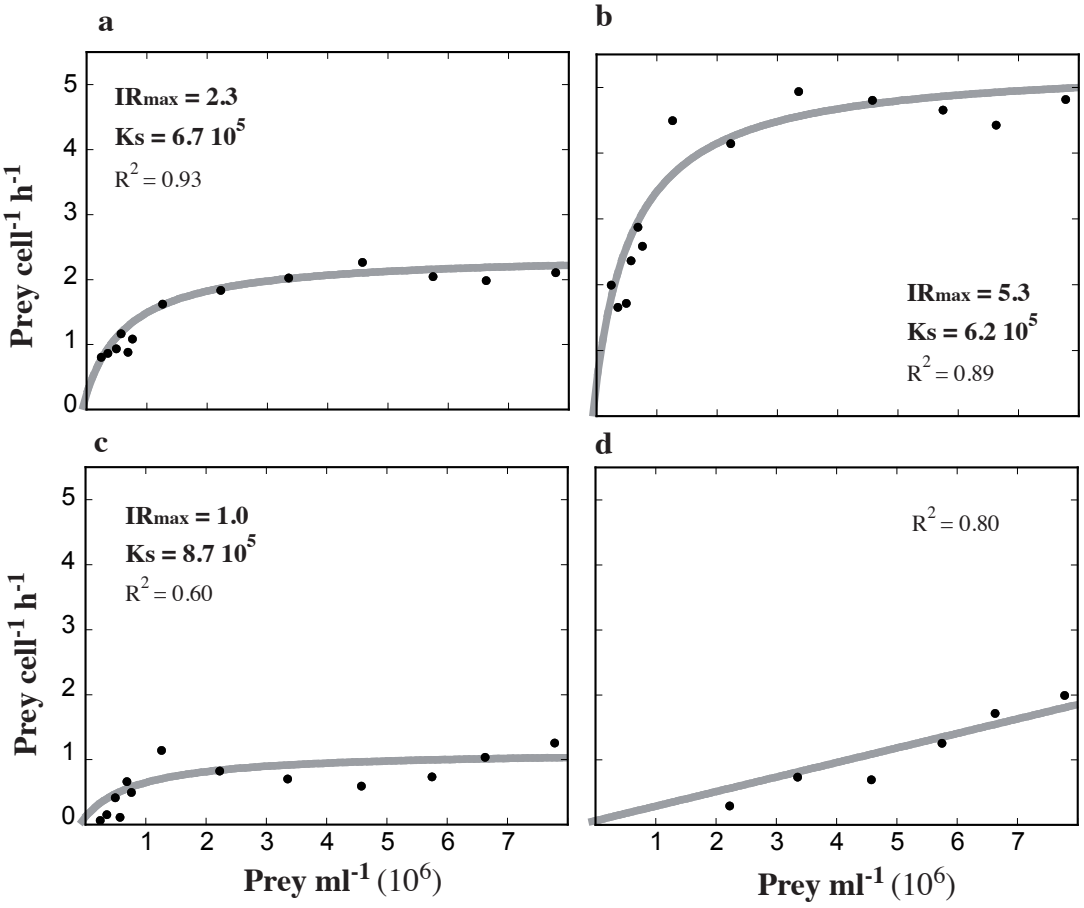


Figure 3. Functional responses (relationship of ingestion rates and prey abundance) of the natural community of heterotrophic flagellates (a), *Minorisa minuta* candidatus (b), uncultured MAST-4 cells (c) and *Paraphysomonas* sp. (d). Ks = half-saturation constant (prey ml^{-1}) and IR_{max}= maximum ingestion rate (prey $cell^{-1}$ h^{-1}).

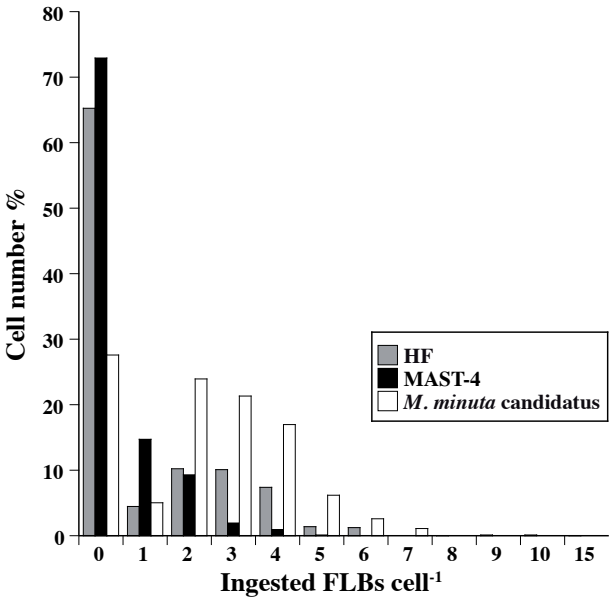


Figure 4. Percentage of cells having different number of FLBs ingested. HF in grey, MAST-4 cells in black and *Minorisa minuta* candidatus cells in white. Values are calculated at saturating food abundances (six last bottles of the gradient).

Another way to present our results was to display the effect of prey abundance on clearance rates, the volume of water processed by the predator. Highest CR should be attained at lower prey abundances, and this was clearly shown for the HF community and for *M. minuta* candidatus (Figs. 5a, b). The highest value measured in both cases, 3.4 and 8.4 nl predator^{-1} h^{-1}, respectively, was close to the maximal value estimated. Then, clearance rates decreased when increasing prey abundance, to values of 0.3 and 0.6 nl predator^{-1} h^{-1} respectively. For MAST-4 the lower prey abundance gave a noise pattern but on average we observed a decrease in clearance rate from a maximum of 0.98 nl predator^{-1} h^{-1} to 0.15 (Fig. 5c). The clearance rate of *Paraphysomonas* sp. did not decrease by increasing prey abundance, being relatively constant around 0.25 nl predator^{-1} h^{-1} (Fig. 5d)

Discussion

A main challenge in microbial ecology is to shed light on the black box approach. Indeed, for decades the abundance and activity of microbial components (such as bacteria, bacterial grazers, picoalgae, etc...), has been studied as bulk properties, ignoring the different features of the building

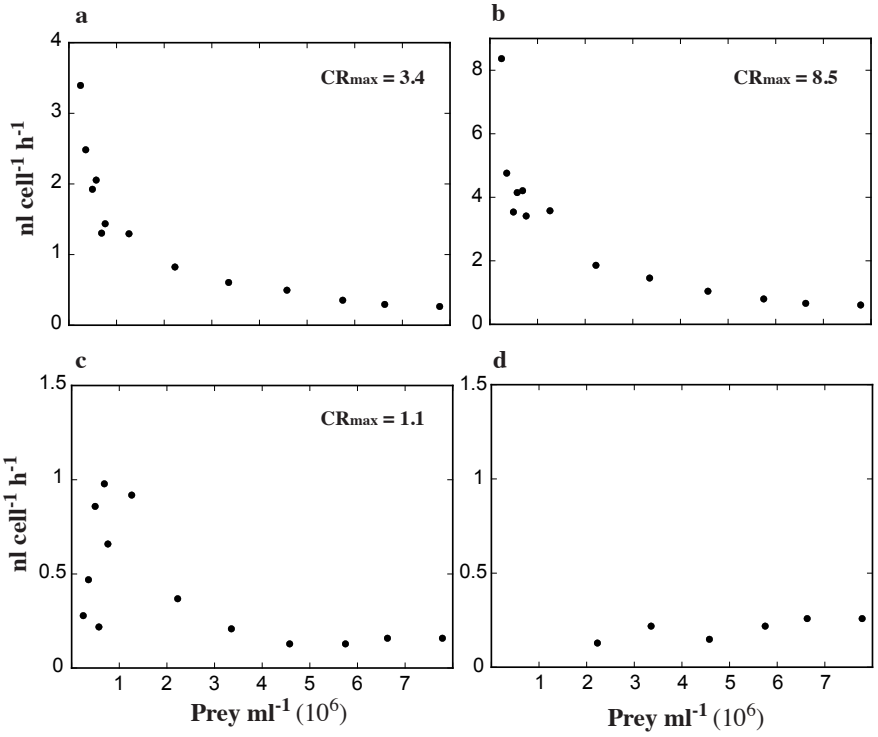


Figure 5. Relationship of clearance rates and prey abundance of the natural community of heterotrophic flagellates (a), *Minorisa minuta* candidatus (b), uncultured MAST-4 cells (c) and *Paraphysomonas* sp. (d).

specific components. A very relevant question when investigating predation in nature is the study of functional or numerical responses. These physiological features have been studied in cultured strains of heterotrophic flagellates eating bacteria (summarized in Table 2). However, to our knowledge, there are no studies analyzing functional responses of natural assemblages, neither of uncultured flagellates. This was the motivation of our study.

Our grazing experiments were based in short-term incubations with FLBs, later counted inside protist food vacuoles by epifluorescence. This technique has been highly used (Jezbera et al 2005, Simek and Chrzanowski 1992, Vaqué et al 1992) and is the best to calculate specific ingestion rates for taxonomic classes of protists. Major problems associated with this approach are the negative selection against fluorochrome labeled and heat-killed bacteria (Fu et al 2003, Landry et al 1991, Massana et al 2009), egestion of vacuole content because fixation (Sanders et al 1989, Sieracki et al 1987), and statistical problems in obtaining reliable counts of ingested FLB at low predator densities (McManus and Okubo 1991). The first problem could had been solved by using alive monospecific bacteria as food and CARD-FISH detection in food vacuoles (Massana et al. 2009), but this would have added an extra layer of complexity to our experiment. At any rate, the understimation because using FLBs would be similar along the prey concentration gradient, therefore not affecting the shape of the functional response. To minimize the problem of low predator densities, we did our grazing experiment with a sample enriched by HF. As seen before (Massana et al 2006a), the unamended seawater incubation performed here selected for heterotrophic flagellates abundant *in situ.* In our case, a large part of the diversity detected corresponded to uncultured heterotrophic flagellates, MAST and MALV cells, whereas a significant fraction corresponded to *M. minuta* candidatus, a flagellate recently cultured using techniques that better mimic natural conditions (del Campo et al submitted). Luckily enough, some typical cultured HF also developed in our incubation, highlighting very interesting and contrasting behavior.

Perhaps the most remarkable finding of this work was that the Ks of the functional responses for the whole community and for two flagellate taxa was in a range of 6-9 10^5 prey ml^{-1} (Fig. 3). Both MAST-4 (Massana et al 2006b, Rodríguez-Martínez et al 2012) and *M. minuta* candidatus (del Campo et al submitted) represent heterotrophic flagellates that are widely distributed and abundant in natural marine assemblages. Interestingly, they appear very well adapted to the bacterial abundances of marine planktonic environments, typically around 10^6 bacteria ml^{-1} (Fuhrman and Hagström 2008). For instance, in the oligotrophic coastal system sampled here, the average bacterial concentration along the complete year is ca. 0.9 10^6 bacteria ml^{-1} (Alonso-Sáez et al 2008, Boras et al 2009). In contrast, the Ks of cultured heterotrophic flagellates is typically at least one order of magnitude higher, ranging from 1.1 to 45 10^6 bacteria ml^{-1} (Table 2). These higher Ks of traditional cultured flagellates are the expected values for organisms that grow efficiently in rich media, and at the same time establish an obvious limitation for their development at the prevailing low in situ bacterial abundances. The results shown here for *Paraphysomonas* sp. agree with this scenario, since this species exhibited a low grazing capacity and no food saturation in the range of prey abundance tested.

The feeding rates of *M. minuta* candidatus, both IR_{max} (Fig. 3) and CR_{max} (Fig. 5), were much higher than those of MAST-4. This probably explains the higher measured growth rates of Minorisa (1.56 d^{-1}; del Campo et al submitted), than MAST-4 (0.62 d^{-1} average; Massana et al 2006a). Using this type of prey, and the concentration range assayed, *M. minuta* results to be very efficient in filtering water (maximal clearance of 8.5 nl h^{-1}), therefore being well adapted to low prey abundances. On the other hand, MAST-4 is less voracious than the whole HF assemblage. One reason could be that MAST-4 may select negatively against heat-killed FLBs, preferring live bacteria in good physiological state (Massana et al 2009), although is it likely that this holds true for most HF taxa in the natural assemblage (Fu et al 2003, Landry et al 1991). In addition, MAST-4 could be adapted to graze on a specific prey, as the small *Pelagibacter ubique* that has been detected within

Table 2. A comparison of maximum specific growth rates (μ_{max}, h^{-1}), maximum ingestion rates (IR_{max}, prey $cell^{-1}$ h^{-1}) and half-saturation constants (Ks, 10^6 prey ml^{-1}) for various cultured flagellate species and prey types

Flagellate cultures	Prey	μ_{max}	IR_{max}	Ks	Ref. [a]
Actinomonas mirabilis	*Pseudomonas sp.*	0.250		1.4	*1*
Bodo designis	*B1*	0.160		3.4	*2*
	Aeromonas sp.	0.120		8.8	*3*
Ciliophrys infusionum	*B1*	0.045		45.0	*2*
Codosiga gracilius	*B1*	0.052		9.7	*2*
Diaphanoeca grandis	*Pseudomonas sp.*	0.120		2.4	*4*
Jakoba libera	*Aeromonas sp.*	0.080		5.3	*3*
	B1	0.036		5.4	*2*
Monosiga sp.	*Pseudomonas sp.*	0.170		13.5	*5*
Ochromonas sp.	*Pseudomonas sp.*	0.190		19.0	*5*
Paraphysomonas vestita	*Pseudomonas sp.*	0.230		14.9	*5*
P. Imperforata	*B1*	0.210		1.1	*2*
	Aeromonas sp.	0.120		4.4	*3*
	Vibrio sp.	0.220		13.0	*6*
Pleuromonas jaculans	*Pseudomonas sp.*	0.160		38.6	*5*
Pseudobodo tremulans	*Pseudomonas sp.*	0.150		8.4	*5*
Stephanoeca diplocostata	*Pseudomonas sp.*	0.076		6.8	*7*
	B1	0.035		2.3	*2*
Cafeteria roenbergensis	*Photobacterium angustum*	0.260		5.8	*8*
	Vibrio vuinificus	0.210		2.7	*8*
	Sphingopyxis alaskensis	0.240		7.4	*8*
Cafeteria sp	*Mixed bacterial communities*	0.041		8.7	*9*
	Flavobacterium sp.	0.041		9.1	*9*
	Alteromonas sp.	0.040		9.2	*9*
	Pseudomonas sp.	0.040		9.1	*9*
Jakoba libera	*Mixed bacterial communities*	0.024		5.1	*9*
	Flavobacterium sp.	0.032		3.7	*9*
	Alteromonas sp.	0.004		9.5	*9*
	Pseudomonas sp.	0.041		1.4	*9*
Poterioochromonas malhamensis	*Polinucleobacter*	0.042		18.2	*10*
	Listonella pelagia	0.071		1.5	*10*
Spumella sp.	*Polinucleobacter*	0.083		20.5	*10*
		0.096		22.0	*11*
	Listonella pelagia	0.100		1.2	*10*
		0.117		1.2	*11*
	Bodo	0.161		2.4	*12*
		0.183		3.9	*12*
			72.8	7.4	*12*
			64.6	3.5	*12*

[a] References: *1*) Fenchel 1982, *2*) Eccleston-Parry and Leadbeater 1994, *3*) Hammond 1991, *4*) Andersen 1989, *5*) Fenchel 1982, *6*) Edwards 1989, *7*) Geider and Leadbeater 1988, *8*) Anderson et al 2011, *9*) Mohapatra and Fukami 2004, *10*) Boenigk et al 2006, *11*) Pfandl and Boenigk 2006, *12*) Jürgens 1995

MAST-4 through a single cell sequencing (Martínez-García et al 2011), therefore occupying a different ecological niche than *M. minuta*. The low feeding rates for MAST-4 measured here are consistent with previous estimates obtained with FLB, with clearance rates of 0.7 nl $predator^{-1}$ h^{-1} and ingestion rates of 1.0-1.5 bacteria $predator^{-1}$ h^{-1} (Massana et al 2009). Finally, *Paraphysomonas* spp. are typically retrieved from cultures where they feed on large bacteria at very high densities. Here, these flagellates did not achieve saturation of the ingestion rate even at the highest prey concentrations assayed (near 10^7 prey ml^{-1}). With these data we should expect a high Ks, similar to other cultured flagellates (Table 2). Studies with *Paraphysomonas* species have shown different results, including a strain that has relatively low Ks, 1.1 10^6 bacteria ml^{-1} (Eccleston-Parry and Leadbeater 1994) and another strain that ceased to multiply when the prey abundance fell below ~2 10^6 cells ml^{-1} (Ishigaki and Sleigh 2001).

Our results seemed to agree more with the later case, since our *Paraphysomonas* sp. did not start to graze until bacterial abundances were above 2.2 10^6 (Fig. 3).

In conclusion, we have shown here that the natural assemblages of heterotrophic flagellates presented a functional response with a Ks of 6.7 10^5 bacteria ml^{-1}, which is much lower than the traditional cultured flagellates. This suggests that they are well adapted to the low in situ marine bacterial abundance. Inside this mixed community, there are different taxa with different functional responses, therefore having different ecological niches, perhaps even with different filtration strategies and prey preferences. We have seen that *Minorisa minuta* candidatus was well adapted to low prey abundances (Ks of 6.2 10^5) being very efficient in filtering water (CR_{max} 8.5 nl h^{-1}) and ingesting FLBs (IR_{max} 5.3 prey h^{-1}). The MAST-4 group was less voracious but was also well adapted to the typical bacterial abundances of marine planktonic environments (Ks of 8.7 10^5 bacteria ml^{-1}). In contrast, *Paraphysomonas* sp. was food limited all along the prey gradient tested, suggesting a general poor performance in natural marine planktonic environments. Our study sets the basis for the fundamental differences between cultured and uncultured bacterial grazers.

Acknowledgements

Funding has been provided by ESTRAMAR (CTM2004-12631/MAR, MEC) and FLAME (CGL2010-16304, MICINN, Spain) projects to R.M. and by a F.P.I. fellowship from the Spanish Ministry of Education and Science to R.R.M. We thank R. Terrado for the MAST-7 CARD-FISH, V. Balagué for molecular help and E. Sainz for advices.

References

Alonso-Sáez L, Vázquez-Domínguez E, Cardelús C, Pinhassi J, Sala MM, Lekunberri I *et al* (2008). Factors Controlling the Year-Round Variability in Carbon Flux ThroughBacteria in a Coastal Marine System. *Ecosystems* **11:** 397-409.

Andersen P (1989). Functional biology of the Choanoflagellate Diaphanoeca grandis Ellis. *Mar Microb Food Webs* **3:** 35-50.

Anderson R, Kjelleberg S, Mcdougald D, Jürgens K (2011). Species-specific patterns in the vulnerability of carbon-starved bacteria to protist grazing. *Aquat Microb Ecol* **64:** 105-116.

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

Boenigk J, Pfandl K, Hansen P (2006). Exploring strategies for nanoflagellates living in a'wet desert'. *Aquat Microb Ecol* **44:** 71.

Boras JA, Sala MM, Vázquez-Domínguez E, Weinbauer MG, Vaqué D (2009). Annual changes of bacterial mortality due to viruses and protists in an oligotrophic coastal environment (NW Mediterranean). *Environ Microbiol* **11:** 1181-1193.

del Campo J (2011). Culturing bias in the study of marine heterotrophic flagellates diversity. *PhD thesis University of Barcelona*.

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

Eccleston-Parry J, Leadbeater B (1994). A comparison of the growth kinetics of six marine heterotrophic nanoflagellates fed with one bacterial species. *Mar Ecol Prog Ser* **105:** 167-177.

Edwards AM (1989). Heterotrophic chrysophytes; their role in energy and carbon turnover in the sea. *Ph D thesis, University of Birmingham*.

Elwood HJ, Olsen GJ, Sogin ML (1985). The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates Oxytricha nova and Stylonychia pustulata. *Mol Biol Evol* **2:** 399-410.

Fenchel T (1980). Suspension Feeding in Ciliated Protozoa: Functional Response and Particle Size Selection. *Microb Ecol* **6:** 1-11.

Fenchel T (1982). Ecology of heterotrophic microflagellates. II. Bioenergetics and growth. *Mar Ecol Prog Ser* **8:** 225-231.

Fenchel T (1986). The Ecology of Heterotrophic Microflagellates. *Adv Microb Ecol* **9:** 57-97.

Fenchel T (1987). Ecology of Protozoa. The Biology of Free-living Phagotrophic Protists. *Brock/Springer Series in Contemporary Bioscience* 47 figs.: 197pp.

Fu Y, O'Kelly C, Sieracki M, Distel DL (2003). Protistan grazing analysis by flow cytometry using prey labeled by in vivo expression of fluorescent proteins. *Appl Environ Microbiol* **69:** 6848-6855.

Fuhrman JA, Hagström A (2008). Bacteria and archaeal community structure and its patterns. *In Kirchman DL (ed) Microbial Ecology of the Oceans Wiley Series***:** 45-80.

Geider RJ, Leadbeater BSC (1988). Kinetics and energetics of growth of the marine choanoflagellate Stephanoeca diplocostata. . *Mar Ecol Prog Ser* **47:** 169-177.

Hammond SJ (1991). Comparative growth characteristics of heterotrophic flagellates. *M Sc thesis, Bristol University***:** 26-37.

Henriksen CI, Saiz E, Calbet A, Hansen BW (2007). Feeding activity and swimming patterns of Acartia grani and Oithona davisae nauplii in the presence of motile and non-motile prey. *Mar Ecol Prog Ser* **331:** 119-129.

Holling CS (1959). The Components of Predation as Revealed by a Study of Small-Mammal Predation of the European Pine Sawfly. *Can Entomol* **91:** 293-320.

Isari S, Saiz E (2011). Feeding performance of the copepod Clausocalanus lividus (Frost and Fleminger, 1968). *J Plankton Res* **33:** 715-728.

Ishigaki T, Sleigh M (2001). Grazing characteristics and growth efficiencies at two different temperatures for three nanoflagellates fed with Vibrio bacteria at three different concentrations. *Microb Ecol* **41:** 264-271.

Jeong HJ, Yoo YD, Park JY, Song JY, Kim ST, Lee SH *et al* (2005). Feeding by phototrophic red-tide dinoflagellates: five species newly revealed and six species previously known to be misotrophic. *Aquat Microb Ecol* **40:** 133-150.

Jezbera J, Hornák K, Simek K (2005). Food selection by bacterivorous protists: insight from the analysis of the food vacuole content by means of fluorescence in situ hybridization. *FEMS Microbiol Ecol* **52:** 351-363.

Jonsson R (1986). Particle size selection, feeding rates and growth dynamics of marine planktonic oligotrichous ciliates (Ciliophora: Oligotrichina). *Mar Ecol Prog Ser* **33:** 265-277.

Jürgens K (1995). Die Bedeutung heterotropher Nanoflagellaten als Bakterienkonsumenten sowie deren Regulation durch Prädation und Ressourcen. *PhD thesis***:** 1-1.

Jürgens K, Simek K (2000). Functional response and particle size selection of Halteria cf. grandinella, a common freshwater oligotrichous ciliate. *Aquat Microb Ecol* **22:** 57-58.

Jürgens K, Matz C (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *A Van Leeuw* **81:** 413-434.

Jürgens K, Massana R (2008). Protist grazing on marine bacterioplankton. *Microbial ecology of the oceans***:** 383-441.

Kim JS, Jeong HJ (2004). Feeding by the heterotrophic dinoflagellates Gyrodinium dominans and G. spirale on the red-tide dinoflagellate Prorocentrum minimum. *Mar Ecol Prog Ser* **280:** 85-94.

Landry M, Lehner-Fournier J, Sundstrom J, Fagerness V, Selph K (1991). Discrimination between living and heat-killed prey by a marine zooflagellate, Paraphysomonas vestita (Stokes). *J Exp Mar Biol Ecol* **146:** 139-151.

Lim E, Dennett M, Caron D (1999). The ecology of Paraphysomonas imperforata based on studies employing oligonucleotide probe identification in coastal water samples and enrichment cultures. *Limnol Oceanogr* **44:** 37-51.

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

Martínez-García M, Brazel D, Poulton… N (2011). Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. *ISME J* **6:** 703-707

Massana R, Stumm CK, Pedrós-Alió C (1994). Effects of temperature, sulfide, and food abundance on growth and feeding of anaerobic ciliates. *Appl Environ Microbiol* **60:** 1317-1324.

Massana R, DeLong EF, Pedrós-Alió C (2000). A few cosmopolitan phylotypes dominateplanktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol* **66:** 1777-1787.

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

Massana R, Guillou L, Terrado R, Forn I, Pedrós-Alió C (2006a). Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol* **45:** 171-180.

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

Massana R, Del Campo J, Dinter C, Sommaruga R (2007). Crash of a population of the marine heterotrophic flagellate Cafeteria roenbergensis by viral infection. *Environ Microbiol* **9:** 2660-2669.

Massana R, Unrein F, Rodríguez-Martínez R, Forn I, Lefort T, Pinhassi J *et al* (2009). Grazing rates and functional diversity of uncultured heterotrophic flagellates. *ISME J* **3:** 588-596.

McManus GB, Okubo A (1991). On the use of surrogate food particles to measure protistan ingestion. *Limnol Oceanogr* **36:** 613-617.

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

Mohapatra B, Fukami K (2004). Comparison of the numerical grazing response of two marine heterotrophic nanoflagellates fed with different bacteria. *J Sea Res* **52:** 99-107.

Montagnes DJS, Berges JA (2004). Determining Parameters of the Numerical Response. *Microb Ecol* **48:** 139-144.

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.

Pernthaler J, Glöckner F, Schönhuber W, Amann R (2001). Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. *Meth Microb* **30:** 207-226.

Pernthaler J (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* **3:** 537-546.

Pfandl K, Boenigk J (2006). Stuck in the mud: suspended sediments as a key issue for survival of chrysomonad flagellates. *Aquat Microb Ecol* **45:** 89.

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

Porter K, Feig Y (1980). The use of DAPI for identifying and counting aquatic microflora. *Limnol Oceanogr* **25:** 943-948.

Rodríguez-Martínez R, Rocap G, Logares R, Romac S, Massana R (2012). Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4). *Mol Biol Evol***:** e-pub ahead of print 13 December 2011, doi: 2010.1093/molbev/msr2303.

Sanders RW, Porter KG, Bennett SJ, DeBiase AE (1989). Seasonal patterns of bacterivory by flagellates, ciliates, rotifers, and cladocerans in a freshwater planktonic community. *Limnol Oceanogr* **34:** 673-687.

Sherr BF, Sherr EB, Fallon RD (1987). Use of monodispersed, fluorescently labeledbacteria to estimate in situ protozoan bacterivory. *Appl Environ Microbiol* **53:** 958-965.

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

Sieracki ME, Haas LW, Caron DA, Lessard EJ (1987). Effect of fixation on particle retention by microflagellates: underestimation of grazing rates. *Mar Ecol Prog Ser* **38:** 251-258.

Simek K, Chrzanowski TH (1992). Direct and Indirect Evidence of Size-Selective Grazing on Pelagic Bacteria by Freshwater Nanoflagellates. *Appl Environ Microbiol* **58:** 3715-3720.

Strom S (2000). Bacterivory: Interactions between bacteria and their grazers. *In: Kirchmand DL (ed) Microbial ecology of the oceans* Wiley-Liss Inc: New York: 351-386.

Unrein F, Massana R, Alonso-Sáez L, Gasol J (2007). Significant year-round effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal system. *Limnol Oceanogr* **52:** 456-469.

Vaqué D, Pace ML, Findlay S, Lints D (1992). Fate of bacterial production in a heterotrophic ecosystem: grazing by protists and metazoans in the Hudson Estuary. *Mar Ecol Prog Ser* **89:** 155-163.

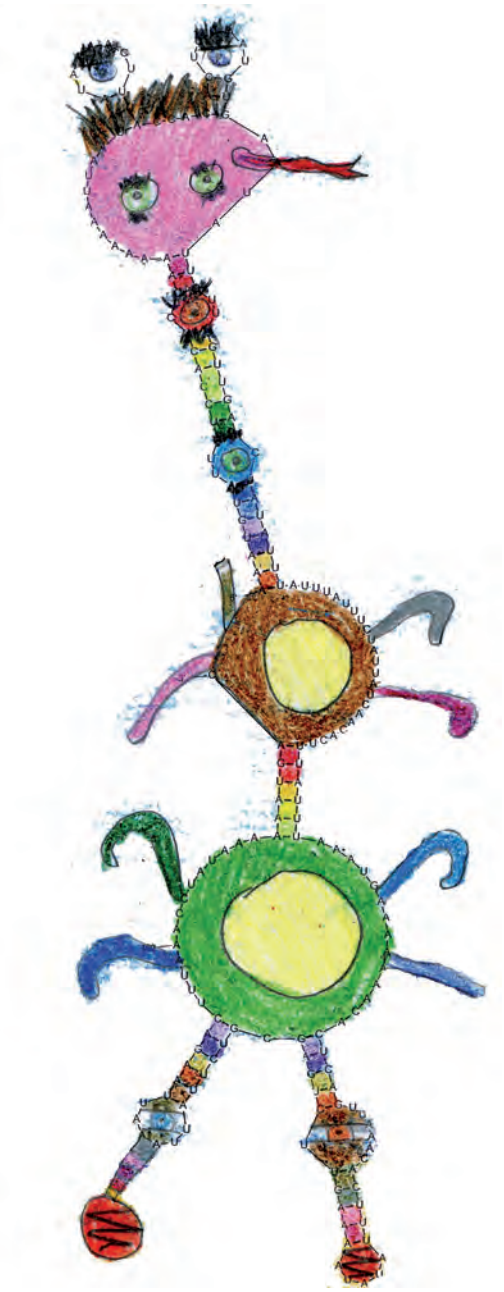
Vaqué D, Gasol JM, Marrasé C (1994). Grazing rates on bacteria: the significance of methodology and ecological factors. *Mar Ecol Prog Ser* **109:** 263-274.

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

Vázquez-Domínguez E, Peters F, Gasol JM, Vaqué D (1999). Measuring the grazing losses of picoplankton: methodological improvements in the use of fluorescently labeled tracers combined with flow cytometry. *Aquat Microb Ecol* **20:** 119-128.

Chapter 4

Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4)



Rodríguez-Martínez R, Rocap G, Logares R, Romac S, Massana R (2012). Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4). *Molecular Biology and Evolution***:** e-pub ahead of print 13 December 2011, doi: 2010.1093/molbev/msr2303.

Low Evolutionary Diversification in a Widespread and Abundant Uncultured Protist (MAST-4)

Raquel Rodríguez-Martínez,*[,1] Gabrielle Rocap,[2] Ramiro Logares,[1] Sarah Romac,[3] and Ramon Massana*[,1]

[1]Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain

[2]School of Oceanography, University of Washington

[3]Station Biologique de Roscoff, Centre National de la Recherche Scientifique (CNRS) et Universite Pierre et Marie Curie (UPMC), Roscoff, France

***Corresponding author:** E-mail: raquelr@icm.csic.es; ramonm@icm.csic.es.

Associate editor: Douglas Crawford

Abstract

Recent culture-independent studies of marine planktonic protists have unveiled a large diversity at all phylogenetic scales and the existence of novel groups. MAST-4 represents one of these novel uncultured lineages, and it is composed of small (~2 μm) bacterivorous eukaryotes that are widely distributed in marine systems. MAST-4 accounts for a significant fraction of the marine heterotrophic flagellates at the global level, playing key roles in the marine ecological network. In this study, we investigated the diversity of MAST-4, aiming to assess its limits and structure. Using ribosomal DNA (rDNA) sequences obtained in this study (both pyrosequencing reads and clones with large rDNA operon coverage), complemented with GenBank sequences, we show that MAST-4 is composed of only five main clades, which are well supported by small subunit and large subunit phylogenies. The differences in the conserved regions of the internal transcribed spacers 1 and 2 (ITS1 and ITS2) secondary structures strongly suggest that these five clades are different biological species. Based on intraclade divergence, ITS secondary structures and comparisons of ITS1 and ITS2 trees, we did not find evidence of more than one species within clade A, whereas as many as three species might be present within other clades. Overall, the genetic divergence of MAST-4 was surprisingly low for an organism with a global population size estimated to be around 10^{24}, indicating a very low evolutionary diversification within the group.

Key words: MAST-4, low evolutionary diversification, uncultured protist, pyrosequencing, ITS secondary structure.

Introduction

Microbes have vital roles for the functioning of the biosphere (Falkowski et al. 2008), but currently, we are far from having acceptable estimates of their diversity. Furthermore, it is unclear how microbial diversity is distributed in space and time, and how diversity ranks are translated into ecologically meaningful interactions or processes. The marine protists of very small size, the picoeukaryotes, are among the underexplored microbes with large ecological importance (Massana 2011). Picoeukaryotes have key ecological roles in the oceans as primary producers, bacterial grazers, or parasites. They are found in all planktonic marine samples at concentrations ranging between 10^3 and 10^4 cells ml^{-1}. During the last 10 years, molecular tools based on sequencing environmental 18S ribosomal DNA (rDNA) genes have revealed a wide diversity of microeukaryote assemblages as well as the existence of novel and uncultured lineages (Díez et al. 2001; López-García et al. 2001; Moon-van der Staay et al. 2001). Still, most of this diversity remains poorly known.

The assignation of this novel and uncultured diversity to taxonomic groups is a challenging task. An approach to address this issue is to explore the correspondence between genetic divergence and species limits using cultured strains and then use that data as a proxy to investigate species limits in uncultured strains. Studies combining molecular and morphological data have been done within different taxonomic groups, such as prasinophytes (Slapeta et al. 2006), prymnesiophytes (Lange et al. 2002), diatoms (Amato et al. 2007; Evans et al. 2007; Casteleyn et al. 2008; Rynearson et al. 2009; Sorhannus et al. 2010), and dinoflagellates (Montresor et al. 2003; Litaker et al. 2007; Lowe et al. 2010). Gene markers used in the mentioned studies generally involve the 18S rDNA and other more variable genes (such as *rbcL* or *cox1*) since the former may be too conserved to differentiate among related but different species (Edvardsen et al. 2000; Logares et al. 2007). For uncultured protists detected in 18S rDNA surveys, the obvious loci for increasing phylogenetic resolution are the contiguous internal transcribed spacer (ITS) regions (ITS1 and ITS2). The above-mentioned functional genes, proposed as more robust phylogenetic markers (Álvarez and Wendel 2003), are currently inaccessible for uncultured microorganisms. ITS regions are noncoding loci that display high sequence variability but also key functionally constrained positions since transcripts need to fold into a secondary structure to permit their own splicing and the correct processing of the rDNA genes (Schlötterer et al. 1994; Côté et al. 2002). They have been proposed as the best tool for barcoding in diatoms (Moniz and Kaczmarska 2010) and are useful for species and genus phylogenetic inferences (Coleman 2003).

The secondary structure of the ITS2 region has been used for delimiting biological species. Compensatory base changes (CBCs) in particular regions of the secondary structure have been associated with sexual incompatibility (Coleman 2007, 2009). Taxa exhibiting at least one CBC in these conserved regions most likely belong to different biological species (Amato et al. 2007). Significant progress has been made in identifying such relevant positions in Volvocaceae, Haliotis, and Fagales (Coleman 2000, 2003; Coleman and Vacquier 2002; Müller et al. 2007). In addition, this hypothesis has been subjected to a large-scale testing using the ITS2 database containing 100,000 secondary structures (Schultz et al. 2006; Selig et al. 2007) and has been supported in 93% of the cases (Müller et al. 2007). However, this is a one-way diagnostic; a lack of CBCs does not mean that organisms are members of the same species.

In this study, we investigate an important and poorly known uncultured picoeukaryote group, the MAST-4 (Massana et al. 2004). This protist group is widespread in surface marine waters (except polar systems), where it represents approximately 9% of heterotrophic flagellates (Massana, Terrado, et al. 2006; Rodríguez-Martínez et al. 2009). Although MAST-4 remains uncultured, it is easily detected in environmental samples using molecular tools. So far, only the 18S rDNA of MAST-4 has been sequenced. To understand the genetic structure and evolutionary patterns of this uncultured model picoeukaryote, we sequenced a large fragment of the rDNA operon, including the ITS region and the beginning of the 28S (using Sanger sequencing) as well as the V4 region of the 18S (454 pyrosequencing). We have also compiled and analyzed all publicly available MAST-4 18S rDNA sequences. The emerging scenario is that despite being hugely abundant and widely distributed, this lineage has experienced a limited evolutionary diversification. A more detailed study of the biogeography of the group will appear elsewhere (Rodríguez-Martínez R, unpublished data).

Materials and Methods

Compilation of Published MAST-4 Sequences

BLAST (basic local alignment search tool) searches against NCBI-nr were done using as seeds different regions (1–500, 501–1000, and 1001–1700) of the published MAST-4 18S rDNA sequences ME1.19, ME1.20 (Díez et al. 2001), UEPACCp4, UEPAC05Cp2 (Worden 2006), and SSRPD78 (Not et al. 2007). Best hits (sorted in decreasing order by identity) were selected until a sequence classified to another group appeared. The retrieved 134 sequences were screened to remove chimeras, identical sequences from the same study, and sequences that did not cover the V4–V5 regions, leaving 72 sequences that aligned at least ~550 bp. Seven partial GenBank sequences, of clones from our own libraries, were completely sequenced. This resulted in 17 complete MAST-4 sequences ("small subunit [SSU]–complete" data set). The remaining 55 partial sequences (between 461 and 1,266 bp) formed the "SSU-partial" data set.

Retrieval of MAST-4 Using 454 Pyrosequencing

Seawater samples were collected through the BioMarKs consortium (http://www.biomarks.org/) in several European coastal stations (offshore Oslo, Naples, Blanes, Roscoff, Gijon, and Varna) with Niskin bottles attached to a conductivity, temperature, and depth rosette at surface and deep chlorophyll maximum depths. Water samples were prefiltered through 20 μm. Afterward, they were sequentially filtered through 3 and 0.8 μm 142 mm polycarbonate filters. Filters were flash frozen and stored at −80 °C. Total DNA and RNA were extracted simultaneously from the same filter using the NucleoSpin RNA L kit (Macherey-Nagel) and quantified using a Nanodrop ND-1000 Spectrophotometer. Extract quality was checked on a 1.5% agarose gel. To remove contaminating DNA from RNA, we used the TurboDNA kit (Ambion). Extracted RNA was immediately reverse transcribed to DNA using the RT Superscript III_random primers kit (Invitrogen). The universal primers TAReuk454FWD1 and TAReukREV3 were used to amplify the V4 region (~380 bp) of eukaryotic 18S rDNA (Stoeck et al. 2010). The primers were adapted for 454 using the manufacturers specifications and had the configuration A adapter Tag (8 bp)–forward primer and B adapter–reverse primer. Polymerase chain reactions (PCRs) were performed in 25 μl and consisted 1× MasterMix Phusion High-Fidelity DNA Polymerase (Finnzymes), 0.35 μM of each primer, and 3% dimethyl sulfoxide. We added a total of 5 ng of template DNA/cDNA to each PCR reaction. PCR reactions consisted of an initial denaturation step at 98 °C during 30 s, followed by 10 cycles of 10 s at 98 °C, 30 s at 53 °C, and 30 s at 72 °C, and afterward by 15 cycles of 10 s at 98 °C, 30 s at 48 °C, and 30 s at 72 °C. Amplicons were checked in a 1.5% agarose gel for successful amplification. Triplicate amplicons were pooled and purified using the NucleoSpin Extract II (Macherey-Nagel). Purified amplicons were eluted in 30 μl of elution buffer and quantified again using a Nanodrop ND-1000 Spectrophotometer. The total final amount of pooled amplicons for 454 tag sequencing was approximately of 5 μg. Amplicon sequencing was carried out on a 454 GS FLX Titanium system (454 Life Sciences, USA) installed at Genoscope (http://www.genoscope.cns.fr/spip/, France).

Only reads having exact forward and reverse primers and an estimated error of $\leq$0.1% were kept (682,390 reads) and were annotated using a custom made and curated 18S rDNA database (Guillou L, unpublished data). Sequences with the MAST-4 as the closest group (similarity >90%) were extracted (2,808 reads). Identical reads were removed with Mothur (Schloss et al. 2009) and then clustered at 0.0049 distance, resulting in 169 unique sequences. Subsequently, 81 chimeras were removed with the Chimera Slayer algorithm (Haas et al. 2011) as implemented in Mothur, using a custom-made protist database as a template. Sixteen remaining chimeras were removed manually, after partial sequence BLASTs against NCBI-nr. The final 72 sequences of ~380 bp formed the "SSU-pyrosequencing" data set.

Clone Libraries Covering the 18S to 28S rDNA Regions

Offshore surface samples were selected from separate oceanographic cruises in the Indian Ocean (IND70),

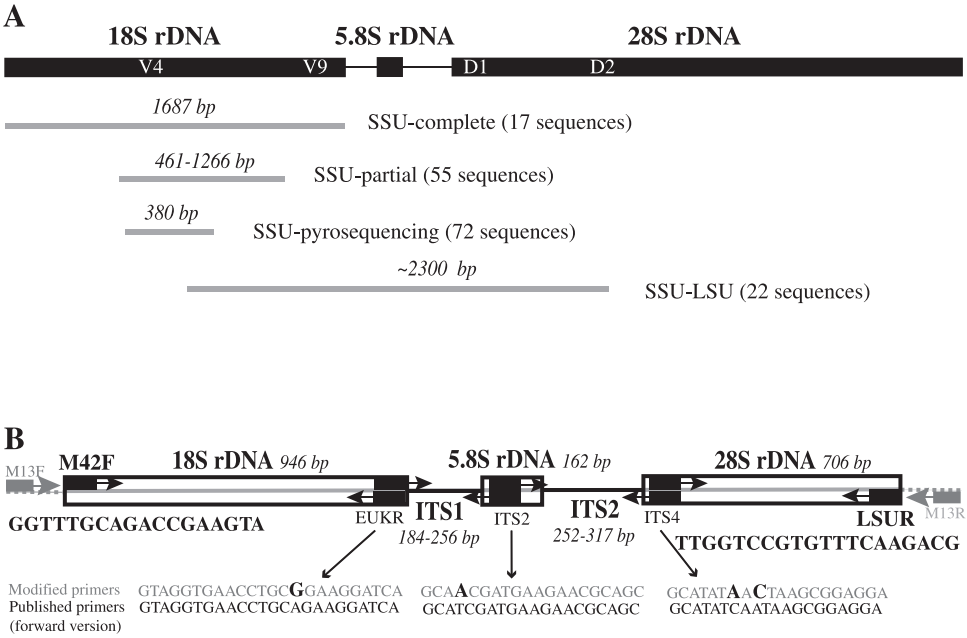


Fig. 1. Map of the rDNA operon showing the covered region of each sequence data set (*A*) and a detailed diagram of the ~2,300 bp MAST-4 rDNA amplicons (*B*). For this last data set, the positions and sequences of primers used are presented.

Sargasso Sea (BE3), North Pacific (WE7), and Mediterranean Sea (BL43, taken on August 2004). These sites correspond to stations INO3, ATL7, PAC1, and MED as shown in Massana, Terrado, et al. (2006). The 0.2–3 μm microbial fraction of surface seawater was collected by peristaltic filtration. A fifth sample was selected (OA4), derived from the peak of heterotrophic flagellates in an unamended incubation from the MED station (March 2006) processed as in Massana, Guillou, et al. (2006). DNA extraction was done using enzymatic and sodium dodecyl sulfate digestion plus phenol purification (Massana et al. 2000).

PCR amplification was done with the MAST-4-specific primer M42F (5′-GGTTTGCAGACCGAAGTA-3′) located in the 18S rDNA (after the V4 region) and the universal eukaryotic primer LSUR (5′-TTGGTCCGTGTTTCAAGACG-3′) located in the 28S rDNA (in the middle of the D2 region) (Jerome and Lynn 1996). Primer M42F is the reverse sequence of the FISH probe NS4 (Massana et al. 2002) and has a good specificity for MAST-4 (Massana, Terrado, et al. 2006). Primer LSUR matches 183 of 187 stramenopile large subunit (LSU) sequences extracted from SILVA database (Pruesse et al. 2007). Primers were checked for formation of primer dimers, GC content, and theoretical melting temperature in the website www.operon.com, using the Oligo Analysis & Plotting Tool. This primer set gave an amplicon size of ~2,300 bp covering the end of the SSU gene, the whole ITS region (ITS1-5.8S-ITS2), and the beginning of the LSU gene (fig. 1).

The PCR mixture (25 μl) contained 1 μl of DNA template, 0.5 μM of each primer, 200 μM of each dNTP, 3 mM $MgCl_2$, 1.25 units of a proofreading *Taq* polymerase (ACCUZYME), and the enzyme buffer. PCR cycling, carried out in an MJ Research thermal cycler, was initial denaturation at 94 °C for 5 min, 30 cycles with denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min, and extension at 72 °C for 3 min and a final extension at 72 °C for 10 min. We added a reconditioning PCR step to eliminate heteroduplexes from mixed-template PCR products (Thompson et al. 2002). The PCR reaction was diluted 5-fold into fresh reaction mixture and cycled three times as above. We tested the $MgCl_2$ concentration (from 1.5 to 3 mM) and the annealing temperature (from 57 to 64 °C) and chose the more stringent conditions giving the expected band. The PCR product from four parallel reactions per sample was pooled and reduced to 25 μl by ethanol precipitation or vacuum concentration and run in a 1% agarose gel electrophoresis. Bands of 2,000–3,000 bp were cut and purified with the QIAquick Gel Extraction kit (QIAGEN). We added 3′ A-overhangs to the final PCR product and cloned it using the TOPO-TA cloning kit (Invitrogen) with the vector (pCR4) following manufacturer's recommendations. Putative positive colonies were picked and transferred to a new Luria–Bertani (LB) plate and finally into LB-glycerol solution for frozen stocks (−80 °C).

Presence of correct insert was checked by PCR reamplification with vector primers M13F and M13R using a small aliquot of culture as template. Amplicons with the right insert size were sequenced in both directions at the Macrogen sequencing service (Korea) with eight primers (fig. 1). After inspecting the first sequences, we modified primers EUKR, ITS2, and ITS4 for a perfect match with MAST-4 sequences (fig. 1). Chromatograms were examined with 4Peaks (A. Griekspoor and T. Groothuis, http://www.mekentosj.com), and sequences for each clone were assembled with Geneious (Drummond et al. 2010), which also allows careful inspection of chromatograms and sequence editing. A total of 22 sequences from clone libraries were

used for subsequent analyses ("SSU-LSU" data set). These sequences were aligned with MAFFT v6.853 (Katoh and Toh 2008) with the E-INS-I algorithm, and the alignment was inspected visually. Boundaries of rDNA genes were determined by comparison with published reference sequences belonging to closely related organisms, resulting in five separate DNA regions: 3′ end of the SSU gene (946 bp), ITS1 (184–256 bp), 5.8S gene (162 bp), ITS2 (252–317 bp), and 5′ beginning of the LSU gene (706 bp). Sequences have been deposited in GenBank under accession numbers JN836289–JN836310.

Sequence Analyses

Sequences from the SSU-complete data set were aligned together with a MAST-7 outgroup using MAFFT as specified above. This alignment (1688 positions) was used as a skeleton, and shorter sequences from GenBank (SSU-partial data set) or from pyrosequencing (SSU-pyrosequencing data set) were incorporated into it using the "-add" option of MAFFT. Alignments with 5.8S and 28S regions were done using *Phytophthora infestans* as outgroup (GenBank accession numbers HQ191489 and EU079637, respectively). Due to the large sequence variability in ITS regions, ITS1 and ITS2 alignments were done separately for each SSU-defined clade, using secondary structure models for alignment improvement (Rocap et al. 2002; Wang et al. 2007; Tippery and Les 2008). All these alignments were used to calculate sequence divergences (uncorrected pairwise distances) using Mothur (Schloss et al. 2009).

Maximum likelihood (ML) phylogenetic trees were reconstructed using RAxML v7.0.4 MPI version (Stamatakis 2006), using the General Time Reversible model of nucleotide substitution and a Gamma distributed rate of variation across sites (GTR+G). As suggested in RaxML, we did not estimate the proportion of invariable sites, and missing data were not considered (i.e., treated as missing data). The shape parameter (α) of the gamma distribution was estimated from the data set using default options. Phylogenies were reconstructed at both the University of Oslo Bioportal (www.bioportal.uio.no) and the Instituto Astrofísico de Canarias (IAC) computer cluster. One thousand alternative ML trees were run, and the tree with the best likelihood was selected and visualized in FigTree v1.3.1 (Rambaut 2009) or iTOL (Letunic and Bork 2007). Bootstrap analyses were run with 1,000 pseudoreplicates, and a consensus tree was constructed with MrBayes (Huelsenbeck and Ronquist 2001).

ITS1 and ITS2 Secondary Structures

ITS1 and ITS2 sequences extracted from the SSU-LSU data set were folded in mFOLD (Zuker 2003), which generates multiple possible secondary structures. We used default settings for a linear molecule with a folding temperature fixed at 37 °C and 1 M NaCl with no divalent ions for ionic conditions. The best conformation for each sequence was the one that possessed the previously defined ITS hallmarks and was also similar between related clones. This generally coincided with the minimum free energy configuration. For ITS2 models, we searched for the familiar four-helix domain seen in eukaryotic taxa, such as green algae and flowering plants (Mai and Coleman 1997), dinoflagellates (Gottschling 2004), and metazoans (Joseph et al. 1999; Coleman and Vacquier 2002; Müller et al. 2007; Wiemers et al. 2009). The core structure and hallmarks for the ITS1 secondary structure are less clear (see Discussion). Exported secondary structures in Vienna format (http://www.tbi.univie.ac.at/~ivo/RNA/) were aligned and visualized as a consensus of each clade with 4SALE version 1.5 (Seibel et al. 2006). Structural models were further analyzed for the presence of CBCs (e.g., a change of paired G-C into paired A-U) in conserved regions (Gutell et al. 1994; Coleman et al. 1998). We used the models proposed by Coleman to identify the ITS2 conserved regions having a biological meaning (Coleman 2003, 2007, 2009).

Results

Low Diversity within the MAST-4 18S rDNA Gene

The phylogenetic tree with the distinct MAST-4 18S rDNA sequences retrieved from our thorough GenBank search (SSU-complete plus SSU-partial data sets) displayed the complete MAST-4 variability published so far. MAST-4 diversity was limited to only five clades (A–E), each one containing at least two complete sequences and being well supported (except clade B) by bootstrap values above 80% (fig. 2). Only clone IND31.115 (Massana, Terrado, et al. 2006) did not belong to a given clade. The intraclade sequence divergence (calculated in the SSU-complete data set) was typically below 0.010, whereas among clades, the average divergence was 0.030 (table 1), with a maximum of 0.044. In addition, a BLAST search of MAST-4 sequences against NCBI-nr displayed a maximum of 91% similarity to the closest outgroup sequence, which belonged to MAST-7 or MAST-8. Sequences with intermediate similarity (i.e., between 91% and 96%) were chimeras.

High-throughput sequencing approaches allow deeper sampling of environmental diversity than traditional cloning and Sanger sequencing. In order to determine whether additional MAST-4 clades were present in the marine plankton, we analyzed 454 reads (~380 bp) obtained from several coastal locations around Europe using eukaryotic universal primers. After processing an initial set of 2,808 MAST-4 sequences, the 72 distinct sequences of the SSU-pyrosequencing data set were used for phylogenetic reconstructions together with a subset of the GenBank sequences (31 remaining sequences after clustering the SSU-complete and SSU-partial data sets at a 0.0049 distance) (fig. 3). Pyrosequences distributed among the five clades reported before, and, most interestingly, no additional clades appeared. The sequence IND31.115 still remained alone.

Analysis of Other rDNA Regions Support Five Main Clades

We obtained good quality sequences of ~2,300 bp (SSU-LSU data set) for 22 clones derived from four

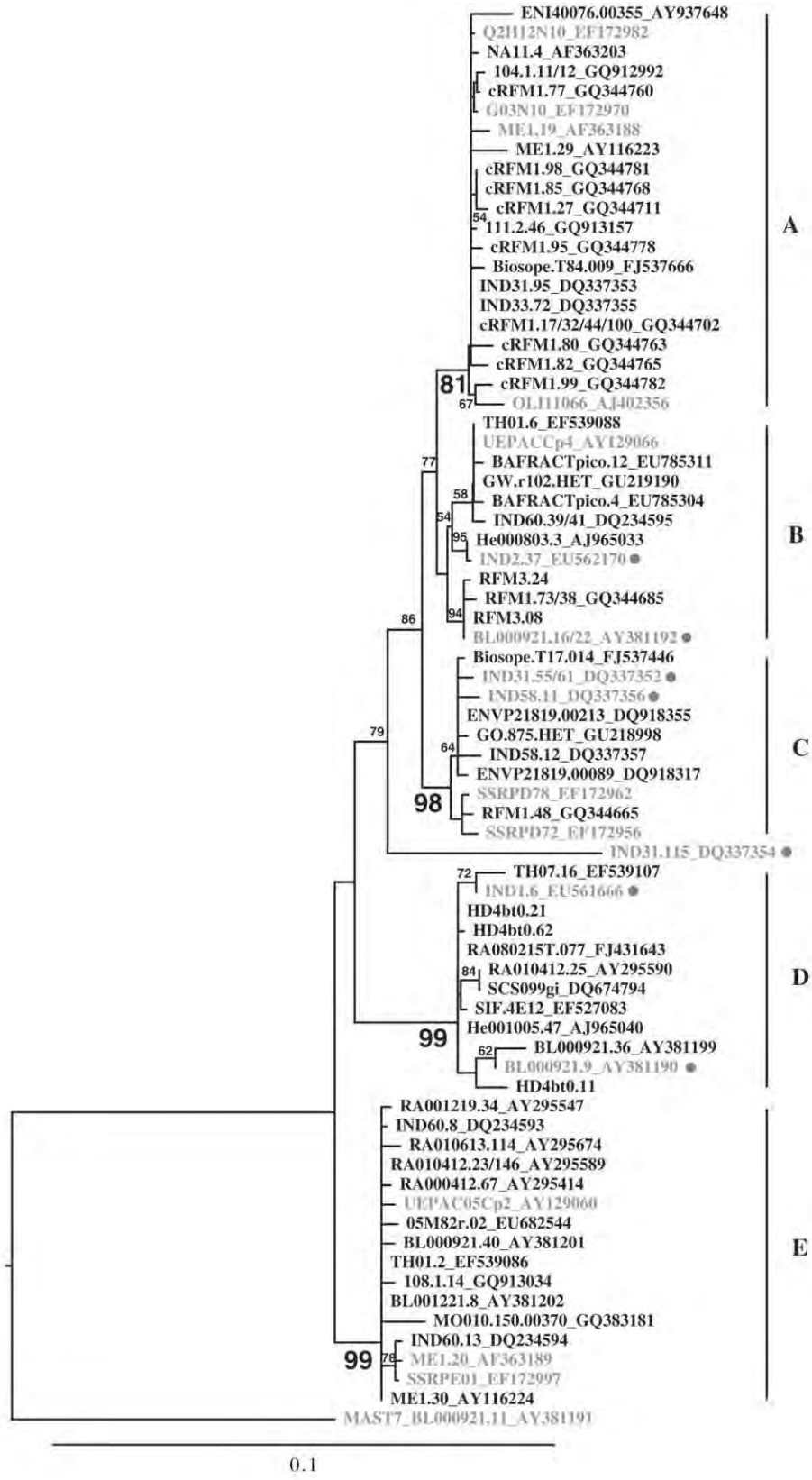


Fig. 2. ML phylogenetic tree of complete (gray) and partial (black) 18S rDNA GenBank sequences affiliating to MAST-4, showing the five main clades (labeled A–E) and their bootstrap support. Complete 18S rDNA sequenced in this study is indicated with a gray dot.

oceanographic regions, the Indian Ocean (six clones), the North Pacific (four clones), the Sargasso Sea (three clones), and the Mediterranean Sea (nine clones, four of them from an unamended enrichment). These were separated into the three genes and the two internal spacers of the rDNA operon for more exhaustive phylogenetic analyses (fig. 4). The

Table 1. Sequence Divergence (uncorrected p-distance) within and among MAST-4 Clades, Shown as Average (minimum–maximum).

	Data Set SSU-Complete		Data Set SSU-LSU						
Clade	*n*	18S rDNA	*n*	18S rDNA	5.8S rDNA	28S rDNA	ITS1	ITS2	ITS
A	4	0.007 (0.002–0.012)	7	0.001 (0–0.002)	0	0.005 (0.001–0.010)	0.044 (0.005–0.071)	0.036 (0.004–0.057)	0.031 (0.011–0.050)
B	3	0.008 (0.008–0.009)	2	0.001	0	0.021	0.204	0.197	0.163
C	4	0.008 (0.004–0.011)	6	0.005 (0–0.010)	0	0.020 (0–0.033)	0.104 (0.005–0.156)	0.121 (0–0.195)	0.078 (0.002–0.117)
D	2	0.011	1		0				
E	3	0.006 (0.002–0.008)	6	0	0	0.003 (0–0.006)	0.020 (0.004–0.032)	0.021 (0–0.031)	0.016 (0.010–0.013)
Interclade	16	0.030 (0.011–0.044)	22	0.027 (0.010–0.047)	0.024 (0–0.049)	0.161 (0.074–0.218)	0.420 (0.200–0.559)	0.418 (0.298–0.562)	0.270 (0.185–0.416)

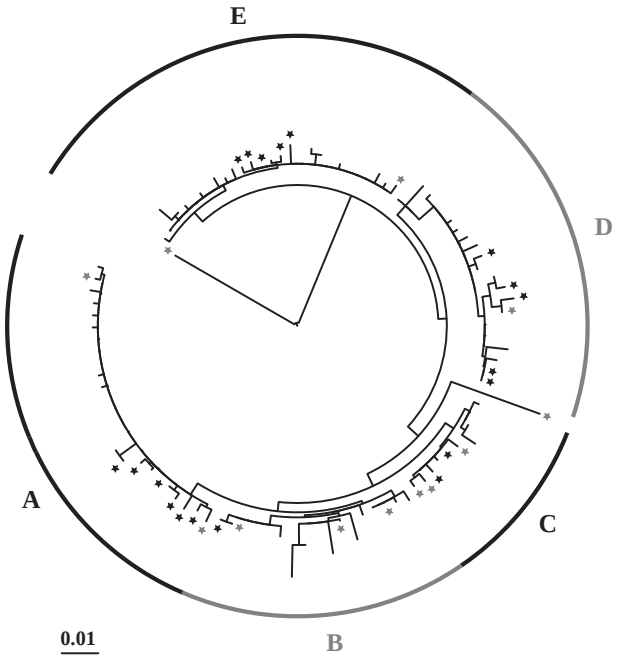


Fig. 3. ML phylogenetic tree of MAST-4 18S rDNA with a subset of GenBank sequences shown in figure 2 (indicated with stars) plus additional unique pyrosequences obtained in our study.

18S rDNA tree (fig. 4*A*) was consistent with that shown in figure 2. The 28S rDNA tree displayed the same five clades as before, but here, the clades appeared better resolved and separated with longer phylogenetic distances (fig. 4*D*). The 5.8S was the least informative of the three genes, since all sequences within clades A–C were identical (fig. 4*C*). LSU rDNA sequences were also used to place MAST-4 within the stramenopiles (tree not shown) and revealed a position consistent with previous 18S rDNA trees (Massana et al. 2004). Giving the variability of ITS regions, we did not attempt to construct a tree with all sequences. Instead, ITS1 and ITS2 trees were done separately for each clade and used to contrast their topology (the order and relative branching of the different clones) (fig. 4*B*). Clade C exhibited a consistent topology, with clones branching in the same way in both ITS trees, whereas clones within clade A were widely mixed when comparing both trees. Clade E would be an intermediate case of the two previous examples.

The averaged 18S rDNA sequence distance between clades for the partial sequences in the SSU-LSU data set was 0.027, very similar to the distance estimated with complete sequences (table 1). The additional resolution provided by 28S rDNA partial sequences was clear, as the interclade distance using this gene was 0.161. Interclade distances using ITS sequences were much higher: 0.420 for ITS1, 0.418 for ITS2 (0.270 for the whole ITS region), although these values were less certain due to the inherent difficulty in aligning ITS regions. It was also clear that not all clades had a comparable intraclade variability, which was low in clades A and E and significantly larger in clades B and C. In fact, the maximal intraclade ITS distance in clade B or C was similar to the minimal interclade distance.

Limited Diversity of a Widespread Picoeukaryote · doi:10.1093/molbev/msr303 MBE

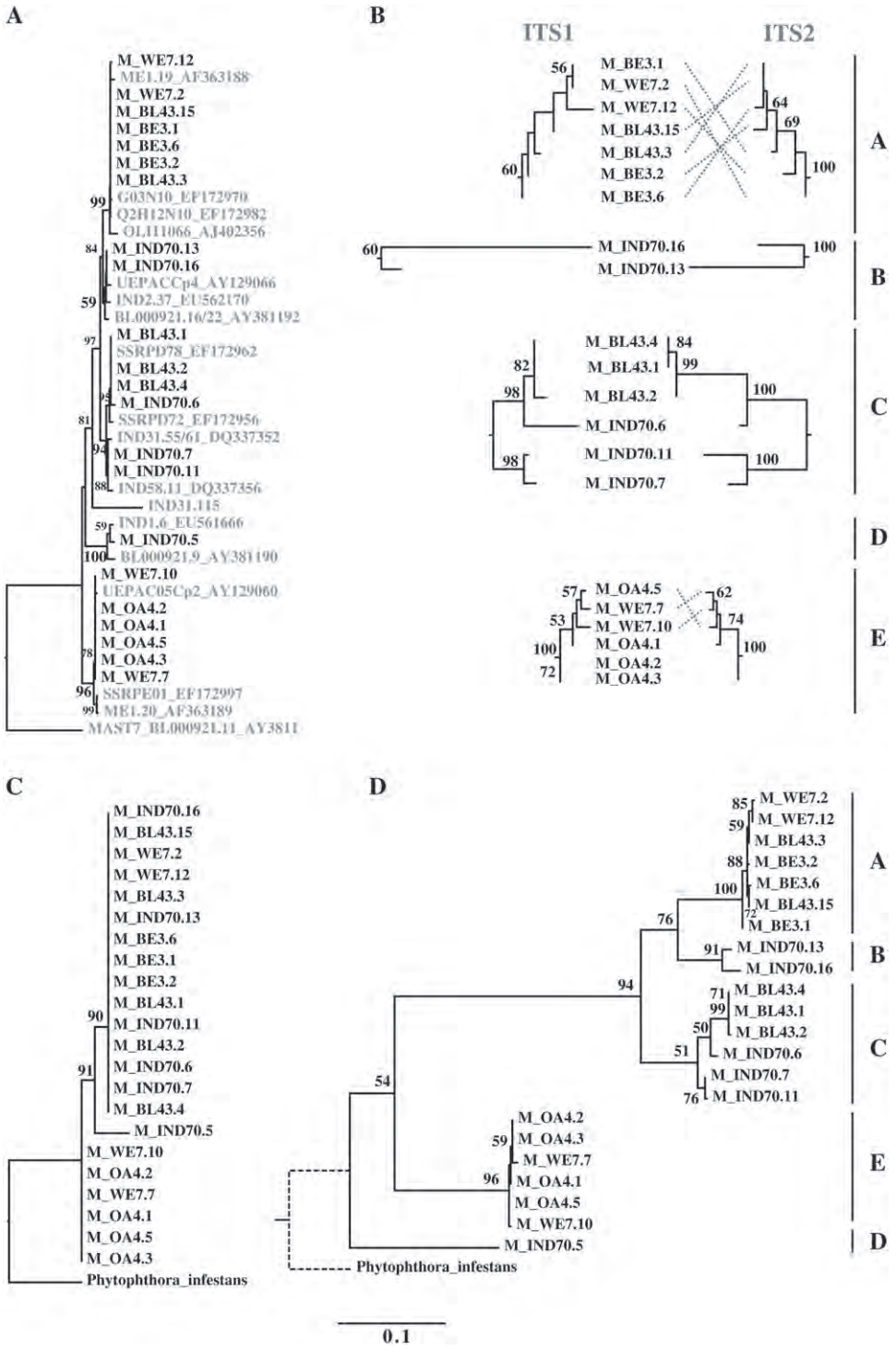


Fig. 4. ML phylogenetic trees constructed with 22 MAST-4 clones considering the partial 18S rDNA sequences (*A*), contrasted ITS1 and ITS2 sequences (*B*), complete 5.8S rDNA sequences (*C*), and partial 28S rDNA sequences (*D*). Tree in *A* also includes complete 18S rDNA sequences (in gray). The scale bar applies to all trees and indicates substitutions per position. Bootstrap values above 50 are shown.

Exploring Intraclade Diversity Using ITS Secondary Structures

ITS secondary structures allow differentiating between constrained or neutral changes in ITS sequences. ITS2 secondary structures of MAST-4 contained the four-helix domains known in many eukaryotic taxa (fig. 5). Helix II included the universal pyrimidine–pyrimidine (U–U) mismatch and had an initial stem of five base pairs that was conserved within each clade, with the exception of one position in clade C (fig. 5, helix II). Differences between clades in this section were supported by compensatory base changes (CBCs). In the 5′-side of helix III, the most conserved part in the ITS2, clades A and E exhibited 30 conserved positions, whereas 24 and 22 positions were conserved in clades B and C, respectively (fig. 5, helix III). The TGGT motif in the middle of the helix III conserved region (Coleman 2007) was observed

Rodríguez-Martínez et al. · doi:10.1093/molbev/msr303 MBE

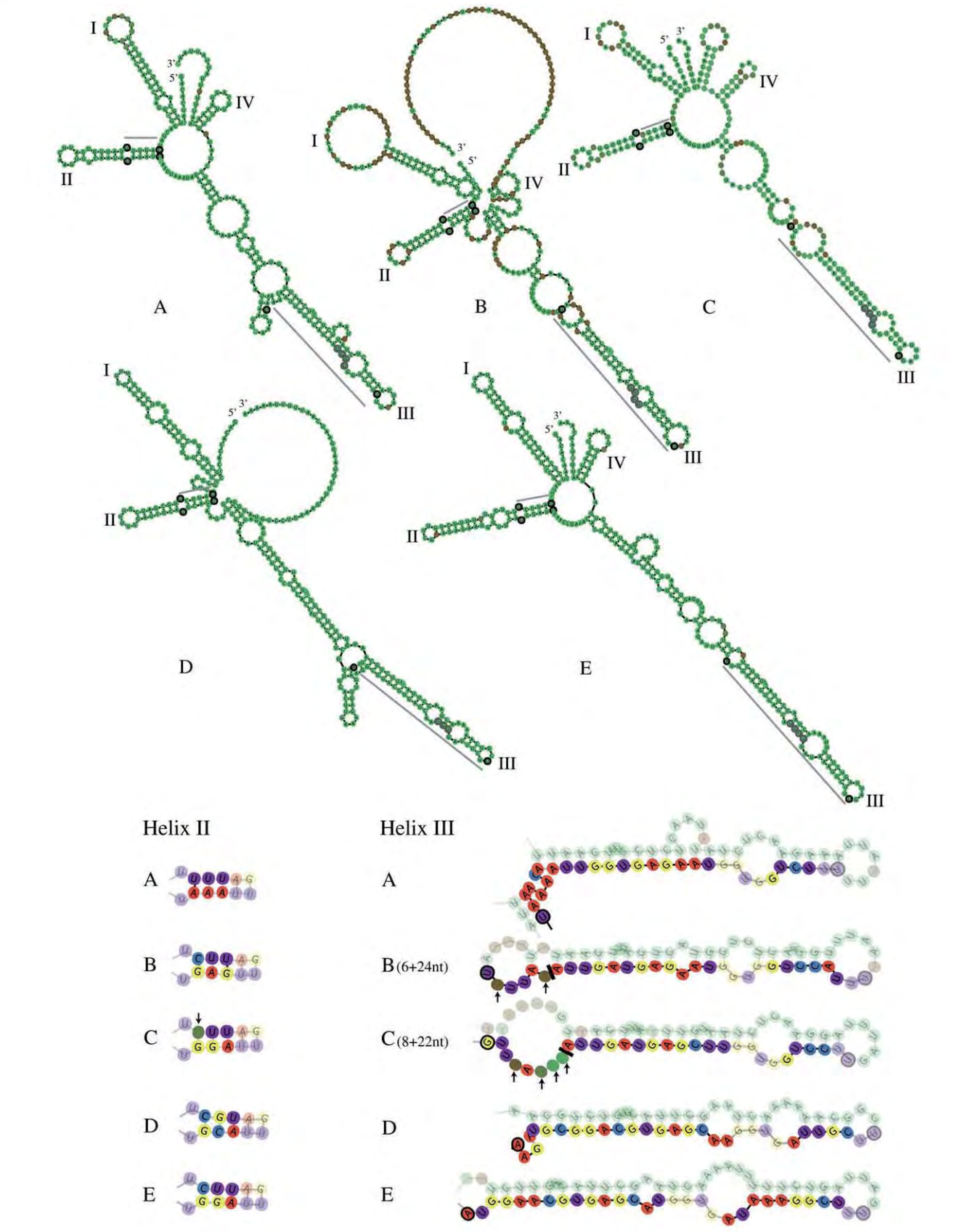


Fig. 5. Consensus ITS2 secondary structures for each of the five MAST-4 clades (*A*-*E*), showing three main helices (I–III). Base pairs highly conserved within each clade are shown in green; variable positions are shown in brown. Nucleotides with gray circles represent the UGGU motif. Details of helices II and III are represented at the bottom of the figure, highlighting in bright color the conserved nucleotides within each clade that differ among clades. Nucleotides conserved in all sequences appear in weak color. Black lines in helix III delimit the longest position (if this is smaller than 30 nt) of this conserved region until polymorphism appears. Arrows point to polymorphisms (in conserved regions) within a clade.

Limited Diversity of a Widespread Picoeukaryote · doi:10.1093/molbev/msr303 MBE

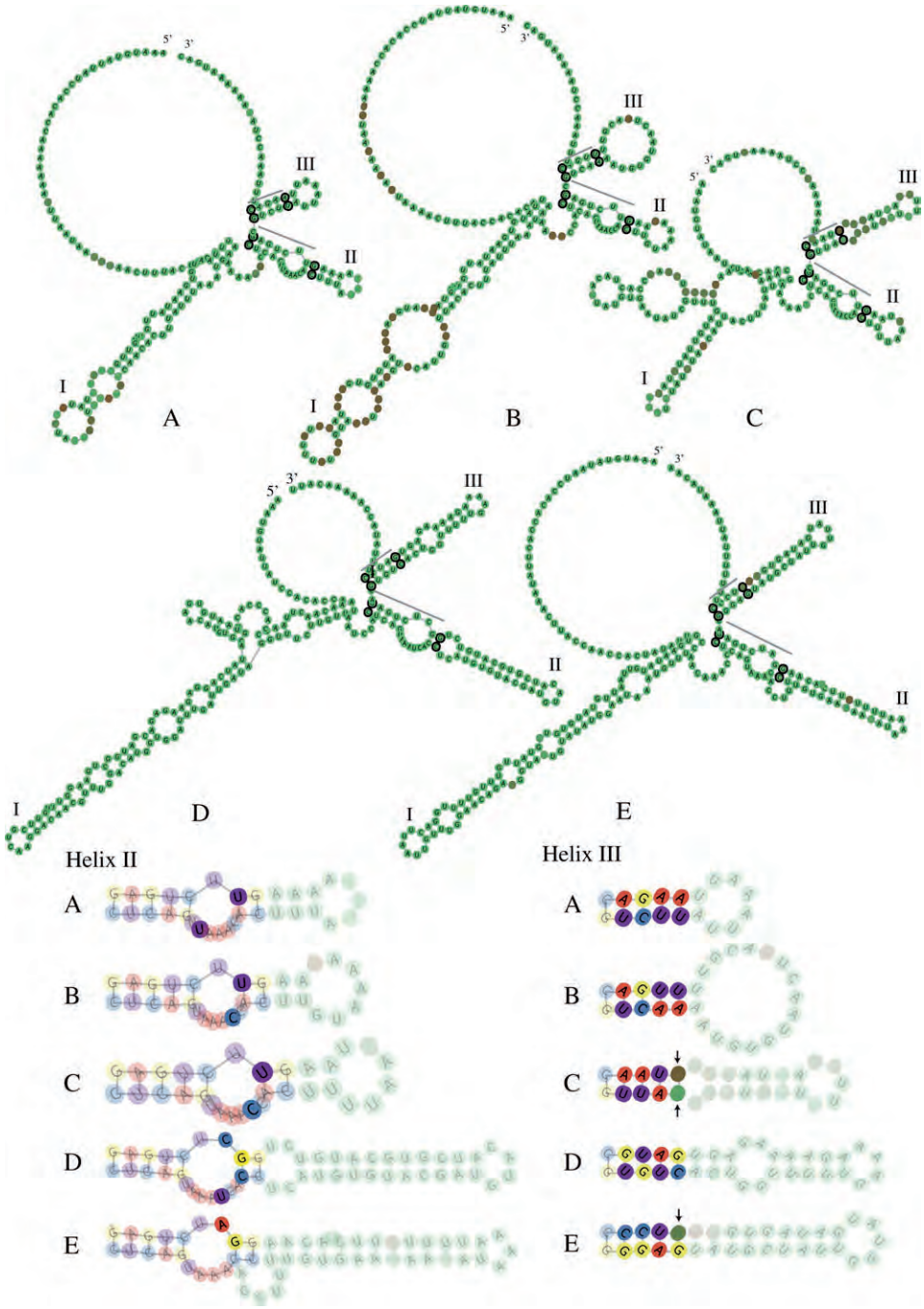


Fig. 6. Consensus ITS1 secondary structure for each of the five MAST-4 clades (*A*-*E*), showing three main helices (I–III). Base pairs highly conserved within each clade are shown in green; variable positions are shown in brown. Details of helices II and III are represented at the bottom of the figure, highlighting in bright color the conserved nucleotides within each clade that differ among clades. Nucleotides conserved in all sequences appear in weak color. Arrows point to polymorphisms (in conserved regions) within a clade.

in all cases except for clade D that had AGGT (fig. 5). As in other taxa studied, helices I and IV were very variable (in some cases, helix IV did not appear in the intraclade consensus structure). Overall, we identified helices II and III as regions in the ITS2 that were conserved within clades and differed among clades (with only a few position exceptions).

For ITS1 sequences, we found a common core secondary structure with three helices. Helix I was the most variable, helix II was the most conserved, even for the primary sequence, and the short helix III had a conserved secondary structure (fig. 6). In helix II, there was an initial stem with 5 bp common for all sequences, a loop with positions differing among clades, a stem conserved in clades A–C but very different in clades D and E and a final nonconserved loop. In Helix III, there was a basal stem of four conserved base pairs that started with CG in all clades and was followed by conserved base pairs within clades but differing

with CBCs between them. If we expanded it to the fifth base pair, polymorphisms appeared within clades C (three groups supported by CBCs and hemiCBCs) and E (two groups with a hemiCBC). As found in ITS2 secondary structures, we identified an ITS1 conserved region (in helix III) that was conserved within a clade and varied among clades.

Discussion

The group MAST-4 accounts for 9% of heterotrophic flagellates in all the oceans (except in polar waters), with an average abundance of 131 cells ml^{-1} (Massana, Terrado, et al. 2006). Its global population size is estimated to be about 10^{24} cells, 2,500 billion times the number of total birds in the world (Gaston and Blackburn 1997), which are divided in around 8,600 species (May and Beverton 1990). In contrast, the MAST-4 18S rDNA phylogeny reveals only five main clades, each one exhibiting low sequence divergence and conservation in specific parts of the ITS secondary structures. This lineage appears as a well-supported discrete group in 18S rDNA phylogenies, and the closest known outgroup sequences are only 91% similar. In addition, the maximal 18S rDNA sequence divergence within MAST-4 is 0.044 (table 1), a very low value as compared with other protist groups (Pernice M, personal communication). Pyrosequencing added more than one order of magnitude of sequences (with respect to currently available GenBank sequences) and confirmed the low MAST-4 diversity, as all found sequences affiliated to the five known clades. It is important to note that the MAST-4 pyrosequences were retrieved from a vast environmental protist survey. Only clone IND31.115 did not match any new pyrosequences, and we confirmed that this divergent sequence was not a chimera or a sequencing error. Instead, this sequence could be a pseudogene (Thornhill et al. 2007). Overall, it is remarkable that such a widespread and abundant protist group appears to have experienced so little evolutionary diversification. A similar scenario of low diversity and cosmopolitan distribution seems to exist in other picoeukaryotes, such as the prasinophyte *Micromonas* (Slapeta et al. 2006).

For a better interpretation of the detected genetic variability in this uncultured group, we sequenced for the first time its complete ITS region, since it has been observed that the ITS2 secondary structure can help in delimiting species (Coleman 2007, 2009). In particular, strains exhibiting at least one CBC in the conserved nucleotides of helices II and III were shown to belong to different biological species. Still the presence of a hemiCBC (one sided) could allow some weak degree of interbreeding (Coleman 2009). For instance, two *Pseudo-nitzschia* strains differing by three hemiCBCs produced zygotes but never gave viable offspring, thus being considered as separate species (Amato et al. 2007). We looked for these conserved regions among the consensus ITS2 secondary structures of each MAST-4 clade. The 5 bp in helix II are identical within each clade but differ by CBCs among clades, suggesting that each clade is a separate biological species. The only exception is clade C, which has a hemiCBC between several clones, so it could include at least two species. With respect to the conserved region in helix III, different size criteria have been invoked in the literature to correlate with sexual incompatibility, 18 positions (Coleman 2007), ~20 (Coleman 2003) or 30 (Coleman 2009). In clades A and E, we identified a region of 30 bp conserved within clades and differing among clades, whereas clades B and C exhibited a slightly shorter conserved fragment. So, using the most restrictive criteria of 30 bp, these two latter clades would include more than one species.

Whereas the ITS2 secondary structure has been widely investigated, the ITS1 still lacks of a universal core secondary structure model. In the eukaryotic taxa examined so far, it is typically represented by an open loop containing multiple double-stranded helices (Coleman et al. 1998; Gottschling et al. 2001; Goertzen 2003; Gottschling 2004; Hoshina 2010; Thornhill and Lord 2010). However, it seems that the generally accepted hypervariability of ITS1 was overestimated (Itskovich et al. 2008), and this region can be used, for example, to define species complex groups on the basis of the conserved helix II motif (Bridge et al. 2008). Similarly to the ITS2, the consensus ITS1 secondary structures for each MAST-4 clade suggest five different species (one per clade), with the presence of CBCs among clades in the conserved 4 bp stem in helix III. If this stem was elongated with one additional base pair, clades C and E would include more putative species.

In order to further investigate whether each MAST-4 clade is composed of one or several species, we contrasted their corresponding tree topologies recovered by the ITS1 and ITS2 regions. These are rapidly evolving spacers that can be used to explore questions related to the speciation process (Coleman 2007; Mullineux and Hausner 2009). We hypothesize that groups that have diversified enough to constitute different species will display congruent topologies in their ITS1 and ITS2 trees, since no recombination would exist among these markers. In contrast, groups that may still constitute one single species, or are in the process of speciation, might display incongruent topologies due to present or recent recombination events. Basically, this is the concordance–discordance principle used for the recognition of phylogenetic species (Taylor et al. 2000). The incongruent topologies displayed by clade A (fig. 4*B*) suggest that it could represent a single biological species. This is consistent with the low intraclade divergence in the ITS region (table 1), which is within the observed variation in other species (table 2). On the contrary, clade C displays the same topology in the two ITS trees, revealing three subclades that could qualify as separate species. In addition, clade C has an intraclade divergence similar to the minimum interclade distance (table 1) and also to the average minimum divergence between species (table 2). A similar scenario of high divergence is seen with the two sequences from clade B. Finally, clade E has some sequences intermixed in the trees and others with a consistent position, suggesting that it may include more than one species.

There are a few alternative explanations, besides speciation, for the ITS variability we observe. It is unlikely

Table 2. Different Examples of ITS Sequence Divergences (uncorrected p-distance; shown as minimum–maximum) among Related Strains or Species of Cultured Eukaryotes.

		Intraclonal			Intraspecies			Interspecies			
	Species	ITS	ITS1	ITS2	ITS	ITS1	ITS2	ITS	ITS1	ITS2	Ref.[a]
Diatoms	*Eunotia bilunaris*	0.000–0.052			0.000–0.123						1
	E. bilunaris	0.000–0.043			0.000–0.044						2
	Pseudo-nitzschia multistriata				0.006	0.010	0.006				3
	P. pungens	0.000–0.070			0.000–0.044	0.000–0.050	0.000–0.064				4
	P. seriata and *P. australis*								0.036	0.027	5
	P. decipiens and *P. dolorosa*				0.000–0.005			0.105–0.108			6
	P. delicatissima and *P. decipiens*				0.000–0.049			0.075–0.090			
	P. dolorosa and *P. delicatissima*				0.000–0.002			0.129–0.151			
	Several species (5.8S+ITS2)				0.000–0.070			0.110–0.260			7
Dinoflagellates	*Symbiodinium*	0.006–0.061	0.009–0.043	0.010–0.124							8
	Several species	0.000–0.017	0.000–0.034	0.000–0.026	0.000–0.021	0.000–0.040	0.000–0.021	0.042–0.577	0.038–0.734	0.020–0.732	9
	Several species							0.000–0.014			10
	Peridinium limbatum and *P. willei*					0.000–0.099	0.000–0.111		0.551–0.566	0.432–0.463	11
	Scrippsiella trochoidea				0.002–0.015						12
Ciliates	*Halteria grandinella*				0.001–0.082						13
Mollusca	*Haliotis*					0.000–0.049	0.000–0.044		0.380–0.590	0.380–0.480	14
Copepod	Several species						0.000–0.008			0.002–0.034	15
Magnoliophyta	Several species								0.000–0.480	0.000–0.440	16
Averages		0.001–0.049	0.005–0.039	0.005–0.075	0.000–0.042	0.002–0.050	0.001–0.042	0.077–0.200	0.201–0.481	0.144–0.363	

[a] References: 1) Vanormelingen et al. (2008), 2) Vanormelingen et al. (2007), 3) D'Alelio et al. (2008), 4) Casteleyn et al. (2008), 5) Fehling et al. (2004), 6) Lundholm et al. (2006), 7) Moniz and Kaczmarska (2009), 8) Thornhill et al. (2007), 9) Litaker et al. (2007), 10) Logares et al. (2008), 11) Kim et al. (2004), 12) Montresor et al. (2003), 13) Katz et al. (2005), 14) Coleman and Vacquier (2002), 15) Goetze (2003), and 16) Goertzen (2003).

that this variability is due to experimental artifacts, as we used very stringent sequencing and analysis methods, but it could be caused by intragenomic polymorphisms (Prokopowich et al. 2003). It is generally assumed that these polymorphisms are rapidly eliminated through a series of homogenizing mechanisms referred to as concerted evolution (Elder and Turner 1995; Ganley and Kobayashi 2007), but it is also known that some species have a fraction of the rDNA units that have escaped the process of concerted evolution (Keller et al. 2006; Simon and Weiss 2008). Intragenomic variation in species where this occurs is very low (lower than interspecific variation, Litaker et al. 2007) and typically only in extremely variable positions that are never paired in secondary structure (Behnke et al. 2004; Orsini et al. 2004; Casteleyn et al. 2008). Thus, the ITS can be treated as single copy region (Coleman 2003). Finally, we observed that MAST-4 cells have a relatively low rDNA copy number (around 30, Rodríguez-Martínez et al. 2009), which reduces the possibility of mutations.

In summary, despite the presence of a huge number of MAST-4 cells in the oceans, its diversity is structured into just five main clades, each representing at least one biological species. Clade A is particularly interesting because it seems to be composed of only one species, appearing in distant oceanic areas. Specifically, clade A showed no polymorphisms in the critical regions of the ITS2 and ITS1 secondary structures (figs. 5 and 6), the topologies of their ITS1 and ITS2 trees were incongruent (fig. 4), and the clade presented a very low sequence divergence (table 1). On the other hand, using the same three criteria, clade B would include at least two species, clade C three species, and clade E two species (clade D is still undersampled with only one ITS sequence). Altogether there is currently evidence of a maximum of only ten separate species within MAST-4. Within each of the clades, diversification appears to be very low, as indicated by the 18S and ITS rDNA markers. This low evolutionary diversification points to either a very recent evolutionary divergence and worldwide dispersal or to a very strong environmental filtering that penalizes any deviation from an optimal cell design.

Acknowledgments

Funding has been provided by FLAME (CGL2010-16304, MICINN, Spain) and BioMarKs (2008-6530, ERA-net Biodiversa, EU) projects to R.M., by an FPI fellowship from the Spanish Ministry of Education and Science to R.R.M., by a Marie Curie Intra-European Fellowship grant PIEF-GA-2009-235365 to R.L., and by NSF ATOL 0629521 to G.R. We thank M. Pernice for Mothur help, V. Balagué and C. Williams for molecular help, R.E. Collins for useful advice, and all people involved in the BioMarKs consortium, in particular, the coordinator Colomban de Vargas.

References

Álvarez I, Wendel J. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol.* 29:417–434.

Amato A, Kooistra W, Levialdi Ghiron J, Mann D, Pröschold T, Montresor M. 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158:193–207.

Behnke A, Friedl T, Chepurnov V, Mann D. 2004. Reproductive compatibility and rDNA sequence analyses in the Sellaphora pupula species complex (Bacillariophyta). *J Phycol.* 40:193–208.

Bridge PD, Schlitt T, Cannon PF, Buddie AG, Baker M, Borman AM. 2008. Domain II hairpin structure in ITS1 sequences as an aid in differentiating recently evolved animal and plant pathogenic fungi. *Mycopathologia* 166:1–16.

Casteleyn G, Chepurnov V, Leliaert F, Mann D, Bates S, Lundholm N, Rhodes L, Sabbe K, Vyverman W. 2008. Pseudo-nitzschia pungens (Bacillariophyceae): a cosmopolitan diatom species? *Harmful Algae* 7:241–257.

Coleman A. 2000. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* 151:1.

Coleman A. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* 19:370–375.

Coleman A. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.* 35:3322–3329.

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.

Coleman AW, Maria Preparata R, Mehrotra B, Mai JC. 1998. Derivation of the secondary structure of the ITS-1 transcript in Volvocales and its taxonomic correlations. *Protist* 149:135–146.

Coleman AW, Vacquier VD. 2002. Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (Haliotis). *J Mol Evol.* 54:246–257.

Côté CA, Greer CL, Peculis BA. 2002. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8:786–797.

D'Alelio D, Amato A, Kooistra W, Procaccini G, Casotti R, Montresor M. 2008. Internal transcribed spacer polymorphism in Pseudo-nitzschia multistriata (Bacillariophyceae) in the Gulf of Naples: recent divergence or intraspecific hybridization? *Protist* 160:9–20.

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

Drummond AJ, Ashton B, Buxton S, et al. (12 co-authors). 2010. Geneious v5.1 [Internet]. Available from: http://www.geneious.com.

Edvardsen B, Eikrem W, Green J, Andersen RA, Moon-van der Staay SY, Medlin LK. 2000. Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia* 39:19–35.

Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol.* 70:297–320.

Evans KM, Wortley AH, Mann DG. 2007. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist* 158:349–364.

Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive earth's biogeochemical cycles. *Science* 320:1034–1039.

Fehling J, Green DH, Davidson K, Bolch CJ, Bates SS. 2004. Domoic acid production by Pseudo-nitzschia seriata (Bacillariophyceae) in Scottish waters. *J Phycol.* 40:622–630.

Ganley ARD, Kobayashi T. 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* 17:184–191.

Gaston KJ, Blackburn TM. 1997. How many birds are there? *Biodivers Conserv.* 6:615–625.

Goertzen L. 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Mol Phylogenet Evol.* 29:216–234.

Goetze E. 2003. Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proc Biol Sci.* 270:2321–2331.

Gottschling M. 2004. Secondary structure models of the nuclear internal transcribed spacer regions and 5.8S rRNA in Calciodinelloideae (Peridiniaceae) and other dinoflagellates. *Nucleic Acids Res.* 32:307–315.

Gottschling M, Hilger H, Wolf M, Diane N. 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boraginales. *Plant Biol.* 3: 629–636.

Gutell R, Larsen N, Woese C. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Mol Biol Rev.* 58:10.

Haas BJ, Gevers D, Earl AM, et al. (15 co-authors). 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21:494–504.

Hoshina R. 2010. Secondary structural analyses of ITS1 in Paramecium. *Microbes Environ.* 25:313–316.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Itskovich V, Gontcharov A, Masuda Y, Nohno T, Belikov S, Efremova S, Meixner M, Janussen D. 2008. Ribosomal ITS sequences allow resolution of freshwater sponge phylogeny with alignments guided by secondary structure prediction. *J Mol Evol.* 67:608–620.

Jerome C, Lynn D. 1996. Identifying and distinguishing sibling species in the Tetrahymena pyriformis complex (Ciliophora, Oligohymenophorea) using PCR/RFLP analysis of nuclear ribosomal DNA. *J Eukaryot Microbiol.* 43:492–497.

Joseph N, Krauskopf E, Vera M, Michot B. 1999. Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res.* 27:4533.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 9:286–298.

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.

Keller I, Chintauan-Marquier IC, Veltsos P, Nichols RA. 2006. Ribosomal DNA in the grasshopper Podisma pedestris: escape from concerted evolution. *Genetics* 174:863–874.

Kim E, Wilcox L, Graham L, Graham J. 2004. Genetically distinct populations of the dinoflagellate Peridinium limbatum in neighboring Northern Wisconsin lakes. *Microb Ecol.* 48:521–527.

Lange M, Chen YQ, Medlin LK. 2002. Molecular genetic delineation of Phaeocystis species (Prymnesiophyceae) using coding and non-coding regions of nuclear and plastid genomes. *Eur J Phycol.* 37:77–92.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.

Litaker RW, Vandersea MW, Kibler SR, Reece KS, Stokes NA, Lutzoni FM, Yonish BA, West MA, Black MND, Tester PA. 2007. Recognizing dinoflagellate species using ITS rDNA sequences. *J Phycol.* 43:344–355.

Logares R, Daugbjerg N, Boltovskoy A, Kremp A, Laybourn-Parry J, Rengefors K. 2008. Recent evolutionary diversification of a protist lineage. *Environ Microbiol.* 10:1231–1243.

Logares R, Rengefors K, Kremp A, Shalchian-Tabrizi K, Boltovskoy A, Tengs T, Shurtleff A, Klaveness D. 2007. Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? *Microb Ecol.* 53:549–561.

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

Lowe CD, Montagnes DJS, Martin LE, Watts PC. 2010. Patterns of genetic diversity in the marine heterotrophic flagellate Oxyrrhis marina (Alveolata: Dinophyceae). *Protist* 161:212–221.

Lundholm N, Moestrup Ø, Kotaki Y, Hoef-Emden K, Scholin C, Miller P. 2006. Inter-and intraspecific variation of the Pseudo-nitzschia delicatissima complex (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic analyses. *J Phycol.* 42:464–481.

Mai JC, Coleman AW. 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J Mol Evol.* 44:258–271.

Massana R. 2011. Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol.* 65:1–47.

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, DeLong EF, Pedrós-Alió C. 2000. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol.* 66:1777–1787.

Massana R, Guillou L, Díez B, Pedrós-Alió C. 2002. Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. *Appl Environ Microbiol.* 68:4554–4558.

Massana R, Guillou L, Terrado R, Forn I, Pedros-Alio C. 2006. Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol.* 45:171–180.

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

May RM, Beverton RJH. 1990. How many species? *Philos Trans R Soc Lond B Biol Sci.* 330:293.

Moniz MBJ, Kaczmarska I. 2009. Barcoding diatoms: is there a good marker? *Mol Ecol Resour.* 9:65–74.

Moniz MBJ, Kaczmarska I. 2010. Barcoding of diatoms: nuclear encoded ITS revisited. *Protist* 161:7–34.

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

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.

Müller T, Philippi N, Dandekar T, Schultz J, Wolf M. 2007. Distinguishing species. *RNA* 13:1469.

Mullineux T, Hausner G. 2009. Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera Grosmannia and Leptographium. *Fungal Genet Biol.* 46:855–867.

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

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.

Prokopowich CD, Gregory TR, Crease TJ. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46:48–50.

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

Rambaut A. 2009. FigTree. Edinburgh (United Kingdom): Institute of Evolutionary Biology, University of Edinburgh. Available from: http://tree.bio.ed.ac.uk/software/figtree.

Rocap G, Distel D, Waterbury J, Chisholm S. 2002. Resolution of Prochlorococcus and Synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol.* 68:1180.

Rodríguez-Martínez R, Labrenz M, Del Campo J, Forn I, Jürgens K, Massana R. 2009. Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environ Microbiol.* 11: 397–408.

Rynearson TA, Lin EO, Armbrust EV. 2009. Metapopulation structure in the planktonic diatom Ditylum brightwellii (Bacillariophyceae). *Protist* 160:111–121.

Schloss PD, Westcott SL, Ryabin T, et al. (15 co-authors). 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 75:7537–7541.

Schlötterer C, Hauser MT, von Haeseler A, Tautz D. 1994. Comparative evolutionary analysis of rDNA ITS regions in Drosophila. *Mol Biol Evol.* 11:513.

Schultz J, Muller T, Achtziger M, Seibel P, Dandekar T, Wolf M. 2006. The internal transcribed spacer 2 database—a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res.* 34:W704.

Seibel P, Müller T, Dandekar T, Schultz J, Wolf M. 2006. 4 SALE—a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* 7:498.

Selig C, Wolf M, Muller T, Dandekar T, Schultz J. 2007. The ITS2 Database II: homology modelling RNA structure for molecular systematics. *Nucleic Acids Res.* 36:D377–D380.

Simon UK, Weiss M. 2008. Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Mol Biol Evol.* 25:2251–2254.

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.

Sorhannus U, Ortiz JD, Wolf M, Fox MG. 2010. Microevolution and speciation in Thalassiosira weissflogii (Bacillariophyta). *Protist* 161:237–249.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

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

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol.* 31:21–32.

Thompson J, Marcelino L, Polz M. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* 30:2083.

Thornhill DJ, Lajeunesse TC, Santos SR. 2007. Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. *Mol Ecol.* 16:5326–5340.

Thornhill DJ, Lord JB. 2010. Secondary structure models for the internal transcribed spacer (ITS) region 1 from symbiotic dinoflagellates. *Protist* 161:434–451.

Tippery N, Les D. 2008. Phylogenetic analysis of the internal transcribed spacer (ITS) region in Menyanthaceae using predicted secondary structure. *Mol Phylogenet Evol.* 49:526–537.

Vanormelingen P, Chepurnov V, Mann D, Cousin S, Vyverman W. 2007. Congruence of morphological, reproductive and ITS rDNA sequence data in some Australasian Eunotia bilunaris (Bacillariophyta). *Eur J Phycol.* 42:61–79.

Vanormelingen P, Chepurnov V, Mann D, Sabbe K, Vyverman W. 2008. Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of Eunotia bilunaris sensu lato (Bacillariophyta). *Protist* 159:73–90.

Wang S, Bao Z, Li N, Zhang L, Hu J. 2007. Analysis of the secondary structure of ITS1 in Pectinidae: implications for phylogenetic reconstruction and structural evolution. *Mar Biotechnol.* 9: 231–242.

Wiemers M, Keller A, Wolf M. 2009. ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus Agrodiaetus (Lepidoptera: Lycaenidae: Polyommatus). *BMC Evol Biol.* 9:300.

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

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

Chapter 5

Biogeography of the uncultured marine picoeukaryote MAST-4: temperature driven distribution patterns

Rodríguez-Martínez R, Rocap G, Salazar G, Massana R. Biogeography of the uncultured marine picoeukaryote MAST4: temperature driven distribution patterns.
Submitted to *The ISME Journal.*

Biogeography of the uncultured marine picoeukaryote MAST-4: temperature driven distribution patterns

Raquel Rodríguez-Martínez[1]*, Gabrielle Rocap[2], Guillem Salazar[1] and Ramon Massana[1]*

[1] Institut de Ciències del Mar, CMIMA (CSIC). Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain

[2] School of Oceanography, University of Washington, Box 357940, Seattle, Washington, USA

Abstract

The MAST-4 is a widespread uncultured picoeukaryote that accounts for an important fraction of marine heterotrophic flagellates. This group has low genetic divergence and is composed of a small number of species. We combined ARISA (automated ribosomal intergenic spacer analysis) and ITS clone libraries to study the biogeography of this marine protist, examining both spatial and temporal trends in MAST-4 assemblages and associated environmental factors. We did not see marine geographical barriers for the dispersal of the most represented MAST-4 clades, which appeared adapted to different temperature ranges. Distant samples sharing the same temperature presented very similar assemblages, especially in cold temperatures, where only one clade dominated. The most highly represented clades, A and E1, had high gene flow between very distant geographic regions and may each consist of a single species. Our results contribute to the general discussion on microbial biogeography by showing no dispersal limitation together with strong environmental selection for some picoeukaryotes in the marine environment.

Keywords: Biogeography, ITS1, ARISA, temperature, gene flow, MAST-4

Introduction

Biogeography is the study of the distribution of biodiversity over space and time. Most research and theories have been established through the study of plants and animals, revealing that their metapopulations are geographically restricted (MacArthur and Wilson 1967), due to the result of both deterministic (environmental) and stochastic (dispersal) processes. The debate about whether or not microorganisms follow similar patterns dates back decades ago (Baas-Becking 1934, Beijerinck 1913), and has seen a renovated interest in recent years (O'Malley 2008). A long-held concept in microbial ecology is that free-living microbes smaller than 1 mm body size (all prokaryotes and most protists) are probably sufficiently abundant to have worldwide distribution owing to their dispersal ability (Fenchel and Finlay 2004, Finlay 2002, Finlay and Fenchel 2004). Microbial cells inevitably will hitch a ride via wind, water, birds, or floating vegetation, and many species have an astonishing ability to hunker down in harsh environments until their moment arises (Whitfield 2005). Consequently, microbial organisms are believed to occur wherever the environment permits: 'everything is everywhere, but the environment selects' (Baas-Becking 1934, De Wit and Bouvier 2006). Implicit in this tenet is that free-living microbial taxa are not randomly distributed, but rather exhibit biogeographic patterns, and in some cases these patterns may be qualitatively similar to those observed for macroorganisms (Green et al 2004, Horner-Devine et al 2004, Horner-Devine et al 2007). However, these patterns would be the result of local environmental selection rather than dispersal limitation.

The current evidence confirms that environmental selection is fundamental for the spatial variation observed in microbial diversity (Martiny et al 2006). The next frontier is to figure out whether these patterns are also influenced by geographical barriers that facilitate evolution and diversification. However, because geographic distance is often correlated with specific environmental characteristics, disentangling the relative influence of these two factors on community divergence represents a major challenge in elucidating whether or not microbes are limited by dispersal. It is known that freshwater diatoms present real dispersal limitations because of desiccation intolerances (Vyverman et al 2007). Other highly specialized microorganisms, such as the hyperthermophiles, are unlikely to make a long dispersal journey, so they would be easily isolated by geographic barriers, resulting in the development of a global diversity structure (Whitaker 2003). However, more dispersal limitation is expected in terrestrial and freshwater systems than in marine systems.

The oceans are an interconnected geophysical fluid that potentially allows planktonic organisms to disperse globally. A global conveyor belt is mixing oceanic waters at scales of thousand years (Broecker 1991). Thus, tectonic and water mass dispersal barriers are often weak and unable to geographically isolate pelagic planktonic populations for extended periods of time (Sexton and Norris 2008). The geographic distribution of marine planktonic diatoms does not seem to be limited by dispersal and it is environmental selection which dominates diatom community structure (Cermeño and Falkowski 2009). The general view is that of a broad dispersal of marine planktonic microbes (Cermeño et al 2010). An optimal target to further investigate biogeographical patterns of marine microbes are the picoeukaryotes. These are small protists 1-3 μm in size populating surface oceans at abundances of 10^2 to 10^4 cells ml^{-1}, playing important ecological roles, and exhibiting a high and underexplored diversity (Massana 2011).

In a previous study (Rodríguez-Martínez et al 2012), we investigated the diversity structure of an important uncultured picoeukaryote, the MAST-4 lineage (Massana et al 2004), which is widespread in surface marine waters (except polar systems) and represents approximately 9% of heterotrophic flagellates (Massana et al 2006, Rodríguez-Martínez et al 2009). Despite its huge number of cells in the oceans, MAST-4 has a very low genetic divergence and is composed of only five main clades, each representing at least one biological species. The small size (~2 μm), high abundance, worldwide distribution and low genetic diversity make MAST-4 a good model to study marine protist biogeography. In this work we determined the MAST-4 community structure and distribution by combining automated ribosomal intergenic spacer analysis (ARISA) (Fisher and Triplett 1999) and 18S-ITS1 gene libraries. MAST-4 diversity was analyzed in 40 different marine locations and we found evidence for a strong environmental selection and no-dispersal limitation for the most represented clades.

Materials and Methods

Study sites and sampling

Environmental samples were selected from different oceanographic cruises performed at the North Atlantic Ocean (NOR, NAT, RG, BE and COC), the North Pacific Ocean (WE), the Mediterranean Sea (BL and AL) and the southern hemisphere (IND and DH) (Figure 1). Details of some of these cruises have been already published: NOR (Not et al 2005), NAT (González et al 2000), COC (Alonso-Sáez et al 2007), WE (del Giorgio et al 2011), BL (Alonso-Sáez et al 2008), AL (Arin et al 2002), IND (Not et al 2008) and DH (Díez et al 2004). Seawater at different depths was collected with Niskin bottles attached to a CTD rosette and filtered for the 0.2-3 μm microbial fraction with a peristaltic pump. DNA extraction was done using enzymatic and SDS digestion plus phenol purification (Massana et al 2000). The quality and quantity of extracted genomic DNA was determined with a NanoDrop 1000 (Thermo Fisher Scientific Inc., Wilmington, DE). Physico-chemical data (temperature, salinity) and chlorophyll concentration from the samples were compiled.

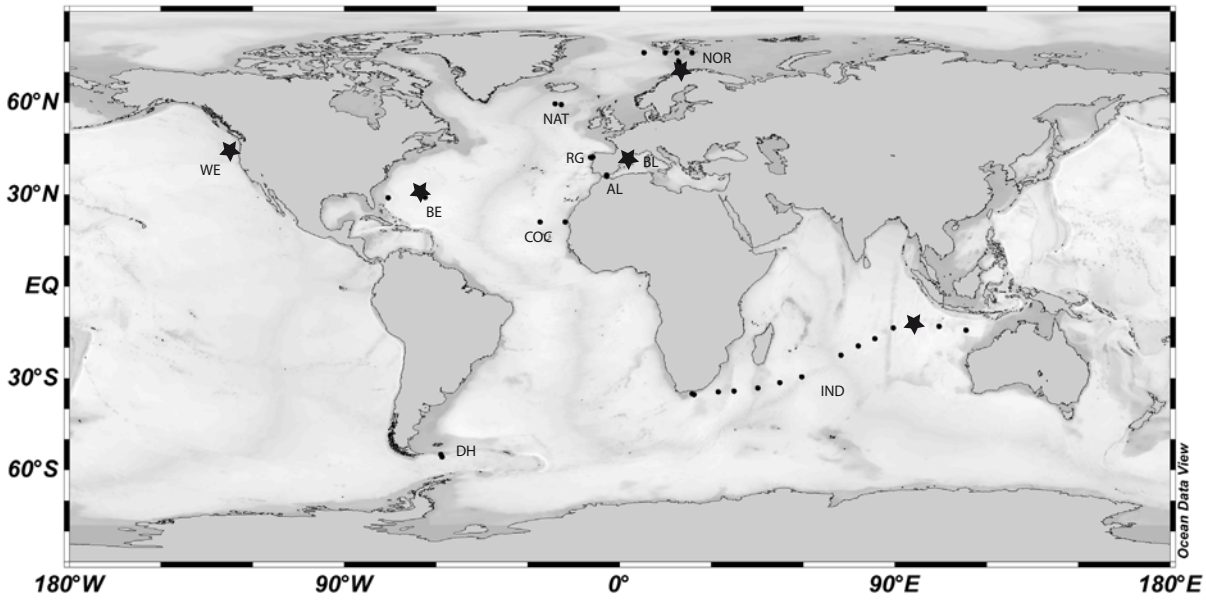


Figure 1. Global map indicating sampling sites used for ARISA fingerprinting (dots) and sites used for clone libraries (stars). The acronym of the cruise is indicated close to the stations.

ARISA fingerprinting was done for surface samples in all the stations marked in Figure 1, for additional depths in most stations, and for a temporal survey at the Blanes Bay Microbial Observatory (BL). Surface samples from stations marked with a star were used for the clone libraries.

Design of PCR primers

Specific primers were designed to amplify the end of the 18S rDNA (290bp) the Internal Transcribed Spacer 1 (ITS1) and the beginning of the 5.8S (39bp). The forward primer M4.18S-F (5'-TGGGTAATCTTTGAACGTGAAT-3'), located before the V9 region of the 18S rDNA, was designed based on all MAST-4 sequences for this region available so far (52 unique clones). It matched perfectly all these clones, except two (ME1.29 and OLI11066) that had an extra nucleotide (likely a sequencing error). It had more than two mismatches (except for 1 sequence with one mismatch and 3 sequences with two mismatches) to non-target sequences from the SILVA database (Pruesse et al 2007). The reverse primer M4.58S-R (5'-GTTGCGAGAACCTAGAC-3'), located in the 5.8S rDNA, was designed to have a perfect match with the 22 MAST-4 sequences (Rodríguez-Martínez et al 2012). This primer had two or more mismatches with all stramenopile sequences extracted from GenBank, except for some Labyrinthulida sequences (3 with no mismatches and 62 with one mismatch). Primers were checked for formation of primer dimers, GC content and theoretical melting temperature in the website www.operon.com, using the Oligo Analysis & Plotting Tool. This primer set gave an amplicon size ranging from 500 to 650 bp.

Construction of clone libraries

The PCR mixture (30 µl) contained 15 ng of DNA template, 0.5 µM of each primer, 200 µM of each dNTP, 1 mM $MgCl_2$, 1.5 units of a *Taq* DNA polymerase (Thermo Scientific ThermoPrime), and the enzyme buffer. PCR cycling, carried out in a BioRad thermocycler, was: initial denaturation at 94ºC for 5 min; 30 cycles with denaturation at 94ºC for 1 min, annealing at 60ºC for 45 sec and extension at 72ºC for 1 min; and a final extension at 72ºC for 10 min. We tested the $MgCl_2$ concentration (from 0.5 to 3 mM) and the annealing temperature (from 55 to 66°C) and chose the most stringent conditions giving the expected band. To check the specificity of the primer set, we confirmed the negative signal with 9 non-target cultures (diatoms, haptophytes, dinoflagellates and cyanobacteria). PCR products were purified with the QIAquick PCR Purification kit (QIAGEN) and cloned using the TOPO-TA cloning kit (Invitrogen) with the vector pCR4 following manufacturer's recommendations and a vector-insert ratio of 1:5. Putative positive bacterial colonies

were picked and transferred to a new LB (Luria-Bertani) plate and finally into LB-glycerol solution for frozen stocks (-80ºC). Presence of correct insert was checked by PCR reamplification with vector primers M13F and M13R using a small aliquot of bacterial culture as template. Amplicons with the right insert size (39 to 49 clones per library) were sequenced at the Macrogen sequencing service in Korea. Chromatograms were examined with 4Peaks (A. Griekspoor and T. Groothuis, mekentosj.com).

Sequence analysis

Complete sequences from the clone libraries together with sequences from the same region in the "SSU-LSU" dataset in (Rodríguez-Martínez et al 2012) were aligned with MAFFT v6.853 (Katoh and Toh 2008) with the E- INS-I algorithm, using a MAST-7 sequence as outgroup. The alignment was inspected visually and modified using secondary structure models folded in mFOLD (Zuker 2003) as in (Rodríguez-Martínez et al 2012) A Maximum Likelihood (ML) phylogenetic tree was reconstructed using RAxML v7.0.4 MPI version (Stamatakis 2006), using the General Time Reversible model of nucleotide substitution and a Gamma distributed rate of variation across sites (GTR+G). We did not estimate the proportion of invariable sites and missing data were not considered (i.e. treated as missing data). The shape parameter (α) of the Gamma distribution was estimated from the dataset using default options. Phylogenies were done at the University of Oslo Bioportal (www.bioportal.uio.no). One thousand alternative ML trees were run, and the tree with the best likelihood was selected and visualized in FigTree v1.3.1 (Rambaut 2009). Bootstrap analyses were run with 1000 pseudo-replicates and a consensus tree was constructed with RAxML. To infer intraspecific phylogenies and visualize alternative potential evolutionary paths we constructed median-joining networks (Bandelt et al 1999) with the Network 4.6.0.0 program (Fluxus Technology). The genetic differentiation between populations was estimated with the fixation index (Fst) computed with DNAsp 5.10.1 and can range from 0 (no genetic differentiation) to a maximum of 1 (complete differentiation).

Generation of ARISA profiles

Environmental DNA samples were PCR-amplified in triplicate for ARISA in a MJ Research cycler. PCR conditions were the same as described before using a volume of 25 µl with 10 ng of DNA template, another enzyme (*Taq* DNA Polymerase "Gene Choice") and the forward primer fluorescently labeled (5-HEX). PCR products stored in the dark at 4ºC were purified with MultiScreen ® $PCR_{\mu 96}$ Plates and quantified using PicoGreen fluorescence (Invitrogen, Carlsbad, CA, USA) in a SpectraMax M2 microplate reader (Molecular Devices Corp., Sunnyvale, CA, USA). Ten ng DNA were ethanol precipitated from triplicates or from pooled PCR products (when the yield of the PCR was low), followed by resuspension with 0.078 µl tween, 9.67 µl water and 0.25 µl fluorescently-labeled internal size standard, CST ROX 60-1500 bp (http://www.bioventures.com/). Samples were run on a MegaBACE 1000 automated capillary sequencer (Molecular Dynamics). The electropherograms were then analyzed using DAx software (v8.0, Van Mierlo Software Consultancy). Only peaks exceeding 4 times the noise signal of the electropherogram curve were taken into account.

Analysis of fingerprinting data

From DAx output tables, peak heights were binned using the "fixed window" binning strategy to take into account the size calling imprecision from ARISA fingerprints (Hewson and Fuhrman 2006). In order to determine the best window size with our data, we applied the "automatic binning algorithm" (Ramette 2009) developed in a R script (The R Foundation for Statistical Computing [http://cran.r-project.org/]); we chose 2 bp. To identify the best window frame (out of the 20 possible starting with a shift value of 0.1), we used the "interactive binning algorithm" (Ramette 2009). This algorithm binned the peaks for each frame, calculated the relative fluorescence intensity of each binned peak by dividing its height by the total peak height of the sample and omitted peaks with values <0.5% (considered as background). We added an option in the script to compare frames considering only triplicate samples (instead of all samples). The frame with the best correlation among triplicates was chosen; starting

with 1.3 in our case. The final output was a table with the relative intensity of each binned peak (each considered as a different operational taxonomic unit, OTU) in the size range of 500 to 650 bp. We then performed a permutational multivariate analysis of variance (PERMANOVA) test using the sample as a grouping factor in order to estimate the variability due to the experimental error. If triplicates were identical, this test would explain 100% of the variability. In our case, it explained 92%, indicating that only 8% of the variability of the samples was due to the inaccuracy of the technique.

A consensus OTU-table (relative intensity of all OTUs in all samples) was obtained by averaging the triplicates in each sample, as long as OTUs appeared at least in two of the three replicates. A distance matrix from the consensus OTU-table was calculated with the Bray-Curtis dissimilarity index (OTUs distance matrix). Patterns were explored using clustering analysis (along with the SIMPROF significance test) and nonmetric multidimensional scaling (NMDS) analysis. A stress value was calculated based on the variance between Bray-Curtis coefficient values and actual distances on the MDS plot, with lower values depicting a better representation of sample distances along a two-dimensional space. Simple and partial Mantel tests (with the Pearson correlation) were done to compare the OTUs distance matrix with a geographic distance matrix (using geographical coordinates in a perfect sphere) and a temperature euclidean distance matrix. The relationship between community composition (the OTU-table) and environmental factors was analyzed by a Constrained Correspondence Analysis (CCA). Automatic forward selection with significance tests of Monte Carlo permutations were used to build the optimal models. Additionally we did a PERMANOVA test to check the importance of the environmental factors in the community composition and performed analysis of similarity (ANOSIM) to test for significant differences between a priori sample groupings. We also assessed the contribution of each OTU to the observed similarity (or dissimilarity) between groups with a similarity percentage analysis (SIMPER). All multivariate analyses were also done with a second OTU-table including triplicates and with transformed data (arcsine of the square root of the relative intensity) to reduce the skew; no differences in the results were seen (data not shown). Statistical tests and graphics were done using R packages (Gmt, Vegan, MASS), CANOCO 4.5 for Windows (ter Braak and Smilauer 2002) and Primer v6.1.2 software.

Results

Validation of the different MAST-4 clades

In a previous study we showed that MAST-4 was composed of five main clades (Rodríguez-Martínez et al 2012). To verify the robustness of these clades, we obtained 228 additional sequences of 500-650 bp encompassing the V9 region of the 18S rDNA and the ITS1 region from five different locations (Figure 1), including the locations from the previous study and one additional. All sequences obtained from the clone libraries belonged to the MAST-4 lineage, demonstrating the specificity of the primers. The ML phylogenetic tree (Figure 2) revealed that the new sequences were distributed in the five clades previously observed, sometimes forming new subclades. Because we used the very variable ITS1 region, bootstrap values for this tree were low, but the purpose behind this locus choice was the definitive assignment of sequences to clades and subclades, rather than a robust interclade topology. Sequences grouped in 12 different helix III types for the ITS1 secondary structure described in (Rodríguez-Martínez et al 2012) (Figure 2). Clade A included sequences from the five different sites, all with the same helix III motif. Clade B also had only one helix III motif but was divided in three phylogenetic subclades. Clade C presented four helix III motifs, including several CBCs (Compensatory Base Changes), that corresponded well with phylogenetic subclades except for subclade C2 that had two motifs differing by a hemiCBC. Each one of these subclades, except C1, included sequences from only one location. Clade D had only four sequences but represented two motifs. Finally, clade E consisted of four motifs, two of them forming well differentiated lineages (E2 and E3), and the other

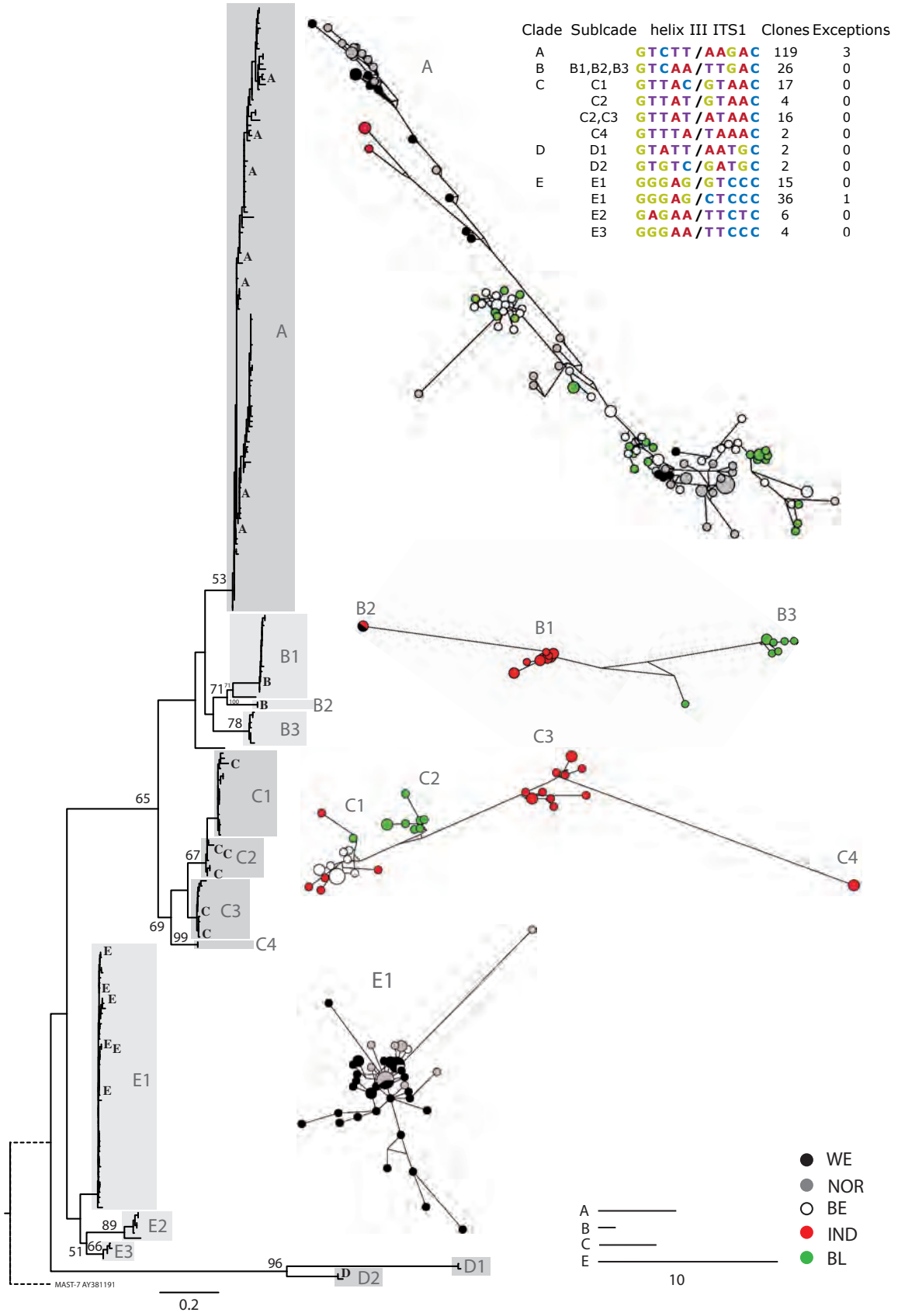


Figure 2. Maximum likelihood phylogenetic tree of MAST-4 built with 228 new sequences (287 bp of 18S rDNA plus the complete ITS1) and 22 clones from a previous study (indicated with letters). Clades and subclades are indicated with gray areas. Bootstraps values above 50% are shown. The scale bar indicates 0.2 substitutions per position. At the right of the tree there are the median-joining networks for each clade, highlighting ribotype differences among sequences (in different color depending the library). Scale bars indicate 10 base-changes between ribotypes (note different scale for each clade). At the top right, sequences for the conserved helix III stem (derived from ITS1 secondary structures) are listed.

two, with only one hemiCBC in the fifth base pair, included in the E1 subclade.

Biogeographical patterns from sequence analyses

To examine biogeographic patterns from the 5 samples, at a finer level assessed by ITS1 sequences, we constructed median joining networks for each clade (Figure 2, right panels), which allow visualization alternative potential evolutionary paths in datasets with large sample sizes and small genetic distances between individuals, and calculated Fixation indices (Fst), a measure of the genetic distance among separate populations. Median joining networks revealed variation in genetic diversity between clades, as the ratio between the estimated number of mutations in the network with respect the number of ribotypes (unique ITS1 sequences) sampled was 2 in clade E, 3 in clade A, 6 in clade C and 15 in clade B.

For clade A, sequences from the five sites were present, and there were few base changes between ribotypes as evidenced by short connecting lines in the median joining network. Fixation indices (Fst), were generally low between the different geographic populations of clade A, indicating a considerable gene flow (Figure 3). The Indian population (IND) appeared the most isolated, with Fst values > 0.40 between other locations, whereas the other four populations had much lower pairwise Fst values (< 0.22). The highest gene flow occurred between two warm sites (Sargasso Sea [BE] and Mediterranean [BL]) and two cold-temperate sites (North Pacific [WE] and North Atlantic [NOR]). In contrast, the median joining network for clades B and C exhibited a strong spatial structuring, with sequences from the different locations typically forming separate subclades (Fst values > 0.80 indicated the lack of gene flow between subclades). Within subclade C1 a high gene flow was seen between BE and IND populations (Fst = 0.18). Finally, sequences from subclade E1, originating from three different regions, were well mixed in the network plot and there were two cases of identical sequences from distant sites (Figure 2). The gene flow between these populations was rather high, with particularly low Fst between WE and NOR (Figure 3).

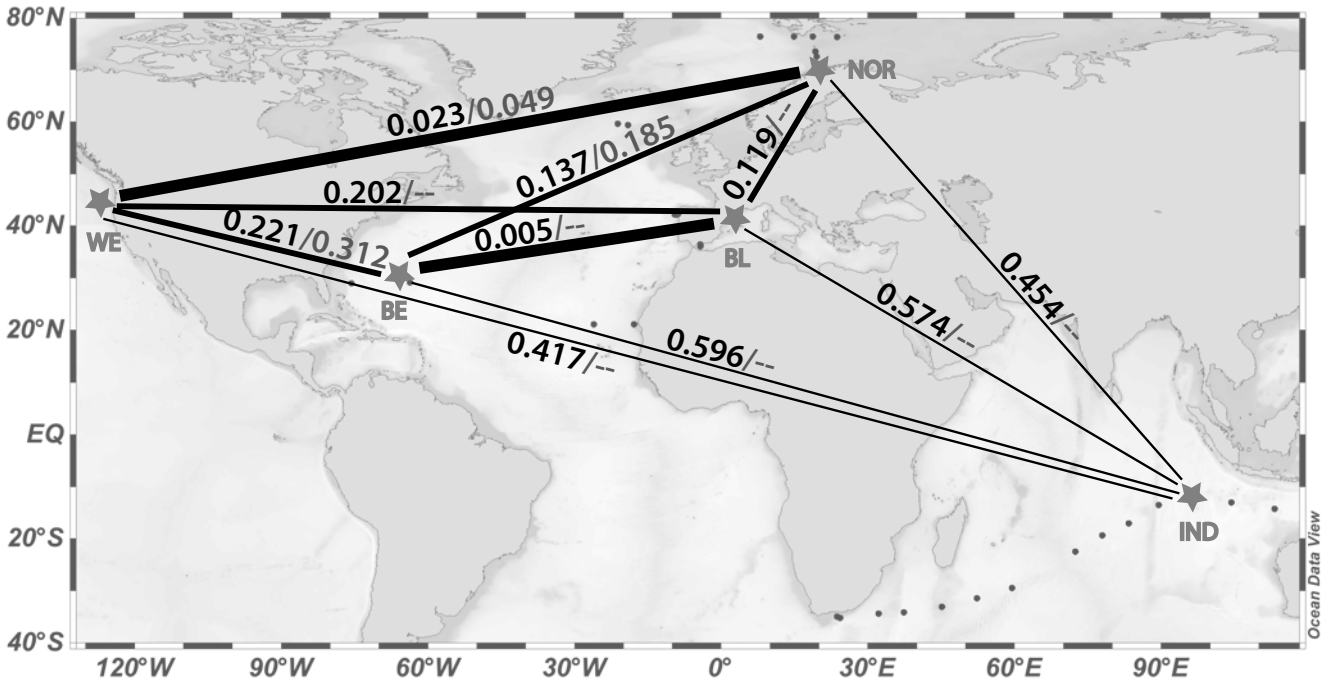


Figure 3. Suggested gene flow within populations of clades A and E1 among the five sites with clone libraries, with the strength of the flow indicated by the thickness of the line. Fst values for clades A (first value) and E1 (second value, when available) are shown above the lines.

Agreement between ARISA profiles and clone libraries

In order to interpret ARISA data, we first analyzed the fingerprints from the same samples that originated the clone libraries. ARISA profiles presented a similar pattern to the size distribution of clones in the libraries (Figure 4), suggesting that all peaks derived from MAST-4 phylotypes (although measured ARISA peaks tended to be 5-9 bp larger than the actual fragment size). We then combined

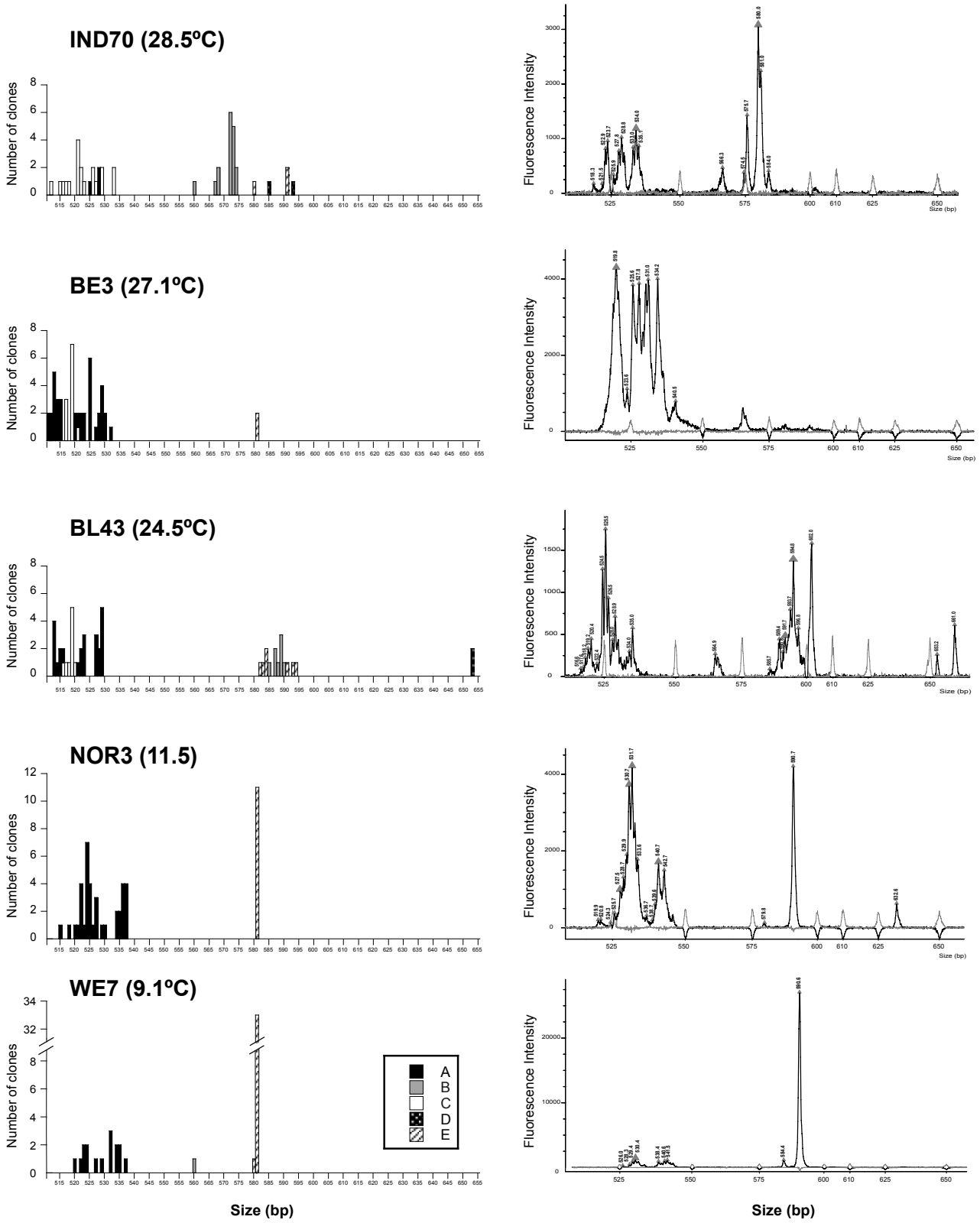


Figure 4. Representation of the sequence size of the clones per each clone library (left) and the corresponding ARISA profile obtained from the same sample (right). The ARISA standard is represented with the gray peaks.

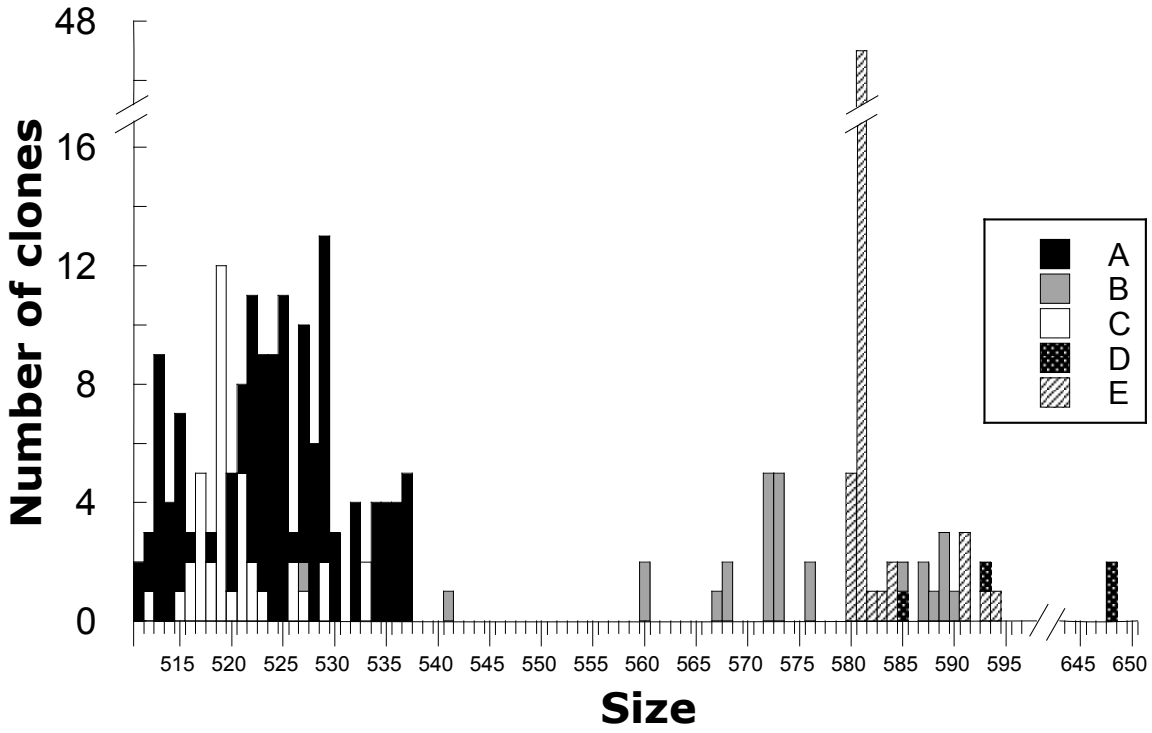


Figure 5. Summary of clone sizes for the 5 libraries (228 sequences) colored according to the corresponding MAST-4 clade

all clones from the different libraries to obtain a taxonomic assignment of ARISA peaks (Figure 5). Clones from clades A and C overlapped at the 511-537 bp region, although some sizes were exclusive to a given clade, such as 513, 525, and 528 (clade A) or 517, and 519 (clade C). Clones from clade B exhibited a wider size range, from 527 to 590 bp, and were the only ones occupying the size spectra from 560 to 574 bp. Clade D, with only four sequences, had the largest clones (648 bp) and also appeared in the 580-594 region intermixed with other clades. Clade E had clones in a restricted region (580-584 bp), and the majority were 581 bp.

Biogeographical patterns from ARISA fingerprints

We investigated the biogeographical patterns of MAST-4 assemblages by comparing the ARISA fingerprints from 107 samples obtained from 40 separate geographic sites. For most sites (25 out of 40), there were several depths sampled (subsurface to 100 or 250 m), whereas a subset of samples (n=23) derived from a temporal study in the Mediterranean site (BL). Samples were analyzed in triplicate, which were averaged for further statistical analyses since they were highly similar (see Materials and Methods). The correlation between genetic and geographic distance matrices done with a Mantel test (Figure 6a) was weak but significant (r= 0.28, p= 0.001), suggesting that geographic distance was not important in explaining sample composition. A partial Mantel test with geographic distance conditioned by temperature indicated a weaker correlation (r= 0.14, p= 0.001) giving even less importance to the geographic distance, once the effect of the main environmental variable (see below) was removed. Interestingly, a group of samples from very distant sites (~15,000 km) were highly similar, whereas samples in the same geographic location could be both, highly similar or very different. In contrast a Mantel test comparing the OTUs distance with the temperature distance (Figure 6b) showed a very high and significant correlation (r=0.60, p=0.001). Analysis of the OTUs distance matrix by NMDS did not reveal any trend for the different cruises (geographic location), but suggested a clear grouping according to sample temperature (Figure 6c). To create these groups, we used a separate dendrogram (not shown) where cold samples (from 2 to 9.4 ºC) grouped in a cluster with 82% similarity, and temperate (from 9.5 to 16.9 ºC) and warm (from 17 to 30ºC) groups

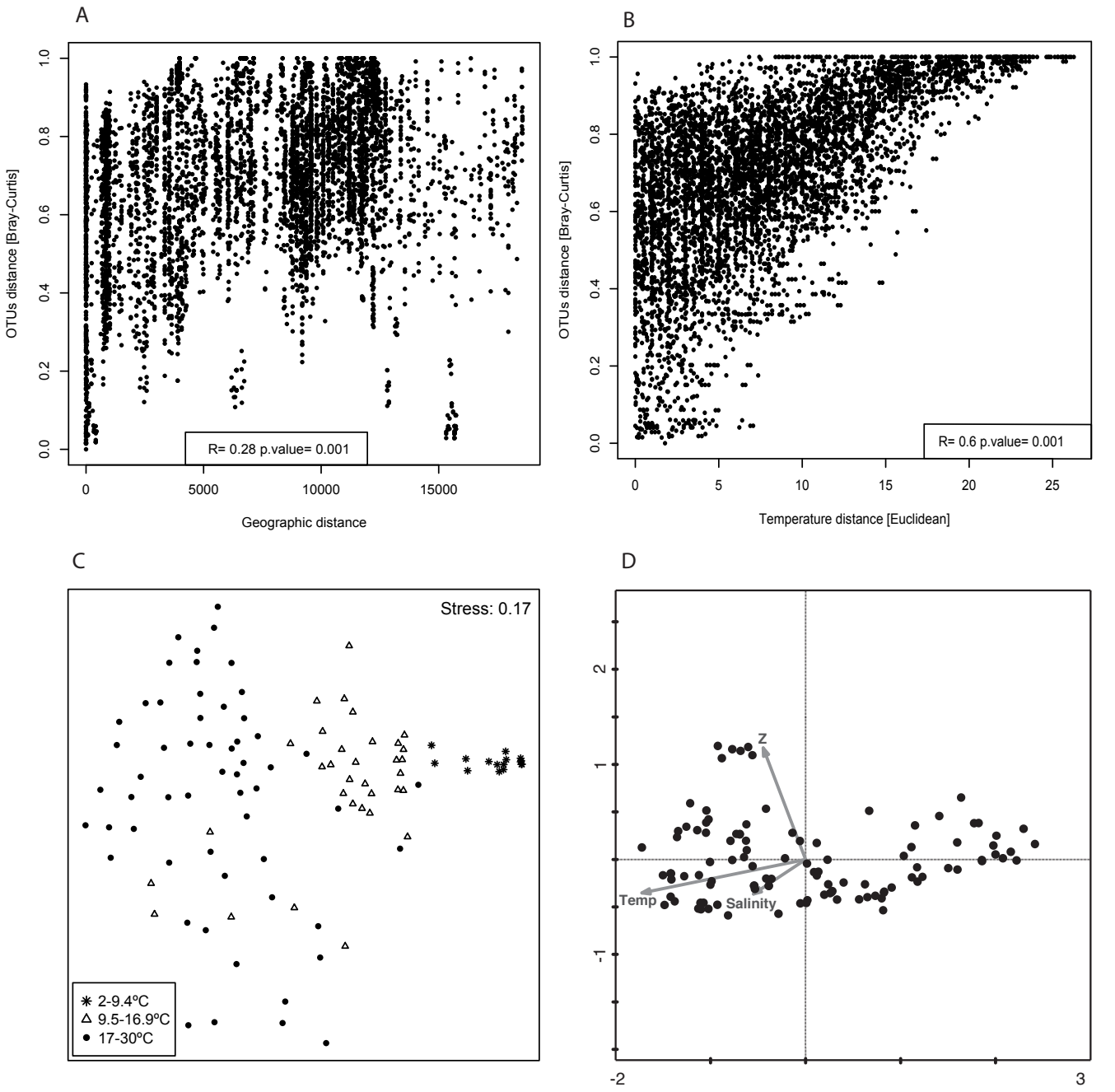


Figure 6. Statistical analysis of the global dataset of ARISA fingerprints. (**a**) Mantel test comparing MAST-4 composition (OTUs distance) and geographic distance. (**b**) Mantel test comparing MAST-4 composition and environmental (temperature) distance. (**c**) Non-metric multidimensional scaling (NMDS) diagram using Bray-Curtis dissimilarity from the OTU-table (OTUs distance matrix). Different symbols represent the temperature grouping. (**d**) Constrained Correspondence Analysis (CCA) diagram of the OTU-table constrained for the three most important factors: temperature (Temp), salinity (Sal) and sampling depth (Z). Arrows represent the direction and magnitude of the environmental factor gradient.

also appeared, although with some exceptions. The actual boundary between temperate and warm groups derived from the clustering of BL samples (see below).

We performed several analyses to identify and quantify the factors driving community changes. A Constrained Correspondence Analysis (CCA) (Figure 6d, Table 1) confirmed that temperature was the factor explaining most of the variance (51%), followed by sampling depth (Z), salinity, bottom depth (Zmax, a proxy of coast-offshore scenario) and chlorophyll concentration. The sum of these factors explained 91% of the variability. A second analysis was done after grouping the factors at discrete

Table 1. Separate statistical analysis (CCA, CCA per groups and PERMANOVA) to estimate the contribution (from R^2 values) of five environmental factors in the ARISA fingerprinting variance.

	CCA		**CCA** (Separate temperature groups)						**PERMANOVA**		
			COLD		**TEMPER.**		**WARM**				
Variable	R^2	Sign	R^2	Sign	R^2	Sign	R^2	Sign	**Range**	R^2	Sign
Temperature (ºC)	**0.51**	***	0.01		0.12	**	0.21	***	2_9.5, 9.5_17, 17_30	**0.36**	***
Z (m)	0.18	***	0.02		**0.41**	***	0.13	***	0-25, 25-75, 75-180, 180-250	0.08	***
Salinity	0.11	***	**0.06**	**	0.10	*	**0.25**	***	32, 33, 34, 35, 36, 37, 38, 39	0.07	***
Zmax (m)	0.06	***	0.01		0.06		0.08	**	0-200, 200-1000, 1000-6000	0.02	*
Chlorophyll (μg L^{-1})	0.05	*	0.01		0.04		0.15	***	0.1-0.5, 0.5-1.5, 1.5-4	0.02	*
Total	0.91		0.06		0.63		0.82			0.55	

Significance codes: ***:< 0.001; **0.001- 0.01; *0.01- 0.05
Bold numbers represent the highest R^2 of the particular analysis

levels. A PERMANOVA test considering grouped factors showed that temperature explained 36% of the variability in community composition, whereas all factors together explained 55% (Table 1). Since temperature appeared as the most important factor, we tested if the established groups (cold, temperate and warm) were statistically different by the ANOSIM pairwise tests. Indeed, they were significantly different in all cases, with a global R of 0.44 (significance level 0.1%) and highest R-value (0.88) for the cold-warm comparison. The ANOSIM test for the other environmental factors generally gave lower R-values, and was significantly different in only a few cases. Finally, when samples within each temperature group were analyzed separately, temperature was less important than salinity and sampling depth (Table 1).

We then searched for the OTUs driving the differences among cold, temperate and warm samples. The SIMPER test identified five OTUs with a dissimilarity contribution between groups higher than 5% (Table 2). OTU 589 had the largest dissimilarity contribution in all pairwise analyses, especially when including cold samples. This was consistent with the 99% similarity contribution of OTU 589 to the cold group (Table 2) and its predominant presence in cold waters (Figure 7). OTU 527 was the most important in warm samples (22% similarity contribution) and contributed 7% to the dissimilarity between all group comparisons with warm samples. OTUs 529 and 531 appeared important in temperate and warm samples, whereas OTU 567 was characteristic for temperate samples and absent in cold samples. Using the agreement between ARISA and clone libraries peak profiles we predicted that OTU 589 corresponded to a clone of 581 bp in size, belonging to clade E, whereas OTUs 527, 529 and 531 (clone sizes between 519 and 524 bp) could belonged to either clades A or C and OTU 567 (clone size 560 bp) belonged to clade B. Therefore, clade E was the best adapted to cold water, clades A and/or C were typical of temperate and warm waters while clade B was most characteristic of temperate waters.

Table 2. Results of the SIMPER analysis to identify the contribution of the five most important OTUs to the dissimilarity and similarity of the groups defined by temperature.

		Dissimilarity contribution %			**Similarity contribution %**		
OTU	Clade assignment	TEMP & COLD	WARM & COLD	TEMP & WARM	COLD	TEMPERATE	WARM
589	E	47.53	47.8	21.38	98.75	43.32	4.59
527	A or C	3.74	7.09	7.37	0.04	6.17	21.89
529	A or C	7.4	4.77	6.42	0.17	12.46	12.22
531	A or C	5.25	3.75	3.98	0.08	8.47	13.06
567	B	4.75	3.11	5.83	0	4.8	3.72

Contributions below 5% are represented in gray

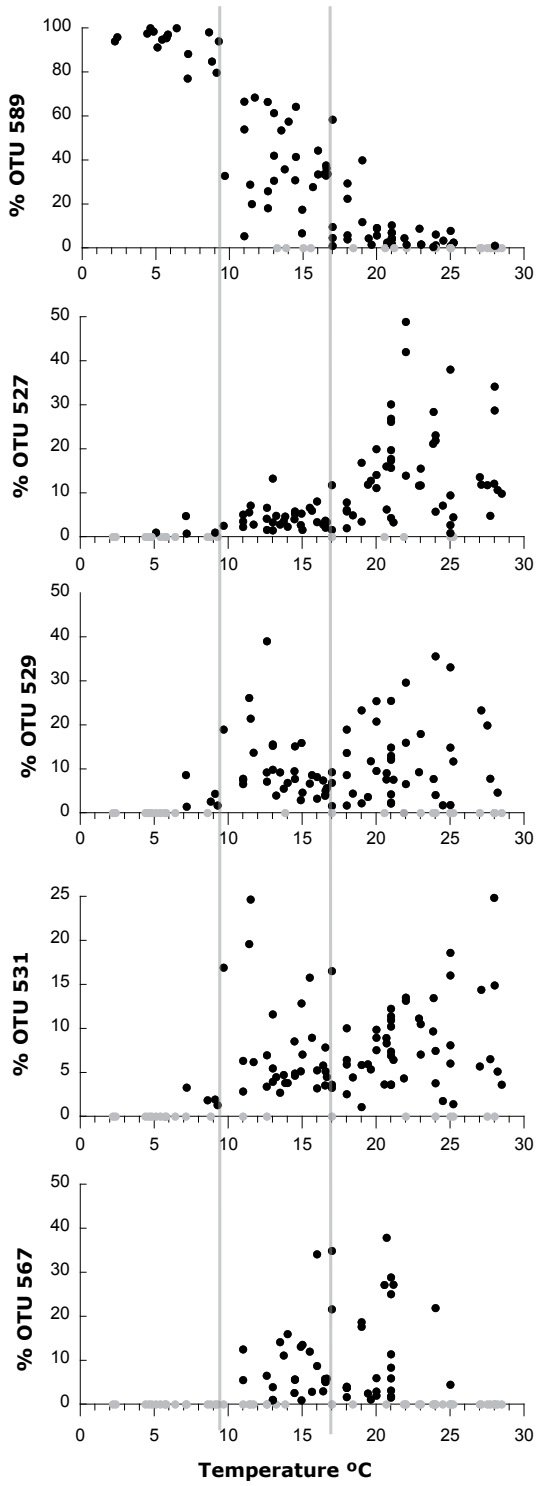


Figure 7. Relative intensity of the most important OTUs in the ARISA fingerprints in all samples displayed according to temperature. Gray circles represent samples where the OTU was not detected.

Temporal changes at Blanes Bay

At the Blanes Bay (BL) station in the NW Mediterranean Coast, we analyzed 23 different ARISA fingerprints covering a monthly seasonal sampling in 2003 and random dates from 2001 to 2006. Samples exhibited a large variability in the composition of MAST-4 and grouped, in an associated dendrogram (not shown), in two clusters differentiated by seawater temperature (above or below 17°C), as highlighted in the NMDS plot (Figure 8). The similarity of temperate samples was above 58% and the similarity of warm samples was above 37%. Each group contained samples from different years, highlighting the importance of temperature in defining community composition along this temporal scale.

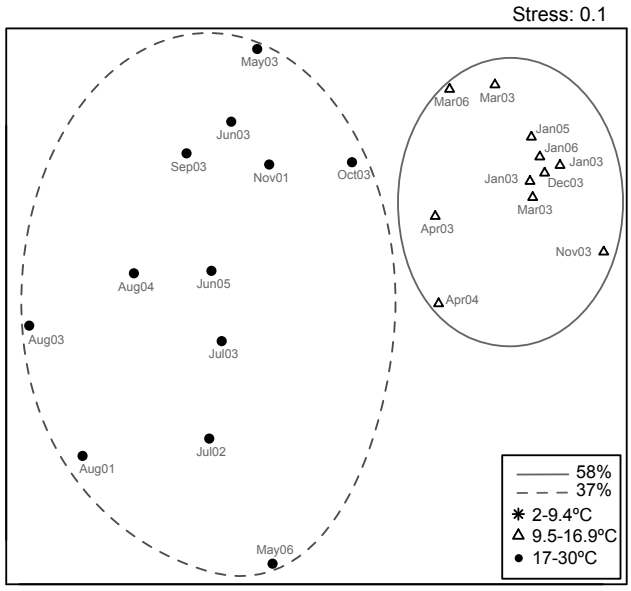


Figure 8. NMDS statistical analysis of ARISA fingerprints in the temporal study at Blanes Bay. Samples are displayed with symbols for temperature as in Figure 6 and with the date of sampling (month-year). Points enclosed by dashed line cluster at 37% similarity and points enclosed by solid line cluster at 58% similarity in a separate (not shown) dendrogram.

Discussion

Despite the recent interest in microbial biogeography as a mechanism to better understand the ecological and evolutionary properties of microbial taxa, this field has suffered from diverse conceptual and methodological limitations that confuse the emerging conclusions (Martiny et al 2006). A first problem can be how the microbial taxa are sampled, as many studies are based in a handful of isolated strains from separate geographic sites (Ki and Han 2005, Kooistra et al 2008) In the present study, the method of sampling is not based on culturing

but on molecular markers easily amplified from environmental DNA. This has the advantage that the sampled diversity is more representative of the natural assemblage (by avoiding culturing bias), and that many individuals (sequences) can be retrieved at once.

Second, it is very important to choose a good taxonomic marker, because the view that no biogeographic patterns exist in microorganisms (Finlay et al 2006) could be caused by a blurry identification of protist species (Lomolino et al 2006). Indeed many protist species appear to be widely distributed when identified via morphology, but these "morphological species" most likely include cryptic species that are not able to interbreed (Dolan 2006). Our approach, initially based on the 18S rDNA, also targeted the highly variable ITS1 region, which provides enhanced taxonomic resolution for diversity and biogeography studies (Brown and Fuhrman 2005). The ITS size variation is the basis of the fingerprinting technique ARISA, which can differentiate OTUs at high taxonomic resolution. ITS1 has been used to assess the diversity of *Pseudo-nitzschia* populations at inter- and intraspecific levels (Orsini et al 2004) and for elucidating *Pseudo-nitzschia* community structure using ARISA fingerprinting on environmental samples (Hubbard et al 2008). Here we amplify the ITS1 region and analyze this marker at the sequence level (in clone libraries), and at a fingerprinting level to compare assemblages. Clone libraries and ARISA profiles showed a remarkable agreement, with a shift of 5-9 bp depending of the length of the sequence, which is similar to previously reported shifts (Collins et al 2010, Hahn et al 2001). Thus, we confirm the ITS region as a good marker for studying microbial biogeography.

Third, it is important to clearly determine the genetic complexity of the studied taxa. We focus here in the uncultured protist MAST-4, which is known to exhibit a limited evolutionary diversification (Rodríguez-Martínez et al 2012). By designing specific primers amplifying the ITS1 region of MAST-4, we increased the sequences within this group by ten-fold, and we still did not have evidence of more than 12 species, as judged by the base of the conserved helix III stem in the ITS1 secondary structure. This confirmed the low diversity of MAST-4, adding only two more putative species. Clade A was still a single species, and all 119 clones (with 3 exceptions) presented the same helix III motif. Subclade E1 seemed to form also a single species. Despite having two helix III motifs it presented very well mixed populations between sites and the difference between these motifs was by only one hemiCBC in the more variable fifth base pair. The structure of MAST-4 in a few genetically distinct species is similar to that found in other marine protists, such as the cosmopolitan diatom *Skeletonema costatum* (Kooistra et al 2008). In addition, MAST-4 has the advantage that it is a very abundant and widespread group, so ideal as a model for biogeographical studies. It has been found in a wide range of temperatures (2-30ºC) as the biliphytes lineage (Cuvelier et al 2008).

Finally, a limitation of many studies is the inability to disentangle the relative contribution of contemporary environmental conditions and historical contingencies in shaping the spatial variations of microbial communities (Martiny et al 2006). Our study addresses this question by analyzing spatial changes together with the associated factors of the samples. In addition, since single sampling may not cover the complete biodiversity of a given place, we also evaluated temporal changes. In fact, it has been suggested that to obtain reliable estimates of biodiversity in a given habitat it is necessary to cover the entire season (Nolte et al 2010).

The analysis of MAST-4 assemblages using ARISA fingerprints showed that temperature was the main factor influencing the distribution patterns, as has been observed in other marine microbes like *Prochlorococcus* (Martiny et al 2009). Distant samples sharing the same temperature could have similar MAST-4 diversity, whereas samples with different temperature were very different. This global pattern was also observable in a temporal study in the same geographic site (Blanes Bay). In this place with a thermal seasonal cycle from 12 to 24°C, samples grouped again by temperature, confirming this as an important structuring factor. Moreover, the Fst analysis revealed a very large gene flow

between samples geographically distant but with a similar temperature. It is worth mentioning that, in theory, a similar pattern could have been obtained due to another driving factor not measured here but strongly correlated with temperature (Martiny et al 2009). Thus, the spatial distribution of MAST-4 seemed to be mainly controlled by contemporary environmental factors with a null or low degree of provincialism. In addition, temperature had a strong positive correlation with OTU richness (r = 0.54), as has been previously reported (Fuhrman et al 2008, Pommier et al 2006). The positive effect of temperature on diversity could be due to the kinetics of biological processes, including rates of reproduction, dispersal, species interaction, adaptive evolution, and speciation (Allen 2002, Rohde 1992). Finally, other abiotic factors that are known to be important in defining microbial community composition, such as salinity (Logares et al 2009) and sampling depth (Winter et al 2008), took importance within each defined temperature group.

A more detailed analysis of clade A, probably constituting a single species, revealed that it was widely distributed (it appeared in all five clone libraries) and very well mixed. The five populations examined from this clade exhibited a high gene flow, and samples from distant places with the same temperature displayed completely mixed populations in the MJ network with few estimated mutations per ribotype. A similar scenario occurred within subclade E1, probably representing a single species as well. This subclade even had a lower proportion of mutations per ribotype. On the contrary, clades B and C, probably composed by more than one species, exhibited some spatial structuring, with subclades appearing in only one library. Moreover, the mutations per ribotype increased two times in clade C or five times in clade B as compared with clade A. This intriguing feature, perhaps indicating some dispersal restriction for clades B and C, could also be due to undersampling and deserves more attention in future surveys. If confirmed by further results, the dispersal limitation of clades B and C would be comparable to that found with the cosmopolitan marine planktonic diatom *Pseudo-nitzschia pungens* (Casteleyn et al 2010).

Our data also highlight specific lineages within MAST-4 that seem adapted to different temperature regimes. Thus, clade B, and either or possibly both clades A or C seemed to be characteristic of temperate and warm waters, whereas clade E1, represented by OTU 589, was the only one able to inhabit cold waters. The genetic structure of MAST-4 with different lineages, some ubiquitous in the oceans and with particular ecological properties, resembles that of other marine picoeukaryotes. Thus, different *Ostreococcus* (Rodríguez et al 2005) and *Synechococcus* lineages (Ahlgren and Rocap 2006) are adapted to different light levels, whereas a *Micromonas pusilla* clade adapted to cold temperature has also been reported (Lovejoy et al 2006). It has been proposed that this ecotypic differentiation can partly explain the success of these picoeukaryotes, allowing them to exploit the whole spectrum of habitat variability.

The uncultured free-living protist MAST-4 is widely abundant and very small, and thus possesses the properties for a worldwide distribution (Finlay 2002). Moreover it lives in a marine habitat, where wind, waves and currents produce mixing events that facilitate the dispersion. However, we did not observe all MAST-4 clades at all locations but saw biogeographical patterns, stressing the importance of the end of the tenet, "but the environment selects". It is reasonable to hypothesize that MAST-4 has a huge dispersal capacity and can arrive everywhere within a marine habitat. For instance, there is one record of a MAST-4 sequence in the Arctic Ocean (Lovejoy and Potvin 2011), showing the potential to arrive to such high latitudes probably dragged by coastal currents of Pacific water. But then, depending the environmental conditions, different organisms will settle and grow up, resulting in different community patterns. These are related to some form of ecological differentiation between related types, as described for *Skeletonema* species (Kooistra et al 2008). Although microorganisms could spread across all suitable habitats, local adaptations eventually reduce the gene flow and promote speciation (Medlin 2007).

As a conclusion, we did not see strong marine geographical barriers for the dispersal of MAST-4, whereas temperature was the main driver for community composition, as has been observed in marine bacterial and archaeal assemblages (Winter et al 2008). In our study the environment seemed to make a taxonomic selection, with clades with supposed physiological adaptations established in different regions. In particular, clade E1 represented by OTU 589 was the only one acclimated to cold waters. Our study also confirmed the low MAST-4 diversity previously detected.

Acknowledgements

Funding has been provided by FPI fellowships from the Spanish Ministry of Education and Science to RRM and GS and by projects FLAME (CGL2010-16304, MICINN, Spain) to RM, GEMMA (CTM2007-63753-C02-01/MAR, MEC) to Carlos Pedrós-Alió, and ATOL (NSF, 0629521) to GR. We thank R.E. Collins for DAx software help, E. Sañé for Primer software help, R. Logares for statistical advices and C. Williams for molecular help.

References

Ahlgren NA, Rocap G (2006). Culture Isolation and Culture-Independent Clone Libraries Reveal New Marine Synechococcus Ecotypes with Distinctive Light and N Physiologies. *Appl Environ Microbiol* **72:** 7193-7204.

Allen AP (2002). Global Biodiversity, Biochemical Kinetics, and the Energetic-Equivalence Rule. *Science* **297:** 1545-1548.

Alonso-Sáez L, Arístegui J, Pinhassi J, Gómez-Consarnau L, González JM, Vaqué D *et al* (2007). Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean. *Aquat Microb Ecol* **46:** 43.

Alonso-Sáez L, Vázquez-Domínguez E, Cardelús C, Pinhassi J, Sala MM, Lekunberri I *et al* (2008). Factors Controlling the Year-Round Variability in Carbon Flux Through Bacteria in a Coastal Marine System. *Ecosystems* **11:** 397-409.

Arin L, Morán X, Estrada M (2002). Phytoplankton size distribution and growth rates in the Alboran Sea (SW Mediterranean): short term variability related to mesoscale hydrodynamics. *J Plankton Res* **24:** 1019.

Baas-Becking LGM (1934). Geobiologie of inleiding tot de milieukunde. *The Hague, the Netherlands: WP Van Stoc- kum & Zoon* (in Dutch) **18**.

Bandelt H, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16:** 37-48.

Beijerinck MW (1913). De infusies en de ontdekking der backteriën. *Jaarboek van de Koninklijke Akademie van Wetenschappen***:** 1–28.

Broecker WS (1991). The great ocean conveyor. *Oceanography* **4:** 79-89.

Brown M, Fuhrman J (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41:** 15-23.

Casteleyn G, Leliaert F, Backeljau T, Debeer A, Kotaki Y, Rhodes L *et al* (2010). Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc Natl Acad Sci USA* **107:** 12952-12957.

Cermeño P, Falkowski PG (2009). Controls on diatom biogeography in the ocean. *Science* **325:** 1539-1541.

Cermeño P, De Vargas C, Abrantes F, Falkowski PG (2010). Phytoplankton Biogeography and Community Stability in the Ocean. *PLoS ONE* **5:** e10037.

Collins RE, Rocap G, Deming JW (2010). Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ Microbiol* **12:** 1828-1841.

Cuvelier ML, Ortiz A, Kim E, Moehlig H, Richardson DE, Heidelberg JF *et al* (2008). Widespread distribution of a unique marine protistan lineage. *Environ Microbiol* **10:** 1621-1634.

De Wit R, Bouvier T (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8:** 755-758.

del Giorgio PA, Condon R, Bouvier T, Longnecker K, Bouvier C, Sherr E *et al* (2011). Coherent patterns in bacterial growth, growth efficiency, and leucine metabolism along a northeastern Pacific inshore–offshore transect. *Limnol Oceanogr* **56:** 1-16.

Díez B, Massana R, Estrada M, Pedrós-Alió C (2004). Distribution of eukaryotic picoplankton assemblages across hydrographic fronts in the Southern Ocean,

studied by denaturing gradient gel electrophoresis. *Limnol Oceanogr* **49:** 1022-1034.

Dolan J (2006). Microbial biogeography? *J Biogeogr* **33:** 199-200.

Fenchel T, Finlay BJ (2004). The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54:** 777-784.

Finlay B, Esteban G, Brown S, Fenchel T, Hoef-Emden K (2006). Multiple cosmopolitan ecotypes within a microbial eukaryote morphospecies. *Protist* **157:** 377-390.

Finlay BJ (2002). Global Dispersal of Free-Living Microbial Eukaryote Species. *Science* **296:** 1061-1063.

Finlay BJ, Fenchel T (2004). Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155:** 237-244.

Fisher MM, Triplett EW (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65:** 4630-4636.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL *et al* (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105:** 7774-7778.

González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C *et al* (2000). Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol* **66:** 4237-4246.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M *et al* (2004). Spatial scaling of microbial eukaryote diversity. *Nature* **432:** 747-750.

Hahn M, Wilhelm J, Pingoud A (2001). Influence of fluorophor dye labels on the migration behavior of polymerase chain reaction--amplified short tandem repeats during denaturing capillary electrophoresis. *Electrophoresis* **22:** 2691-2700.

Hewson I, Fuhrman J (2006). Improved strategy for comparing microbial assemblage fingerprints. *Microb Ecol* **51:** 147-153.

Horner-Devine MC, Carney KM, Bohannan BJM (2004). An ecological perspective on bacterial biodiversity. *P Roy Soc Lond B Bio* **271:** 113-122.

Horner-Devine MC, Silver JM, Leibold MA, Bohannan BJM, Colwell RK, Fuhrman JA *et al* (2007). A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* **88:** 1345-1353.

Hubbard K, Rocap G, Armbrust E (2008). Inter- and intraspecific community structure within the diatom genus Pseudo-nitzschia (Bacillariophyceae). *J Phycol* **44:** 637-649.

Katoh K, Toh H (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* **9:** 286-298.

Ki J, Han M (2005). Molecular analysis of complete SSU to LSU rDNA sequence in the harmful dinoflagellate *Alexandrium tamarense* (korean isolate, HY970328M). *Ocean Sci J* **40:** 43-54.

Kooistra WHCF, Sarno D, Balzano S, Gu H, Andersen RA, Zingone A (2008). Global Diversity and Biogeography of Skeletonema Species (Bacillariophyta). *Protist* **159:** 177-193.

Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K (2009). Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* **17:** 414-422.

Lomolino MV, Riddle BR, Brown JH (2006). Biogeography. *Sinauer Associates, Inc, Sunderland, Massachusetts, ed 3*.

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

Lovejoy C, Potvin M (2011). Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. *J Plankton Res* **3:** 431-444.

MacArthur RH, Wilson EO (1967). *The theory of island biogeography*.

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

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

Massana R, DeLong EF, Pedrós-Alió C (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol* **66:** 1777-1787.

Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A *et al* (2004). Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol* **70:** 3528-3534.

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

Massana R (2011). Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65:** 1-47.

Medlin L (2007). If everything is everywhere, do they share a common gene pool? *Gene***:** 1-4.

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

Not F, Massana R, Latasa M, Marie D, Colson C, Eikrem W *et al* (2005). Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50:** 1677-1686.

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

O'Malley MA (2008). 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Stud His Phil Biol Biomed Sci* **39:** 314-325.

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.

Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P *et al* (2006). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16:** 867-880.

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

Rambaut A (2009). FigTree. Institute of Evolutionary Biology, University of Edinburgh, United Kingdom. http://tree.bio.ed.ac.uk/software/figtree.

Ramette A (2009). Quantitative Community Fingerprinting Methods for Estimating the Abundance of Operational Taxonomic Units in Natural Microbial Communities. *Appl Environ Microbiol* **75:** 2495-2505.

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

Rodríguez-Martínez R, Labrenz M, Del Campo J, Forn I, Jürgens K, Massana R (2009). Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environ Microbiol* **11:** 397-408.

Rodríguez-Martínez R, Rocap G, Logares R, Romac S, Massana R (2012). Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4). *Mol Biol Evol***:** e-pub ahead of print 13 December 2011, doi: 2010.1093/molbev/msr2303.

Rohde K (1992). Latitudinal gradients in species diversity: the search for the primary cause. *Oikos***:** 514-527.

Sexton PF, Norris RD (2008). Dispersal and biogeography of marine plankton: Long-distance dispersal of the foraminifer Truncorotalia truncatulinoides. *Geology* **36:** 899-902.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688-2690.

ter Braak C, Smilauer P (2002). CANOCO reference manual on CanoDraw for Windows User's guide: Software for Canonical Community Ordination (version 4.5). *Ithaca, NY: Microcomputer Power*.

Vyverman W, Verleyen E, Sabbe K, Vanhoutte K, Sterken M, Hodgson DA *et al* (2007). Historical processes constrain patterns in global diatom diversity. *Ecology* **88:** 1924-1931.

Whitaker RJ (2003). Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science* **301:** 976-978.

Whitfield J (2005). Biogeography: Is Everything Everywhere? *Science* **310:** 960-961.

Winter C, Moeseneder MM, Herndl GJ, Weinbauer MG (2008). Relationship of Geographic Distance, Depth, Temperature, and Viruses with Prokaryotic Communities in the Eastern Tropical Atlantic Ocean. *Microb Ecol* **56:** 383-389.

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

Synthesis of Results and General Discussion

The main goal of this thesis was to study the ecology of the widespread uncultured marine stramenopiles MAST-4, as a model of a marine heterotrophic flagellate. Different topics have been addressed, including the extent of its distribution, abundance, trophic role, genetic structure, evolutionary patterns and biogeography. In the following discussion I will evaluate the results obtained in each of these topics separately. Overall, this thesis confirms the important role of the MAST-4 taxon in marine planktonic environments.

Abundance and distribution

To assess the abundance and distribution of the uncultured heterotrophic flagellate MAST-4 in large sets of environmental samples we present a robust Q-PCR (quantitative PCR) protocol for a fast quantification of rDNA molecules **(chapter 1)**. A difficulty when dealing with uncultured organisms is that target molecules are quantified instead of target cells. To convert the abundance of molecules to cells, the gene copy number must be known. The rDNA operon copy number of MAST-4, estimated by comparing Q-PCR and FISH (Fluorescent In Situ Hybridization) signals from the same samples, was around 30. This relatively low number is consistent with its small size, 2-3 µm in diameter (Massana et al 2006a), and fits well within the relationship described for 18 phytoplankton strains (Zhu et al 2005) (figure S.1). However the copy number was not the same in the different sets of samples analyzed. We consider that the main cause of the larger variability and the lower rDNA copy number in environmental samples was that they were not collected to be quantitatively processed. Indeed, the extent to which this protocol gives absolute abundances depends on the care devoted during sample collection and DNA extraction. Nevertheless, an alternative explanation for this noisier signal in environmental samples could be that they are composed by distinct MAST-4 lineages whereas incubations could be selecting a single ribotype. As seen in **chapter 4**, the MAST-4 group was indeed composed by several species, which could have a different copy number in a similar range and appeared mixed in environmental samples. In addition, the Blanes incubation seemed to select a single ribotype.

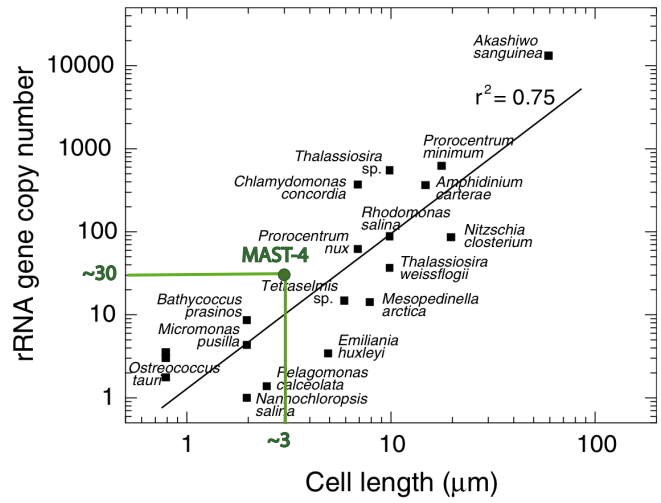


Figure S.1. Correlation between rDNA copy number estimated by QPCR and cell length from 18 strains of phytoplankton (Zhu et al 2005). The estimated position of MAST-4 is included in the figure.

Taking into account our environmental samples constrains, it is clear that the Q-PCR approach **(chapter 1)** provides useful global views of MAST-4 distributional patterns (presence and predominance) and minimal estimates of its rDNA abundance. This was applied to a large sample collection at different depths from different oceanographic cruises, including a transect from South Africa to Australia crossing the Indian Ocean Gyre, a coastal-offshore section at the northern part of the Western Alboran Gyre and a transect across the Polar Front in the Drake Passage, from South Atlantic to Antarctic waters (figure 3, **chapter 1**). A similar pattern was detected at the Indian Ocean and Alboran Sea studies, with higher abundances in coastal than offshore samples and with a presence all along the photic zone, being almost twice more abundant at the DCM (deep chlorophyll maximum) than at surface. MAST-4 presented very low abundance in the upper aphotic zone and was hardly detected deeper below, which is in accordance to the fact that heterotrophic flagellates (HF) and their bacterial food are becoming scarcer in mesopelagic waters (Fukuda et al 2007, Tanaka and Rassoulzadegan 2002). Seawater temperature seemed to be the second constrain in the distribution of MAST-4. Thus, MAST-4 molecules were only detected in the two northernmost stations in the Drake Passage, those with warmer temperatures. In the other stations, with temperatures below 5°C, target molecules were undetected. Thus, MAST-4 seemed to be excluded in waters below 5°C, an intriguing feature shared by other microorganisms such as marine picocyanobacteria (Partensky et al 1999).

Besides the three oceanographic cruises, monthly samples from the coastal Blanes Bay Microbial Observatory were analyzed for a period of six years (figure 4, **chapter 1**). A clear seasonal pattern of MAST-4 abundance was not observed and abrupt changes occurred between consecutive dates. One reason could be that MAST-4 varies on a shorter time scale and sampling once a month does not properly describe its temporal variation. Alternatively, as proposed in (Piwosz and Pernthaler 2010), the apparent lack of seasonality might derive successions of different lineages. In **chapter 4** we determined that MAST-4 was composed by at least 5 different biological species. And in **chapter 5** we observed that the MAST-4 assemblage in Blanes Bay depended on sample temperature. So, different MAST-4 lineages succeeded in this coastal station along the thermal seasonal cycle and this could cause the lack of seasonal pattern when analyzing the whole group. Disentangling the particular dynamics of these lineages is a challenge for future studies.

The average abundances of MAST-4 in the Indian Ocean and the Mediterranean Sea, systems with comparable mean temperatures (16-24°C), were similar, around 800 rDNA molecules ml^{-1} (table 1, **chapter 1**). MAST-4 was absent from the cold Antarctic waters and present at low abundances north of the South Atlantic Front. Virtually all samples from epipelagic waters (surface to 120 m) and with temperatures above 5°C have MAST-4 molecules. Moreover additional FISH counts presented in **chapter 1** were added and the average abundance of 131 cells ml^{-1} (Massana

et al 2006b) was supported with a mean of 100 cells ml^{-1} at surface and 150 cells ml^{-1} at the DCM. This broad and systematic presence of MAST-4 in marine planktonic systems is similar to that described in some marine bacteria like SAR11 (Morris et al 2002), *Roseobacter* (Selje et al 2004) and *Prochlorococcus* (Partensky et al 1999). Nevertheless, we have to keep in mind the phylogenetic diversity of MAST-4. So, although they look the same by FISH (Massana et al 2006a), they include distinct lineages with different and complementary ecological adaptations **(chapter 4** and **5)** that might explain this broad distribution, as has been proposed for other picoeukaryotes (Rodríguez et al 2005).

Trophic role

To study the trophic role of uncultured HF living in natural assemblages we present an approach based on the estimation of the feeding activity of specific grazers detected by FISH after short-term ingestion experiments with tracer preys. As tracers we included the commonly used fluorescently labeled bacteria (FLB) **(chapter 2** and **3)** and two bacterial strains of the family Rhodobacteraceae (MED479) and the family Flavobacteriaceae (MED134), of various cell sizes, which were offered alive and detected after the ingestion by a secondary FISH step **(chapter 2)**. The aim of this section was to determine grazing rates and prey preferences of MAST-4.

Grazing rates of MAST-4 in *in situ* and incubated samples were comparable (**chapter 2**). This suggests that MAST-4 cells were not artificially stimulated by the incubation and the higher ingestion rates in the incubated sample were due to higher prey abundance. Moreover, important differences in grazing rates were seen when using different tracers. FLBs gave the lowest ingestion rates (1 bacteria $predator^{-1}$ h^{-1}) and MED134 the highest rates (3 bacteria $predator^{-1}$ h^{-1}). These differences could not be explained by tracer cell sizes, since rates varied highly using tracers of similar biovolume. A clear and statistically significant pattern emerged when relating grazing rates with the percentage of live cells in tracers (figure 3, **chapter 2**). So, MAST-4 appeared to prefer bacteria in good physiological state, with grazing rates 2-3 times higher on live bacteria than dead FLB. These results fit well with the general view that the use of live bacteria as tracers in ingestion experiments results in higher grazing rates than the use of FLB (Boenigk et al 2001, Landry et al 1991) and the preferably consumption of growing versus starving bacteria (González et al 1993). An extreme case of negative selection was seen in the experiments using heat-killed MED134 cells that were not ingested at all. Finally, besides differences related to cell viability, no other differences in measured grazing rates were seen when using the two bacterial strains. Therefore, the Rhodobacteraceae and Flavobacteriaceae cells were ingested equally by MAST-4, so the putatively important differences in phylogeny and life strategy between both strains did not determine prey preference.

The specific grazing rates of MAST-4 measured in **chapter 2** are comparable to *in situ* rates measured in other environmental studies (Unrein et al 2007, Vaqué et al 1994), but much lower than most estimates derived from cultured strains. However, these comparisons suffer from the actual value of prey abundance during the grazing estimations, which are known to influence dramatically the rates measured, giving the form of functional response (Holling 1959). The functional responses and numerical responses (relationship of growth rates with prey abundance) have been recurrently studied in cultured strains of heterotrophic flagellates. However, to our knowledge, there are no studies analyzing functional responses of uncultured flagellates. To study it, we provided FLB at different abundances, becoming in most bottles the main prey, and analyzed the short-term ingestion experiments with FISH counts of grazers in mixed assemblages **(chapter 3)**. To minimize the problem of low predator densities, we did our grazing experiments using an unamended seawater incubation that are known to promote the growth of MAST-4. For the first time, the functional response (maximum ingestion rate and half-saturation constant, Ks) of an uncultured heterotrophic flagellate has been determined. The maximum ingestion rate was comparable with previous grazing estimates (figure S.2), and the most remarkable finding was that the Ks for MAST-4 was 8.7 10^5 prey ml^{-1} (figure 3, **chapter 3**). Thus, this widely distributed and abundant heterotrophic flagellate appears well adapted to the bacterial abundances of marine planktonic environments, typically around 10^6 bacteria ml^{-1} (Fuhrman and Hagström 2008). In contrast, the Ks of cultured HF are typically at least one order of magnitude higher, ranging from 1.1 to 45 10^6 bacteria ml^{-1} (table 2, **chapter 3**). These higher Ks are the expected values for organisms that grow efficiently in rich media, and at the same time establish an obvious limitation for their development at the prevailing low *in situ* bacterial abundances.

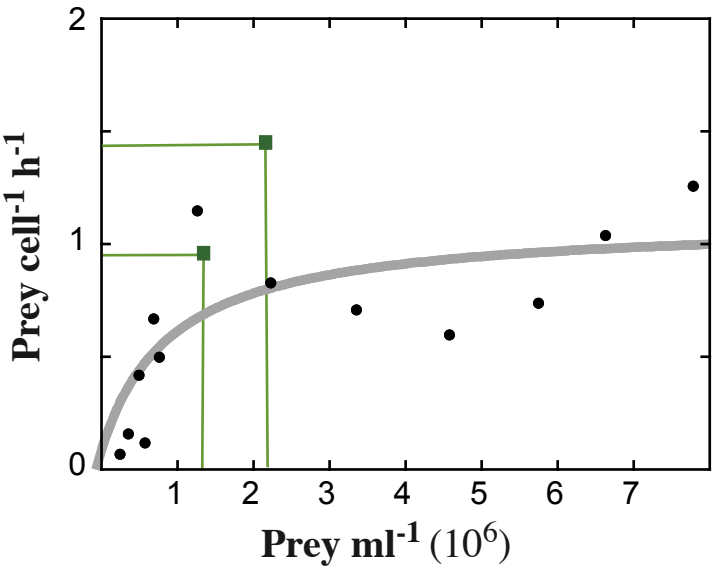


Figure S.2. Functional response of MAST-4 using FLBs (**chapter 3**), with separate grazing rates done in **chapter 2** overlayed.

In **chapter 2** we also compared the grazing rates of MAST-4 with those of another uncultured flagellate, MAST-1C, and the whole eukaryotic assemblage. We clearly showed functional differences between both taxa, although the underlying mechanisms for such differences remained hidden. Comparing grazing rates using FLBs, MAST-4 appeared less active and MAST-1C more active than the eukaryotic assemblage (figure 4, **chapter 2**). This was consistent with MAST-1C being larger than MAST-4. With respect to tracer cell viability, the pattern observed for the eukaryotic assemblage was similar to that of MAST-4, with MED134 giving highest rates than FLBs. In contrast, MAST-1C deviated from this picture, with FLB yielding the highest rates and live bacteria being ingested at lower rates (MED134 was not ingested at all). Clearly, the food preferences of MAST-4 and MAST-1C were distinct and the latter did not behave as the typical bacterivores in the assemblage. A plausible explanation was that the boundary of optimal prey size for MAST-1C fell within the size range of the tested bacteria. MED134, being the smallest of the bacteria tested, could be outside the size range of edible bacteria and escape predation. The fact that MAST-1C could be adapted to feed on larger bacteria than MAST-4 is consistent with its larger size, following the established relationship between predator and prey size (Fenchel 1987). It is commonly accepted that prey size is the main factor in grazing vulnerability (González et al 1990), with sharp size boundaries outside which preys can not be ingested (Fenchel 1987, Jürgens and Matz 2002). Then, when all preys are within the edible size range, other factors interplay with a smaller impact. For instance, MAST-4 preferred live bacteria in good physiological state, but still fed on dead FLB at one third of the maximal rate. Our data did not reveal differential feeding behavior related to the phylogenetic affiliation of the tested bacteria.

Moreover, different taxa inside the mixed community presented different functional responses (figure 3, **chapter 3**), therefore had different ecological niches, perhaps even with different filtration strategies and prey preferences. *Minorisa minuta* candidatus, a recently cultured flagellate using techniques mimicking natural conditions (del Campo 2011), was well adapted to low prey abundances being very efficient in ingesting FLBs and filtering water. Together with MAST-4, both taxa appear as typical bacterivores in natural assemblages, as their functional responses were comparable to that determined for the whole eukaryotic assemblage. Nevertheless, *Minorisa* appeared to be more voracious than MAST-4, with maximal ingestion rates five times higher, and this probably explains the higher measured growth rates of *Minorisa* than MAST-4 (1.56 versus 0.62 d^{-1}). One reason for the low feeding rates of MAST-4, seen in both **chapters 2** and **3,** could be that it selects negatively against heat-killed FLBs, preferring live bacteria in good physiological state, although is it likely that this holds true for most HF taxa in the natural assemblage (Fu et al 2003, Landry et al 1991). Finally, we were lucky enough to include a typical cultured heterotrophic flagellate in the grazing experiments. *Paraphysomonas* spp. did not achieve saturation of the ingestion rate even at the highest prey concentrations assayed (near 10^7). With these data we should expect a high

Ks, similar to that of other cultured flagellates. Strains of *Paraphysomonas* are typically retrieved from cultures where they feed on large bacteria at very high densities. No wonder that they are not abundant in the sea.

Our studies reveal important functional differences between distinct uncultured protists and can be a useful starting point to disentangle the complexity of microbial food webs. Different flagellate taxa seem to have different grazing rates and prey preferences, so, the functional diversity observed here gives an ecological meaning to the large phylogenetic diversity of marine heterotrophic protists (Vaulot et al 2002). For instance, recent data obtained through single cell sequencing (Martínez-García et al 2012) suggests that MAST-4 could be adapted to graze on a specific prey, as the small *Pelagibacter ubique*. In addition, our results set the basis for the fundamental differences between cultured and uncultured bacterial grazers and might explain why the dominant marine bacterivores are not cultured by classical approaches. Natural HF had a Ks of 6-9 10^5, lower than that of the traditional cultured flagellates, thus being well adapted to typical bacterial abundances of marine planktonic environments.

Phylogenetic structure and evolutionary patterns

The MAST-4 global population size is estimated to be about 10^{24} cells. The way that this huge amount of cells group in different lineages can be assessed by the phylogenetic structure of the extant rDNA sequences. In turn, this structure can be used to infer the evolutionary patterns of the group. For a better understanding of the phylogenetic structure of MAST-4, and of their biological meaning, in **chapter 4** we sequenced a large fragment of the rDNA operon, including the highly variable ITS (Internal Transcribed Spacer) region and the beginning of the 28S. We also compiled all publicly available MAST-4 18S rDNA sequences and included new 454 reads from the V4 region of the 18S rDNA. Moreover in **chapter 5** we increased ten times the number of sequences within this group, encompassing the V9 region of the 18S rDNA and the ITS1 region.

Establishing the number of main clades

The phylogenetic tree with the distinct MAST-4 18S rDNA sequences retrieved from our exhaustive GenBank search revealed only five main clades (figure 2, **chapter 4**), each one exhibiting low sequence divergence (uncorrected genetic distance below 0.012). Pyrosequences, retrieved from a vast environmental protist survey, added more than one order of magnitude of sequences and confirmed the low MAST-4 diversity, as all 454 reads affiliated to the five known clades, without any additional clade. The only divergent GenBank sequence that did not match any new pyrosequence could be a pseudogene (Thornhill et al 2007). The 18S rDNA tree for the new clones sequenced in

chapter 4 was consistent with the same clades. The 28S rDNA tree displayed the same five clades as the 18S rDNA tree, although they appeared better resolved and separated with longer phylogenetic distances. Finally, the tree based on the end of the 18S rDNA and the very variable ITS1 region revealed the same five clades previously observed, sometimes organized in separate subclades (figure 2, **chapter 5**). So, despite the presence of a huge number of MAST-4 cells in the oceans and the analysis of hundreds to thousands of environmental sequences, its diversity is structured into only five main clades.

The hypothetical number of species

Different approaches have been used for the theoretical delimitation of different biological species within MAST-4 **(chapters 4 and 5)**. First, we considered the ITS2 secondary structure (figure 5, **chapter 4**), since it has been shown that strains exhibiting at least one compensatory base change (CBC) in the conserved nucleotides of helices II (5 paired bases) and III (18-30 continuous positions) belonged to different biological species (Coleman 2003, Coleman 2007, Coleman 2009). The presence of a hemiCBC (one sided) could allow some weak degree of interbreeding (Coleman 2009). For instance, two *Pseudo-nitzschia* strains differing by three hemiCBCs produced zygotes but never gave viable offspring, thus being considered as separate species (Amato et al 2007). A second criterion was the inspection of analogous regions in the ITS1 secondary structure**.** Whereas the ITS2 has been widely investigated, the ITS1 still lacks of a universal core secondary structure model. For our ITS1 sequences, we established a common core secondary structure with three helices and found a region with a similar behavior than the ITS2 with respect to species delimitation. It consisted in four base pairs at the stem III that were conserved within defined clades but differed with CBCs between them. When expanded to the fifth base pair, polymorphisms appeared within some clades (figure 6, **chapter 4**). Third, we contrasted the tree topologies recovered by the ITS1 and ITS2 regions (figure 4B, **chapter 4**). We hypothesize that groups that have diversified enough to constitute different species will display congruent topologies in their ITS1 and ITS2 trees, since no recombination would exist among these markers. In contrast, groups that may still constitute one single species might display incongruent topologies due to recent recombination events. Basically, this is the concordance–discordance principle used for the recognition of phylogenetic species (Taylor et al 2000), which can be readily applied using these rapidly evolving spacers (Coleman 2007, Mullineux and Hausner 2009). Forth, we compared the intraclade divergence in the ITS region of MAST-4 clades with that observed in other species (tables 1 and 2, **chapter 4**). Finally, we applied tools from population genetics in the species delimitation: median-joining (MJ) networks (Bandelt et al 1999) that infer intraspecific phylogenies and visualize alternative potential evolutionary paths (figure 2, **chapter 5**), and the fixation index (Fst) that estimates the genetic differentiation between populations (figure 3, **chapter 5**).

We combined these five criteria to investigate the correspondence of phylogenetic clades and biological species within MAST-4. ITS1 and ITS2 secondary structures differed by CBCs among clades in the conserved regions, indicating that each clade is clearly a separate biological species. Then, we looked within each particular clade. Clade A seemed to be composed of only one species. It showed no polymorphism in the critical ITS regions, the tree topologies were incongruent and the sequence divergence was very low. Moreover, it presented a well mixed MJ network and a high gene flow between different sites. Clade B presented only one motif in the ITS conserved regions, although the ITS2 conserved fragment in the helix III was short. Sequence divergence was high, forming three phylogenetic subclades and exhibiting a strong separated structuring in the MJ network. For these reasons this clade may include two or three species. Clade C had one hemiCBC between several clones in helix II and a short conserved fragment for helix III of the ITS2 and four helix III motifs of the ITS1. The tree topology revealed three subclades and the MJ network presented a strong population structuring between sites. In addition, the intraclade divergence was similar to the average minimum divergence between species. So, we concluded that clade C might include three to four species. Clade D was undersampled, and the four available sequences already suggest two different species. Clade E presented four ITS1 motifs, two of them forming well differentiated lineages, and the other two, included in the E1 subclade, with only one hemiCBC in the more variable fifth base pair. Sequences within subclade E1 showed no polymorphisms in ITS2 conserved regions, had low sequence divergence, some incongruent ITS tree topologies, and well mixed populations between distant sites with a low Fst, so it seemed to form a single species. Overall, three species may be present in clade E. Despite the huge number of MAST-4 cells in the oceans, the current evidence after the inspection of 250 sequences derived from 5 different sites indicates a maximum of 13 separate biological species.

Low evolutionary diversification

MAST-4 diversity is structured into just five main clades, each representing at least one biological species with a current evidence of a maximum of 13 **(chapters 4** and **5)**. This lineage appears as a well-supported discrete group in 18S rDNA phylogenies, and the closest relative sequences are only 91% similar. In addition, the maximal 18S rDNA sequence divergence **(chapter 4)** within MAST-4 is 0.044, a very low value as compared with other protist groups (Pernice M, personal communication). Overall, the genetic divergence of MAST-4 is surprisingly low for such a widespread and abundant organism, indicating a very low evolutionary diversification within the group. This low evolutionary diversification points to either a very recent evolutionary divergence and worldwide dispersal or to a very strong environmental filtering that penalizes any deviation from an optimal cell design. A similar scenario of low diversity and cosmopolitan distribution seems to exist in other picoeukaryotes, such as the prasinophyte *Micromonas* (Slapeta et al 2006).

Biogeography

MAST-4 was present in virtually all samples from epipelagic waters and with temperatures above 5°C **(chapter 1)**. However, it has been determined that the MAST-4 group includes several species with sequences retrieved coming from distant oceanic sites **(chapter 4)**. The question of the distribution of each of these species still remains to be elucidated and defining particular specific distributions can also give some clues to ecotypic differentiation. Sampling of MAST-4 diversity was based on molecular markers easily amplified from natural samples. This approach, initially based on the 18S rDNA, also targeted the highly variable ITS1 region, which provides enhanced taxonomic resolution for diversity and biogeography studies (Brown and Fuhrman 2005). The ITS1 region has been used for studying inter- and intraspecific population variations (Orsini et al 2004) or to compare assemblages through ARISA (Automated Ribosomal Intergenic Spacer Analysis) fingerprinting (Hubbard et al 2008). In **chapter 5**, to study the biogeography of MAST-4, we combined ARISA and 18S-ITS1 clone libraries and examined both spatial and temporal community changes and associated environmental factors. We confirmed the ITS region as a good marker for studying microbial biogeography.

The analysis of MAST-4 assemblages using ARISA fingerprints showed that temperature was the main factor influencing the distribution patterns (figure 6, **chapter 5**). This has been observed in other marine microbes like *Prochlorococcus* (Martiny et al 2009). Mantel tests between genetic and geographic distances gave weak correlations, indicating that spatial distance was not important in explaining sample composition. In contrast the correlation was high comparing temperature instead of geographic distances. In a second analysis, distributing the samples in a 2-D space (NMDS plot) did not reveal any trend for the different cruises (geographic location), but suggested a clear grouping according to sample temperature. Distant samples sharing the same temperature could have similar MAST-4 diversity, whereas samples with different temperature were very different. Further statistical tests confirmed that temperature was the factor explaining most of the variance (CCA and PERMANOVA) and that temperature groups were significantly different among them (ANOSIM). This global pattern was seen in a temporal study in a single geographic site (with a 12-24°C thermal cycle) (figure 8, **chapter 5**). Samples from different years, grouped again by temperature, confirmed the importance of this structuring factor also in a temporal scale. Moreover, the Fst analysis revealed a very large gene flow (within clades A and E1) between distant samples at similar temperatures. Other abiotic factors that are known to be important in defining microbial community composition, such as salinity (Logares et al 2009) and sampling depth (Winter et al 2008), took importance within each defined temperature group. Thus, the spatial distribution of MAST-4 seemed to be mainly controlled by contemporary environmental factors with a null or low degree of provincialism.

A more detailed analysis of clade A, probably constituting a single species, revealed that it was widely distributed (it appeared in all five clone libraries) and very well mixed (figure 2, **chapter 5**). The five populations examined from this clade exhibited a high gene flow (figure 3, **chapter 5**) and samples from distant places with the same temperature displayed completely mixed populations in the MJ network with few estimated mutations per ribotype. A similar scenario occurred within subclade E1, probably a single species as well. This subclade even had a lower proportion of mutations per ribotype. On the contrary, clades B and C, probably composed by more than one species, exhibited some spatial structuring, with subclades appearing in only one library and with many more mutations per ribotype among subclades. This intriguing feature, perhaps indicating some dispersal restriction for clades B and C, could also be due to undersampling and deserves more attention in future surveys. If confirmed by further results, the dispersal limitation of clades B and C would be comparable to that found with the cosmopolitan marine planktonic diatom *Pseudo-nitzschia pungens* (Casteleyn et al 2010).

There were five OTUs (ARISA peaks) that drove the differences among cold, temperate and warm samples, highlighting specific MAST-4 lineages that seem adapted to different temperature regimes. Thus, clade B, and either or possibly both clades A or C, seemed to be characteristic of temperate and warm waters, whereas clade E1, represented by OTU 589, was the only one able to inhabit cold waters. The genetic structure of MAST-4 with different lineages, some ubiquitous in the oceans and with particular ecological properties, resembles that of other marine picoeukaryotes. Thus, different *Ostreococcus* (Rodríguez et al 2005) and *Synechococcus* lineages (Ahlgren and Rocap 2006) are adapted to different light levels, whereas a *Micromonas pusilla* clade adapted to cold temperature has also been reported (Lovejoy et al 2006). It has been proposed that this ecotypic differentiation can partly explain the success of these picoeukaryotes, allowing them to exploit the whole spectrum of habitat variability.

The uncultured free-living protist MAST-4 is widely abundant and very small, and thus possesses the properties for a worldwide distribution (Finlay 2002). Moreover it lives in a marine habitat, where wind, waves and currents produce mixing events that facilitate the dispersion. However, we did not observe all MAST-4 clades at all locations but saw biogeographical patterns, stressing the importance of the end of the tenet, "but the environment selects". It is reasonable to hypothesize that MAST-4 has a huge dispersal capacity and can arrive everywhere within a marine habitat. For instance, there is one record of a MAST-4 sequence in the Arctic Ocean (Lovejoy and Potvin 2011), showing the potential to arrive to such high latitudes probably dragged by coastal currents of Pacific water. But then, depending on the environmental conditions, different organisms will settle and grow up, resulting in different community patterns. These are related to some form of ecological differentiation between related types, as described for *Skeletonema* species (Kooistra

et al 2008). Although microorganisms could spread across all suitable habitats, local adaptations eventually reduce the gene flow and promote speciation (Medlin 2007). As conclusion, the widespread MAST-4 protist is an ideal model for studying microbial biogeography, displaying no dispersal limitation in marine systems together with strong environmental selection.

Conclusions

1) MAST-4 is a structural component of protist assemblages in marine temperate photic waters. It is present in virtually all samples from epipelagic waters (surface to 120 m) and with temperatures above ~5°C.

2) The average abundance of MAST-4 in systems with warm temperatures (16-24°C) was similar, at the range of 100 to 150 cells ml^{-1}. At colder temperatures it was less abundant.

3) Bacteria tested in this thesis (from 0.07 to 0.18 μm^3) have a cell size that is within the edible range for MAST-4. It typically eats 1 to 3 bacteria per hour, which defines it as not very voracious predator. Moreover, MAST-4 appears to prefer bacteria that are in good physiological state, with 2-3 times higher grazing rates of live bacteria versus dead FLB.

4) Natural heterotrophic flagellates had a Ks lower than that of the traditional cultured flagellates, thus being well adapted to typical bacterial abundances of marine planktonic environments. Specifically, MAST-4 presented a Ks of 8.7 10^5 bacteria ml^{-1}.

5) The functional diversity observed between MAST-4 and other heterotrophic flagellates (such as MAST-1C, *M. minuta* candidatus and *Paraphysomonas*) gives an ecological meaning to the high phylogenetic diversity of marine heterotrophic protists, with different taxa adapted to different ecological niches.

6) Despite the presence of a huge number of MAST-4 cells in the oceans its diversity is structured into just five main clades, each representing at least one biological species. The current evidence indicates a maximum of 13 separate species.

7) The genetic divergence of MAST-4 was surprisingly low for an organism so widespread and abundant, indicating a very low evolutionary diversification, pointing to either a very recent evolutionary divergence or to a very strong environmental filtering that penalizes any deviation from an optimal cell design.

8) We did not see marine geographical barriers for the dispersal of the most represented MAST-4 clades, whereas temperature was the main factor influencing the distribution patterns.

9) The environment seemed to make a taxonomic selection, with different clades of MAST-4 with supposed physiological adaptations established in different conditions. In particular, clade E1, represented by OTU 589, was the only one acclimated to cold waters. This ecotypic differentiation could partly explain the success of this picoeukaryote, allowing it to exploit the whole spectrum of habitat variability.

Future perspectives

1) We have seen throughout this thesis the importance of the MAST-4 lineage. Therefore, there is an urgent need to have it in culture. We think that this is feasible, specially when taking advantage of the new data gathered about this uncultured microbe, such as its feeding habits, environmental constrains, and optimized probes for fast molecular detection.

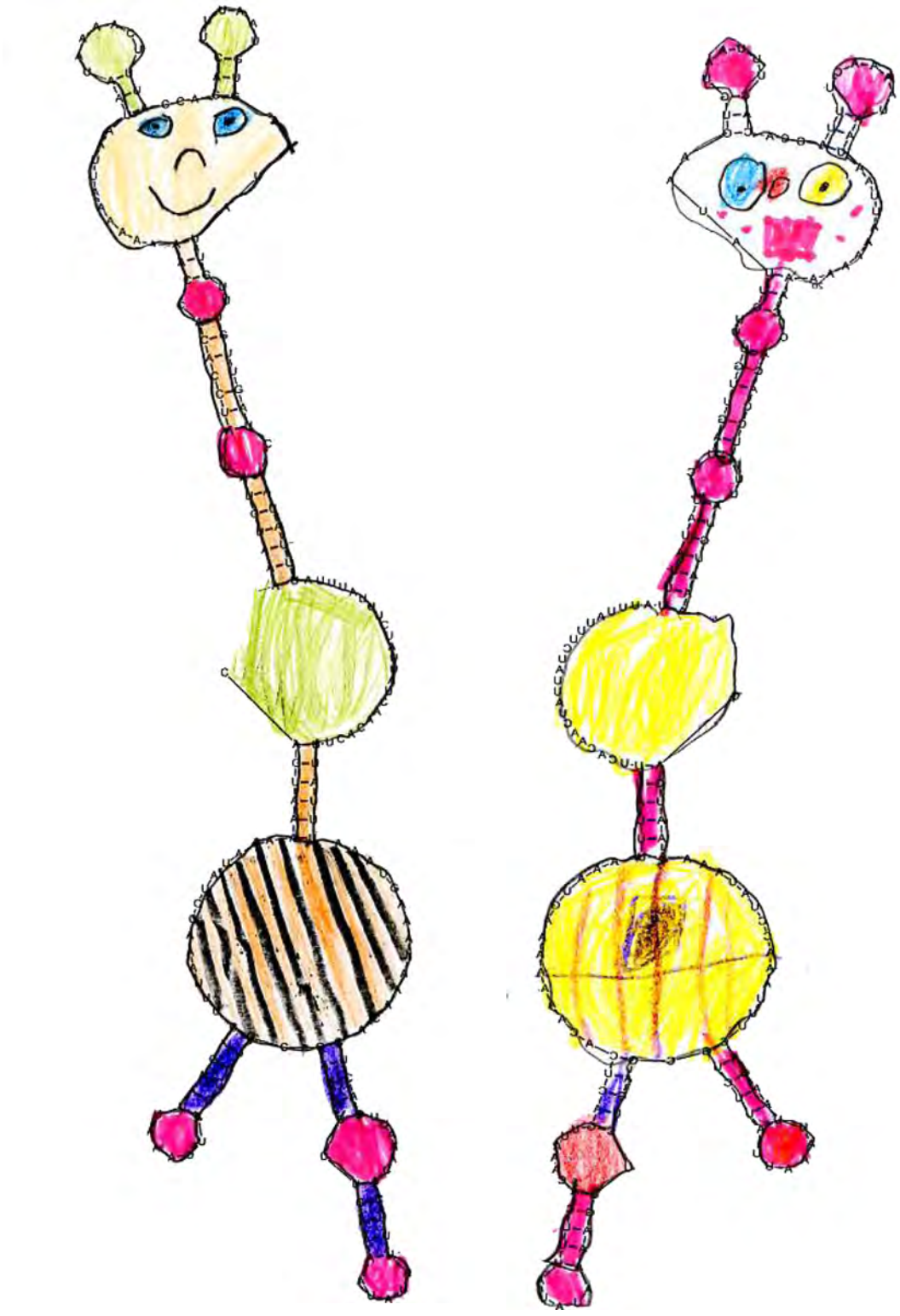
2) Clades A and E provide a general picture of biogeographic patterns, governed by temperature and not geographic distance. However, less clear results are seen in the other clades, probably due to undersampling. It would be very motivating to complete this study adding many more sequences from different sites. This could be done by exploiting the primers and PCR sets designed here for the use in massive sequencing like pyrosequencing. Increasing the number of sequences will indeed refine the population genetics within this protist.

3) We have proposed a conserved region in the ITS1 secondary structure for the delimitation of biological species, similarly to conserved regions at the ITS2. It would be interesting to corroborate this with breeding experiments.

4) Single amplified genomics is a promising approach to get the genomic content of these uncultured microbes and to identify specific biological interactions.

5) It would be interesting to integrate MAST-4 in microbial food webs, and answer how much carbon it is processing and, more intriguingly, what are its mortality factors (likely predation and perhaps viral infection).

Resumen de la tesis (Spanish summary)



Ecología de un linaje de flagelados heterotróficos no cultivados, el MAST-4

Abundancia, papel trófico, diversificación y biogeografía

Raquel Rodríguez Martínez

Barcelona, Mayo 2012

Glosario

18S ADNr. Gen codificante del ARN que compone la subunidad pequeña del ribosoma, es usado habitualmente para identificar y clasificar microorganismos eucariotas.

Alveolados marinos (MALV). Clados sin representantes cultivados que pertenecen al supergrupo eucariota de los alveolados. Han sido detectados a través de análisis moleculares del picoplancton marino.

Análisis automatizado del espaciador interno ribosómico (ARISA). Técnica de identificación genética basada en la variación del tamaño del ITS.

Análisis medioambientales moleculares. Recuperación de señales genéticas de ensamblajes microbianos complejos para diferentes estudios.

Cambio compensatorio de base (CBC). Mutaciones que ocurren en ambos nucleótidos de una posición estructural apareada del ARN ribosomal, siempre que se mantenga el enlace del par de nucleótidos.

Clado filogenético (o linaje). Conjunto de secuencias relacionadas originadas a partir de un único antecesor común.

Espaciador interno transcrito (ITS). Regiones no codificantes que separan los componentes individuales de las unidades del ADN ribosómico. Tienen un mayor grado de variación que las regiones génicas.

Estramenópilos marinos (MAST). Clados sin representantes cultivados que pertenecen al supergrupo eucariota de los estramenópilos. Han sido detectados a través de análisis moleculares del picoplancton marino.

Estructura secundaria. Interacciones entre los pares de bases dentro de una molécula de ARN que se pueden descomponer en tallos y bucles.

Flagelados. Protistas unicelulares, fototróficos o heterotróficos, con uno o más orgánulos, llamados flagelos, habitualmente usados para propulsarse o crear corrientes de alimentación.

Hibridación fluorescente *In Situ* (FISH). Método de microscopía para la detección de células microbianas por medio de una sonda fluorescente que se adhiere específicamente a los ribosomas.

Librería de clones. Colección heterogénea de secuencias clonadas (usualmente ADNr 18S) derivada de un ensamblaje complejo de microorganismos.

Microscopía de epifluorescencia. Técnica que permite la observación por fluorescencia de células pequeñas teñidas o autofluorescentes retenidas en un filtro.

Picoeucariotas. Protistas fototróficos y heterotróficos menores de 3 μm.

Población. Todos los organismos pertenecientes al mismo grupo o especie que viven en la misma área geográfica.

Protistas. Término general para los eucariotas que no pertenecen a las plantas, animales, hongos o algas macroscópicas. Generalmente son organismos unicelulares con un tamaño comprendido entre 1 μm y más de 100 μm.

Respuesta funcional. Tasa de depredación de un consumidor en función de la densidad de alimento.

Respuesta numérica. Tasa de crecimiento de un consumidor en función de la densidad de alimento.

Reacción cuantitativa en cadena de la polimerasa (Q-PCR). También llamada reacción en cadena de la polimerasa en tiempo real (RT-PCR). Técnica molecular basada en la PCR, la cual se usa para amplificar y cuantificar al mismo tiempo una molécula de ADN diana.

Taxón. Unidad sistemática para designar un nivel jerárquico en la clasificación de los organismos.

Acrónimos

ARISA	Análisis automatizado del espaciador interno ribosómico
CBC	Cambio compensatorio de base
CCTH	Criptófitas, Centrohelida, Telonemida y Haptófitas
DAPI	4,6-diamidino-2-fenilindol
FISH	Hibridación fluorescente *in situ*
FLB	Bacterias marcadas con fluorescencia
Fst	Índice de fijación
HF	Flagelados heterotróficos
ITS	Espaciador interno transcrito
Ks	Constante de semi-saturación
MALV	Alveolados marinos
MAST	Estramenópilos marinos
MJ	Unión media
OTU	Unidad taxonómica operativa
PF	Flagelados fototróficos
PCR	Reacción en cadena de la polimerasa
Q-PCR	Reacción cuantitativa en cadena de la polimerasa
RAS	Rhizaria, Alveolados y Estramenópilos
rDNA	Ácido desoxirribonucleico ribosomal
SAGs	Genomas aislados amplificados

Introducción general

El árbol de la vida eucariota

Los eucariotas son solo uno de los tres ámbitos de la vida, junto con las bacterias y las arqueas. Una de las razones de nuestra gran curiosidad por ellos es que incluyen los organismos que podemos ver. Nuestra comprensión de la biología, ecología y evolución de los eucariotas está dominada por el estudio de las plantas terrestres, animales y hongos. Sin embargo, estos son solo tres fragmentos aislados de la gran diversidad de los eucariotas existentes. La mayoría de los eucariotas, refiriéndonos a linajes principales, número de taxones diferentes y número total de células, están formados principalmente por linajes unicelulares. Un número sorprendente de estos linajes están escasamente caracterizados. No obstante, el conocimiento de la diversidad morfológica, funcional y ecológica de los eucariotas microbianos es fundamental para nuestra comprensión de la biología de los eucariotas y las fuerzas subyacentes que la forman (Baldauf 2008).

Durante la segunda mitad del siglo XX, los avances moleculares proporcionaron una manera sistemática de relacionar todos los organismos vivos a través de comparaciones de secuencias de ADN, utilizando inicialmente el gen de ARN de la subunidad pequeña ribosomal: ADNr 16S en procariotas y ADNr 18S en eucariotas (Woese 1987). Usando las secuencias de este gen, el árbol filogenético de la vida eucariota aparecía dividido en tan solo unos pocos supergrupos (Adl et al 2005, Baldauf 2003). Incluyendo varias revisiones y actualizaciones (Baldauf 2008), casi todos los eucariotas pueden ahora ser asignados a alguno de los supergrupos, los cuales configuran una estructura radial sin una clasificación clara entre ellos y con una raíz incierta (figura I.1). Hay pocos caracteres morfológicos o ultraestructurales que conecten los diversos linajes dentro de cada supergrupo, sin embargo las indicaciones filogenéticas son robustas (Jürgens y Massana 2008). Aunque existe una variación a la hora de configurar los supergrupos, el consenso general incluye (1) Unicontos, (2) Arqueoplástidos, (3) Rhizaria + Alveolados + Estramenópilos (RAS), (4) Excavados y (5) Criptófitos, Centrohelida, Telonemida más Haptófitos (CCTH). Los Unicontos incluyen todos los eucariotas que se consideran primitivamente uniflagelados, es decir, Opistocontos (incluidos los animales, hongos y algunos protistas tales como coanoflagelados) y Amebozoos (Cavalier-Smith 2002). Arqueoplástidos es el grupo en el que surgió por primera vez la fotosíntesis en los eucariotas e incluye las algas verdes y las plantas terrestres (Adl et al 2005, Archibald y Keeling 2005). El grupo RAS se ha propuesto recientemente (Burki et al 2007, Hackett et al 2007) para unir tres supergrupos muy heterogéneos, los Rhizaria (Cercozoos, Radiolarios y Foraminíferos), los Alveolados (Dinoflagelados, Ciliados y Apicomplejos) y los Estramenópilos (ver más abajo). Los Excavados están formados por dos grupos distintos, los Excavados mitocondriales que incluyen Euglénidos, Heteroloboseos y Jakóbidos y los Excavados amitocondriales (por ejemplo, Diplomonádidos y Parabasálidos), una colección de

taxones altamente derivados con la estructura celular interna simplificada y carentes de mitocondria aeróbica. CCTH es un nuevo supergrupo (Burki et al 2009) propuesto para relacionar varios linajes filogenéticos importantes pero difíciles de localizar, como los Haptófitos, Criptófitos y Telonemida (Shalchian-Tabrizi et al 2006). Además el CCTH también incluye los Ketablefáridos (sabiendo que están emparentados con los Criptófitos) (Okamoto y Inouye 2005) y quizás los Picobilifitas, una nueva clase de fitoplancton que inicialmente no aparecía relacionada con ningún supergrupo (Not et al 2007). Una consecuencia del marco molecular es que muchos protistas *incertae sedis* (Patterson y Zöffel 1991) están encontrando su posición filogenética en el árbol eucariota. Además, las secuencias del gen 18S ARNr de los representantes cultivados son cruciales en la colocación de los protistas dentro de este contexto filogenético (Cavalier-Smith y Chao 2003, Scheckenbach et al 2005).

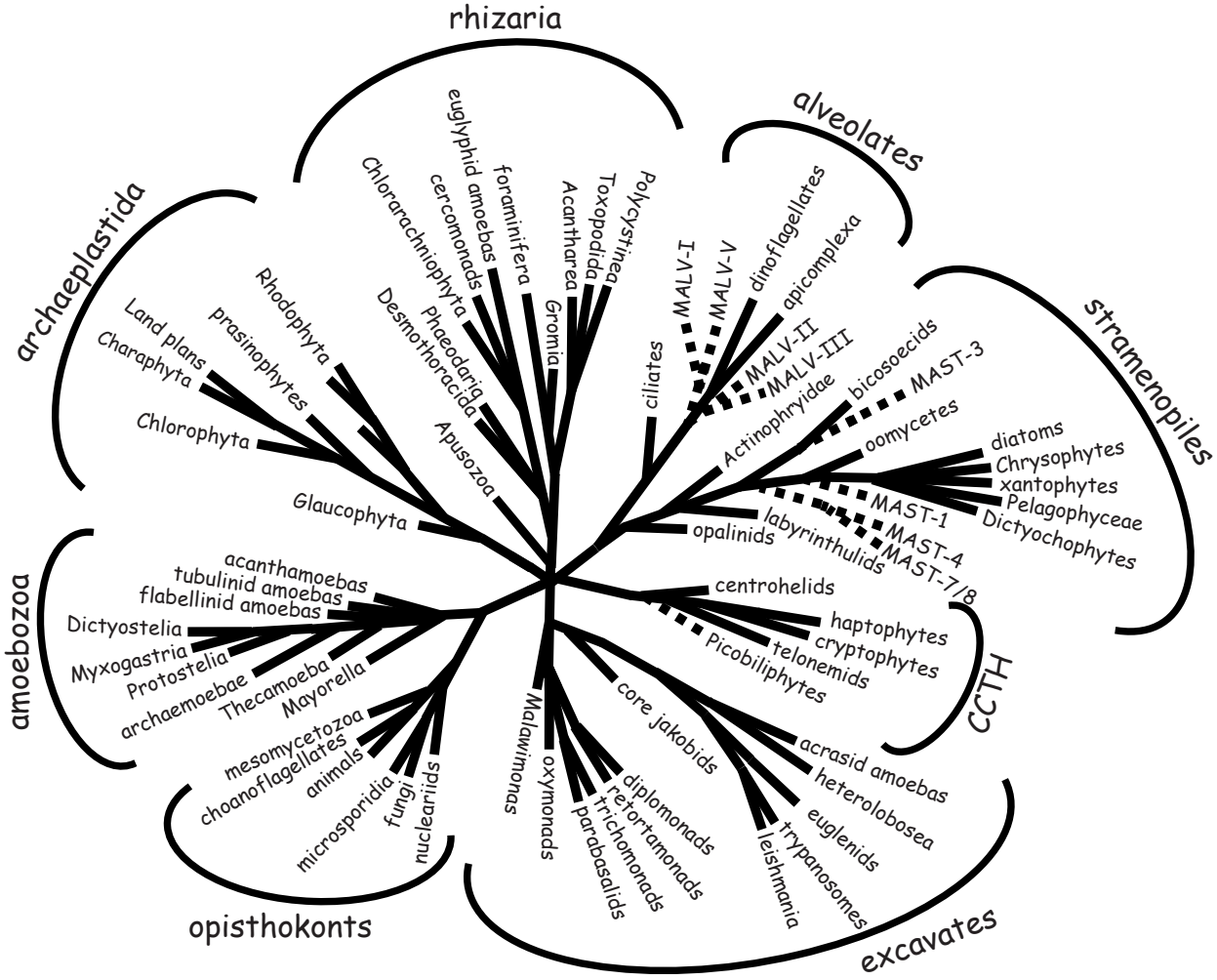


Figura I.1. Árbol de la vida eucariota. Filogenia consenso de los principales grupos eucariotas basada en datos publicados de filogenia molecular y utraestructural. Las líneas punteadas indican las posiciones de los principales linajes, conocidos principalmente mediante técnicas moleculares independientes de cultivo. MALV (alveolados marinos), MAST (estramenópilos marinos) y CCTH (Criptófitos, Centrohelida, Telonemida más Haptófitos). Figura adaptada de (Baldauf 2003).

Un plan corporal muy común en el árbol de la vida eucariota es el de los microorganismos unicelulares no coloreados con uno o pocos flagelos. Este tipo de organización, que por lo general se refiere a los protozoos (o heterótrofos) flagelados, puede ser observado en 27 de los 60 linajes de protistas dentro de los eucariotas (Patterson y Larsen 1991). Por lo tanto, los flagelados son un grado de organización y no un conjunto monofilético consistente. Son organismos que pasan la mayor parte de su existencia moviéndose o alimentándose con un número reducido de flagelos. Su tamaño oscila entre 1-2 μm y 20 μm. El flagelo surgió temprano en la evolución de los eucariotas y no somos capaces de identificar ningún grupo de protistas que primitivamente no tengan flagelo. Se cree que el último ancestro común eucariota era también una especie de flagelado originado por una fusión simbiogenética entre eubacterias y arqueobacterias (Margulis et al 2006). Y obviamente, este eucariota primitivo era incoloro y heterótrofo.

Estudios moleculares microbianos incrementan la diversidad eucariota

El uso de la biología molecular en la ecología microbiana, desarrollado durante el final del siglo XX, ha transformado el campo de la diversidad de los protistas. En general, la identidad de la mayoría de los frágiles y diminutos protistas era muy difícil de evaluar mediante el examen directo de las muestras naturales. Por lo tanto, un método clásico de identificación era, y sigue siendo, obtener estos organismos en cultivo para una clasificación adecuada. En el caso de los protistas autotróficos, las cepas cultivadas son más o menos representativas de las comunidades naturales. Probablemente esto se debe a que es fácil simular las condiciones naturales para ellos en una botella de cultivo, ya que estas células requieren principalmente nutrientes inorgánicos y luz. No obstante, es probable que los cultivos no cubran la diversidad total de los protistas autotróficos *in situ* (Vaulot et al 2008). En el caso de los protistas heterotróficos, la larga lista de especies formalmente descritas (Lee y Patterson 1998) deriva principalmente de cultivos o enriquecimientos que parten del suministro de un sustrato para el crecimiento de las bacterias, que a su vez, son el alimento de los protistas. Las células cultivadas proporcionan información ecofisiológica fundamental, pero al mismo tiempo que es obvio que estas cepas fácilmente enriquecidas viven en el mar, es dudoso que sean los miembros dominantes de los ensamblajes naturales. Un estudio ya clásico demostró que los protistas bacterívoros dominantes en varios enriquecimientos eran raros en las muestras originales (Lim et al 1999). Estudios más recientes han confirmado estos resultados y han proporcionado la explicación mecanicista de este sesgo del cultivo de los flagelados heterotróficos (HF) (del Campo 2011).

Estudios ambientales de secuenciación de los genes del ARNr 18S (independientes de cultivo) se han traducido en una mayor apreciación de la diversidad de los protistas en la naturaleza. De esta manera, los estudios moleculares han revelado numerosas secuencias de protistas desconocidos,

indicando niveles no previstos de la diversidad de los protistas en muchos ambientes y recuperando muy pocas secuencias relacionadas con protistas cultivados (Amaral-Zettler et al 2009, Brown et al 2009, Countway et al 2007, Díez et al 2001, Head et al 1998, Lim 1996, López-García et al 2001, Moon-van der Staay et al 2001, Richards et al 2005, Stoeck et al 2006, Vigil et al 2009). En estudios marinos estos análisis han revelado un gran número de linajes no cultivados, como los alveolados marinos (MALV) y los estramenópilos marinos (MAST) (Massana et al 2004a), los cuales prácticamente aparecen en todos los estudios, aunque todavía la mayor parte de esta diversidad sigue siendo poco conocida. Por lo tanto, está claro que el aislamiento de cepas en cultivo y los estudios moleculares proporcionan diferentes puntos de vista en la composición de especies de protistas marinos en general y de HF en particular. Mientras que estos estudios ambientales fueron a menudo utilizados con el fin ecológico de identificar los miembros dominantes de los ensamblajes naturales, es obvio que también han aportado nuevos conocimientos fundamentales de la filogenia eucariota, así como nuevas ramas en el árbol de la vida formadas exclusivamente por estos linajes no cultivados.

Los estramenópilos, un importante supergrupo en los sistemas marinos

El supergrupo de los Estramenópilos (Adl et al 2005) está formado por muchos linajes heterogéneos, algunos de ellos de vital importancia en los sistemas marinos. Una de las pocas características compartidas por la mayoría de las células móviles de los estramenópilos es la presencia de un flagelo con dos filas opuestas de mastigonemas, pelos tripartitos ("estramenópilos"), los cuales invierten el flujo alrededor del flagelo de manera que la célula se arrastra hacia delante más que se impulsa. La mayoría también tienen un segundo flagelo más corto y liso (de ahí el nombre alternativo "heterokontos"). Este grupo extraordinariamente diverso incluye numerosos linajes unicelulares de heterótrofos (Bicosoécidos) y fotótrofos (Diatomeas), hongos mucilaginosos (Laberintúlidos), parásitos plasmodiales (Oomycetes) y algas multicelulares que pueden alcanzar un gran tamaño (Feófitas).

Existen por lo menos cinco linajes conocidos de estramenópilos no fotosintéticos (figura I.2) (Baldauf 2008). Los Oomycetes (mohos marinos y mohos vellosos) fueron previamente clasificados como hongos e incluyen numerosos parásitos de plantas, extremadamente destructivos, como *Phytophthora infestans*, causante de la plaga de la patata y *Plasmopara viticola*, causante del moho de la vid. Los Bicosoécidos son pequeños biflagelados heterotróficos tales como la bien conocida *Cafeteria* (Fenchel 1988). Los *Blastocystis* spp. son comensales en los intestinos de animales (Stechmann et al 2008) y algunas especies, como el *Blastocystis hominis,* pueden infectar a humanos. Los Laberintúlidos (hongos mucilaginosos) forman redes filamentosas, en forma de raíles, creadas por células ameboides. Se colocaron junto con los Traustoquítridos (Cavalier-Smith et al 1994),

que también tienen la tendencia de formar agregados celulares. Una taxonomía más fina de ambos grupos requiere comparaciones del gen ribosomal del 18S (Honda et al 1999).

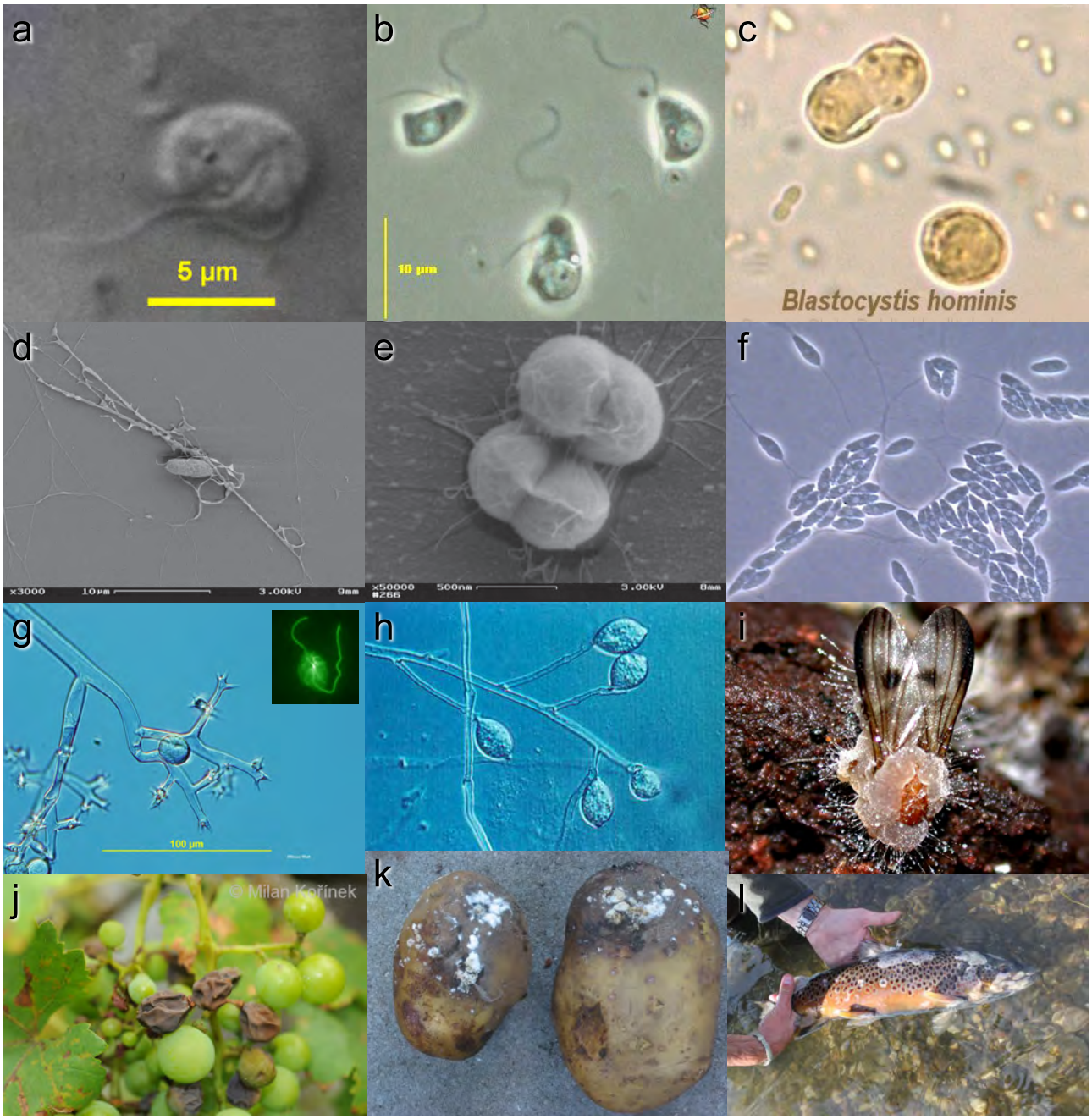


Figura I.2. Fotografías de varios ejemplos de estramenópilos heterotróficos. a) *Developayella elegans*; b) el bicosoecido *Cafeteria roenbergensis*; c) *Blastocystis hominis*; Laberintulidos: d) *Aplanochytrium,* e) *Thraustochytrium* y f) *Labyrinthula terrestris*; Oomycetes: g) *Pasmopara viticola*, h) *Phytophthora infestans*, i) *Saprolegnia*. Las imágenes inferiores representan organismos infectados por los oomycetes superiores, j) mildiú de la vid, k) patatas y l) trucha. Las fotos son cortesía de WJ. Lee, D. Patterson, L.A. Zettler, V. Edgcomb, C. Leander, D. Porter, J. Harper, S. Lew y E. Haugen.

Los estramenópilos fotosintéticos (figura I.3) están formados por al menos once linajes diferentes, donde se incluyen algunas de las algas más importantes y abundantes (Baldauf 2008). Las Diatomeas tienen generalmente dos tecas de sílice con patrones intrincados que encajan entre sí como una caja y su tapa. Son ubicuas y a menudo abundantes en aguas dulces y marinas, con ≈11.000 especies descritas y posiblemente hasta 10^7 especies no descritas (Fehling et al 2007). Las Crisófitas (algas doradas) son generalmente organismos unicelulares de vida libre, pero también forman colonias y filamentos. Las Crisófitas pigmentadas contienen clorofila y un carotenoide llamado fucoxantina que les confiere un color marrón amarillento. Se consideraron mayormente de agua dulce, pero estudios recientes sugieren que también podrían ser bastante abundantes en el plancton marino (Fuller et al 2006, Lepère et al 2009, Shi et al 2011). Las Feófitas (algas pardas) están particularmente extendidas en zonas templadas intermareales y submareales. Tienen verdadero parénquima y forman extensiones de aspecto boscoso en aguas cercanas a la costa, como los llamados bosques de kelp, algas marinas gigantes que contienen ecosistemas complejos incluyendo peces y mamíferos marinos. Las Xantofíceas (algas verde-amarillentas) son los principales productores en algunas marismas de agua salobre y forman también organismos multicelulares. Los grupos restantes están formados por algas muy pequeñas, como las Dicteocófitas, Eustigmatófitas, Faeotamniófitas, Pelagofíceas y Pinguiófitas (Vaulot et al 2008). Las Pelagofíceas son una clase descrita hace poco tiempo (Andersen et al 1993), clasificadas previamente dentro de las Crisofíceas, y podrían ser importantes en el picoplancton oceánico.

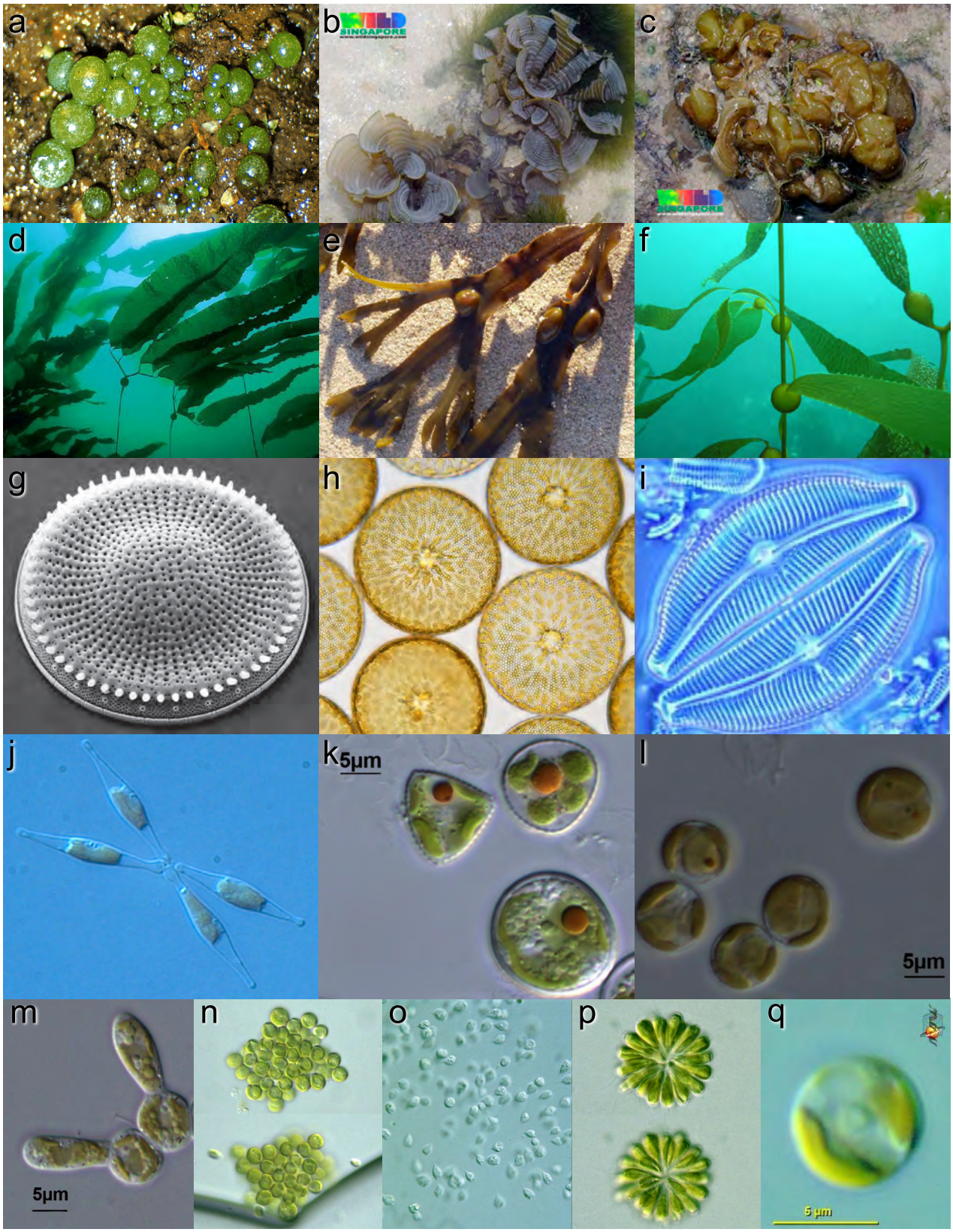


Figura I.3. Fotografías de varios ejemplos de estramenópilos fotosintéticos. a) El xantófito *Botrydium*; Feofitos: b) *Padina,* c) *Colpomenia*, d) *Pelagophycus porra*, e) *Fucus vesiculosus* y f) *Macrocystis integrifolia*; Diatomeas: g) *Stephanodiscus,* h) *Coscinodiscus*, i) *Cymbella tumida* y j) *Phaeodactylum tricornutum*; k) el eustigmatofito cepa 29.96; Phaeothamniophytes: l) *Stichogloea doederleinii* y m) *Phaeothamnion confervicola*; Crisofitas: n) *Chrysocapsa epiphytica,* o) *Spumella* sp. y p) *Synura* y q) la pinguiofita *Pinguiococcus pyrenoidosus*. Las fotos son cortesía de I. Inouye, R. Tan, E. Bierman, D. Mann, A. de Martino, C. Bowler, J.C. Baley, Y. Tsukii, D. Patterson, B. Andersen y U.S. Geological Survey.

MAST, linajes de estramenópilos marinos no cultivados

Los estramenópilos marinos (MAST) fueron detectados por primera vez a partir de secuencias del ADNr 18S obtenidas de ambientes marinos y sin una ubicación filogenética clara. Forman por lo menos 10 clados en la parte basal de los estramenópilos (Massana et al 2004b), en donde todos los protistas son heterótrofos, e incluyen flagelados fagotróficos de vida libre (bicosoécidos), parásitos (blastocistos) u osmótrofos (oomycetes y laberintúlidos) (figura I.4). Los MAST son muy recurrentes en los estudios moleculares, encontrándose en los cinco océanos mundiales, con la mayoría de las secuencias afiliadas a unos pocos clados (MAST-1, MAST-3, MAST-4 y MAST-7). La naturaleza heterotrófica de los MAST, sospechada inicialmente por su ubicación filogenética, fue confirmada posteriormente mediante FISH (hibridación fluorescente *in situ*) para el clado-1, clado-2 y clado-4 (Massana et al 2006b) y para el clado-6 (Piwosz y Pernthaler 2010). Las células MAST que forman estos grupos son pequeños protistas (2-8 µm de tamaño), capaces de crecer en la oscuridad y de ingerir bacterias. Además, son bastante abundantes en el plancton marino y representan una fracción significativa de los HF a nivel mundial (hasta un 35%).

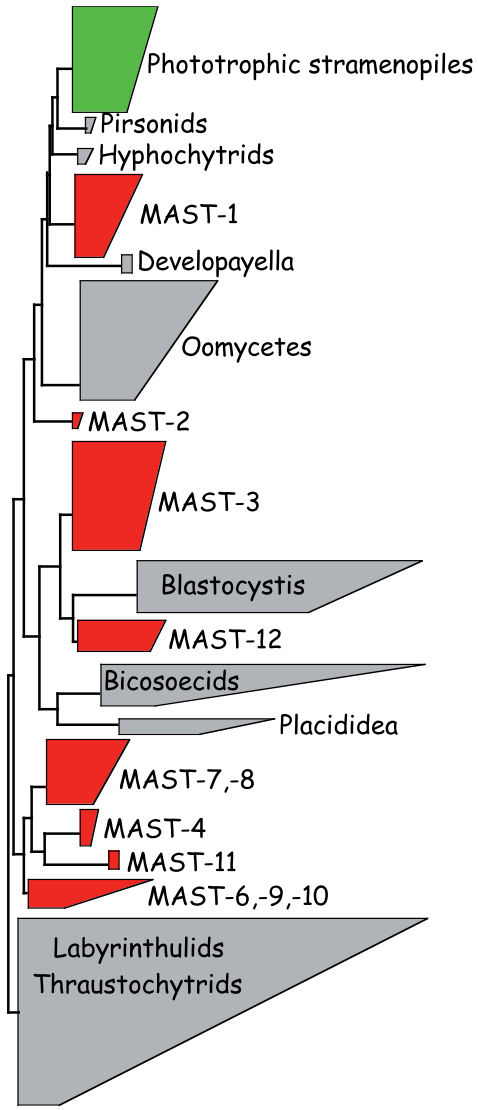


Figura I.4. Posición filogenética de los estramenópilos marinos (MAST) dentro del supergrupo de los estramenópilos. Árbol con secuencias completas del ADNr 18S [modificado de (Massana et al 2004b)] mostrando las posiciones del los linajes de MAST (rojo) entre los fotótrofos cultivados (verde) y los heterótrofos (gris).

Un grupo en particular, el MAST-4, se ha encontrado en todas las muestras analizadas (excepto las polares) (figura I.5 e I.6). Es un protista muy pequeño (2-3 µm de tamaño), por lo tanto clasificado como picoeucariota. Su promedio de abundancia es de 130 cél ml^{-1} y representa el 9% de los protistas heterotróficos en un amplio rango de sistemas marinos (Massana et al 2006b). Este grupo muestra un tamaño corporal consistente en todas las muestras analizadas para un extenso registro de temperaturas (de 5 a 28ºC), incumpliendo la regla de "temperatura-tamaño" según la cual el tamaño corporal disminuye en función del incremento de la temperatura (Atkinson et al 2003). Se definió como HF debido a su rápido crecimiento en la oscuridad, la ausencia de cloroplastos, la observación de vacuolas alimentarias que contenían bacterias y la presencia de un flagelo (Massana et al 2006a) (figura I.7). Debido a su amplia distribución y su abundancia global es probable que las células MAST-4 contribuyan sustancialmente a la cadena alimentaria marina en extensas áreas oceánicas. Hasta el momento se han realizado numerosos esfuerzos, aunque sin éxito, para obtener un representante en cultivo. Es destacable que grupos dominantes en los océanos todavía no estén cultivados y esto remarca la importancia ecológica de la nueva diversidad detectada por métodos moleculares.

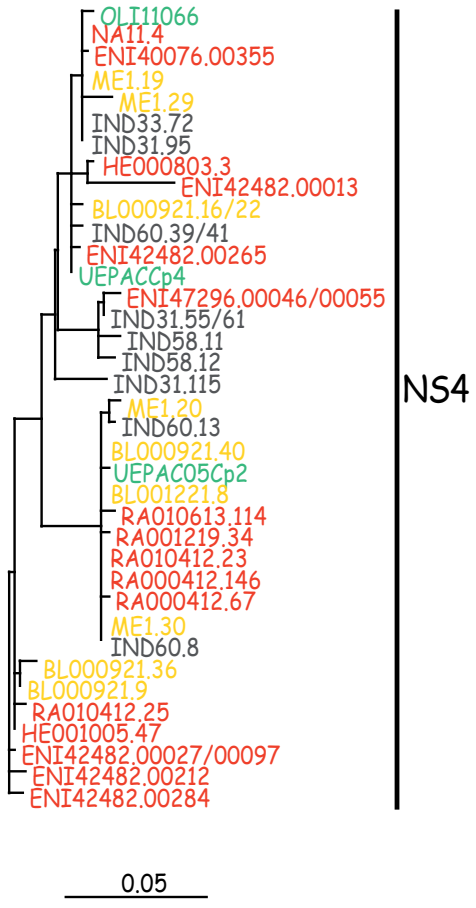


Figura I.5. Árbol filogenético formado por secuencias parciales del ADNr 18S del MAST-4. Cada color identifica secuencias de diferentes regiones (Atlántico: rojo; Pacífico: verde; Índico: gris; Mediterráneo: amarillo). La línea vertical negra muestra la cobertura de la sonda NS4 para FISH. La barra de escala indica 0.05 sustituciones por posición. Figura tomada de (Massana et al 2006b).

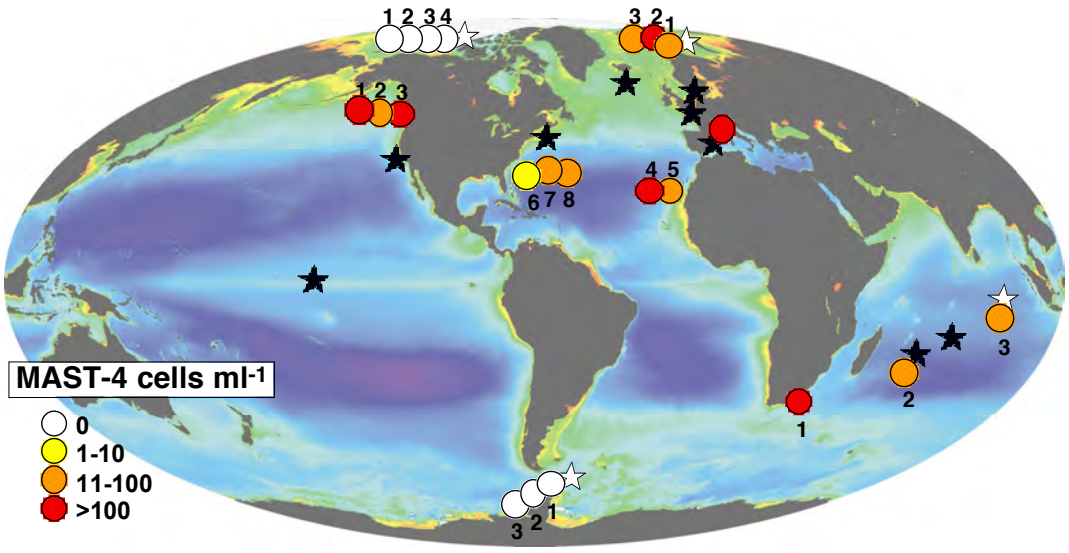


Figura I.6. Distribución global y abundancia de las células del MAST-4 en los océanos. Las estrellas indican lugares donde se han realizado bibliotecas de clones del ADNr 18S, en negro si la biblioteca contiene secuencias del MAST y en blanco si no. Los puntos indican lugares donde se han realizado recuentos de FISH, con diferente color dependiendo de la abundancia de células encontrada. Figura tomada de (Massana et al 2006b).

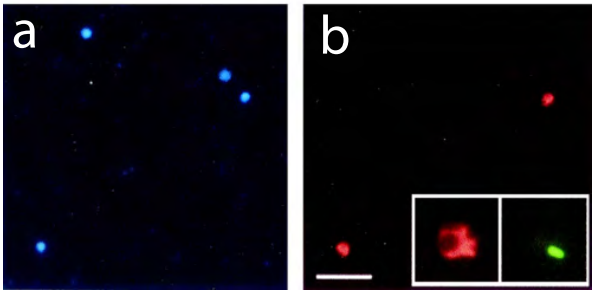


Figura I.7. Micrografías de epifluorescencia de las células del MAST-4. (a) células teñidas con DAPI y su correspondiente visión microscópica bajo luz ultravioleta, (b) células positivas utilizando la técnica de FISH observadas con luz azul. Al comparar a y b se muestra que dos de las cuatro células eucariotas eran MAST-4. La barra de escala es de 10 µm. Se observa que la región nuclear, la más brillante por el DAPI (a), es más tenue para la fluorescencia de FISH (b), coincidiendo con la localización de los ribosomas. El inserto en el panel b muestra una célula del MAST-4 (amplificada 3 veces) con una FLB ingerida. Fotos tomadas de (Massana et al 2002).

Resumen de las características del MAST-4:

- Eucariota unicelular (protista)
- Estramenópilo
- Picoeucariota: tamaño de la célula inferior a 3 µm
- Flagelado heterotrófico marino
- Depredador de bacterias
- Mundialmente distribuido excepto en los sistemas polares
- Abundante en los ecosistemas marinos (~10% de los flagelados heterotróficos)
- Todavía no cultivados

La importancia de los picoeucariotas marinos

La existencia de vida microbiana en suspensión en la columna de agua marina se conoce desde hace mucho tiempo, pero sólo en las últimas décadas ha surgido una apreciación de la importancia de su significado ecológico y biológico. La mayor parte del conocimiento de los protistas marinos se ha limitado a los grandes taxones microscópicos reconocibles, como las microalgas, los ciliados y los flagelados más grandes. Por el contrario, los picoeucariotas marinos son en gran parte indistinguibles al microscopio óptico. Forman un conjunto de pequeñas células inconspicuas, sólo un poco más grandes que las bacterias marinas (Massana 2011). Los picoeucariotas fototróficos (células pigmentadas) son importantes productores primarios que están en la base de la cadena trófica. Los picoeucariotas heterotróficos (células incoloras) son en su mayoría bacterívoros y desempeñan un papel clave en la canalización de las bacterias hacia niveles tróficos superiores, así como en el reciclaje de nutrientes. La mixotrofía y el parasitismo también son relevantes pero sus relaciones tróficas no han sido tan estudiadas. Hasta ahora, sólo unos pocos picoeucariotas han sido aislados y caracterizados, por lo que permanecen en gran parte sin describir.

El picoplancton eucariota es una colección heterogénea de pequeños protistas con un diámetro comprendido entre 0.8 μm en el caso de *Ostreococcus tauri*, el eucariota más pequeño conocido, hasta un rango superior de 2-3 μm. En 1978 se definió un sistema para la clasificación de los organismos marinos en función de su tamaño, basado principalmente en el proceso de tamizado. De esta forma, los microorganismos fueron divididos en tres categorías: picoplancton (0.2-2 μm de diámetro celular), nanoplancton (2-20 μm) y microplancton (20-200 μm). Inicialmente el picoplancton se pensó para estar casi exclusivamente formado por procariotas y el nanoplancton en su mayoría formado por pequeños eucariotas unicelulares. Sin embargo, pronto se reconoció la existencia y abundancia de los protistas dentro del tamaño del picoplancton. Hoy en día, el término picoeucariota es a menudo utilizado menos rigurosamente para incluir a los protistas con un tamaño de hasta 3 μm (Vaulot et al 2008). Observaciones directas del conjunto de los protistas indican que el límite previamente establecido de 2 μm se sitúa frecuentemente en el medio del espectro de tamaño y que se delimita un grupo más coherente si utilizamos un límite de hasta 3 μm (Massana 2011). Muchos picoeucariotas, tanto fototróficos (como *Micromonas pusilla*) como heterotróficos (como el MAST-4), son células flageladas (Patterson y Larsen 1991). Por lo tanto una gran proporción de los ensamblajes, conocidos como flagelados fototróficos y heterotróficos, se clasificarían dentro de los picoeucariotas. De esta manera, los picoeucariotas son muy importantes tanto como productores primarios como depredadores de bacterias en los sistemas acuáticos. Hoy en día sabemos que los picoeucariotas son ubicuos en los ambientes marinos poblando la superficie oceánica con una abundancia de alrededor de 1000 cél ml^{-1}. Los picoeucariotas son, sin lugar a dudas, miembros esenciales de los ecosistemas marinos en términos de abundancia celular, biomasa, actividad y

diversidad y juegan un papel decisivo en la cadena trófica y los ciclos biogeoquímicos (Massana 2011).

Los picoeucariotas tienen la típica estructura de célula eucariota en un estado miniaturizado. Debido a su pequeño tamaño, son atractivos modelos para proyectos de secuenciación genómica, en particular, para la búsqueda de las bases genéticas de la miniaturización de las células y su éxito ecológico. Su reducido tamaño tiene implicaciones en las tasas metabólicas y en la transferencia trófica a través de las redes tróficas microbianas. Aunque existen algunos ejemplos de picoeucariotas cultivados, la mayoría, como el grupo de los MAST-4, se han detectado en los últimos años únicamente por técnicas independientes de cultivo y por lo tanto están escasamente caracterizados. Estudios moleculares de los picoeucariotas han dado a conocer una gran diversidad filogenética, incluyendo algunos linajes nuevos y son fundamentales para entender el significado ecológico y evolutivo de esta amplia y nueva diversidad. Un objetivo importante es evaluar cómo los individuos se organizan en unidades taxonómicas y cómo participan en los procesos ecológicos.

En resumen, los picoeucariotas, definidos como protistas menores de 3 μm, son abundantes y ecológicamente importantes en los ecosistemas marinos planctónicos. Incluyen diversas células fototróficas y heterotróficas y juegan papeles esenciales como productores primarios, consumidores de bacterias y parásitos. En los últimos años, su diversidad filogenética y funcional, así como su abundancia y amplia distribución han comenzado a ser reconocidas y están atrayendo más la atención. Esta tesis aborda el estudio de un linaje específico no cultivado, el MAST-4, que podría considerarse como un flagelado heterotrófico modelo.

Enfoques para la autoecología de linajes microbianos específicos

El estudio de la ecología de una sola especie (o especies afines) es una disciplina normalmente llamada autoecología. Esto incluye el estudio de la distribución, la función y las interacciones de esta especie con el medio ambiente y es un enfoque aplicado normalmente en animales y plantas. Un ejemplo es el estudio de la diversidad y abundancia de la comunidad en las poblaciones de aves (Poirazidis et al 2011). Un estudio típico puede comenzar por establecer diferentes puntos de observación y puede realizarse con unos prismáticos y un buen conocimiento de la morfología de las aves (figura I.8). Si se requiere un registro más preciso de sus movimientos y actividad es posible incluso capturar y marcar determinados individuos con una anilla o etiqueta en el ala para distinguirlos en posteriores observaciones. También se les puede colocar una mochila radiotranmisora para seguir sus movimientos (Schindler et al 2006), como se ha hecho con el buitre negro *(Aegypius monachus*) para estudiar la temporada de cría (Vasilakis et al 2005) y su distribución espacial (Vasilakis et al 2008) en el Parque Nacional de Dadia, Grecia.

Figura I.8. Seguimiento del buitre negro *(Aegypius monachus*) en el Parque Nacional de Dadia, Grecia. Buitres en un lugar de alimentación (arriba-izquierda). Observación de aves por ornitólogos (arriba-derecha). Antena para buscar la señal del transmisor de telemetría (abajo-izquierda). Buitres con anilla y etiqueta en el ala (Fotos cedidas por J. Elorriaga) (abajo-derecha).

Estos métodos de estudio resultan mucho más complicados, incluso imposibles, en el mundo microbiano. La limitación principal es que las características observadas a través del microscopio óptico, para los microorganismos de interés, no son suficientes para su identificación. Esto sucede tanto en los procariotas, como en los picoeucariotas. Si usamos el microscopio de epifluorescencia con una tinción de DAPI (4,6-diamidino-2-phenylindole) podemos obtener datos cuantitativos de las células fototróficas (pigmentadas) o heterotróficas (incoloras), pero no los podemos clasificar. Dentro del conjunto de las células heterotróficas (figura I.9) solo se pueden distinguir claramente dos grupos: los coanoflagelados, por su evidente collar con un único flagelo y los dinoflagelados, por el aspecto granulado del núcleo. Una herramienta que facilita el estudio es su obtención en

cultivos puros, pero hoy en día sabemos que un gran número de microorganismos dominantes en las comunidades naturales todavía no están cultivados. Su investigación es muy importante para tener un mejor conocimiento de la naturaleza (son los más abundantes) pero esto resulta difícil debido a la ausencia de cultivos. Particularmente, el MAST-4 (objeto de estudio de esta tesis) es uno de estos grupos no cultivados. Por lo tanto, es necesario estudiarlos directamente de muestras ambientales, independientemente de cultivos y microscopía directa. A continuación exponemos tres enfoques diferentes para el estudio ambiental de linajes microbianos específicos.

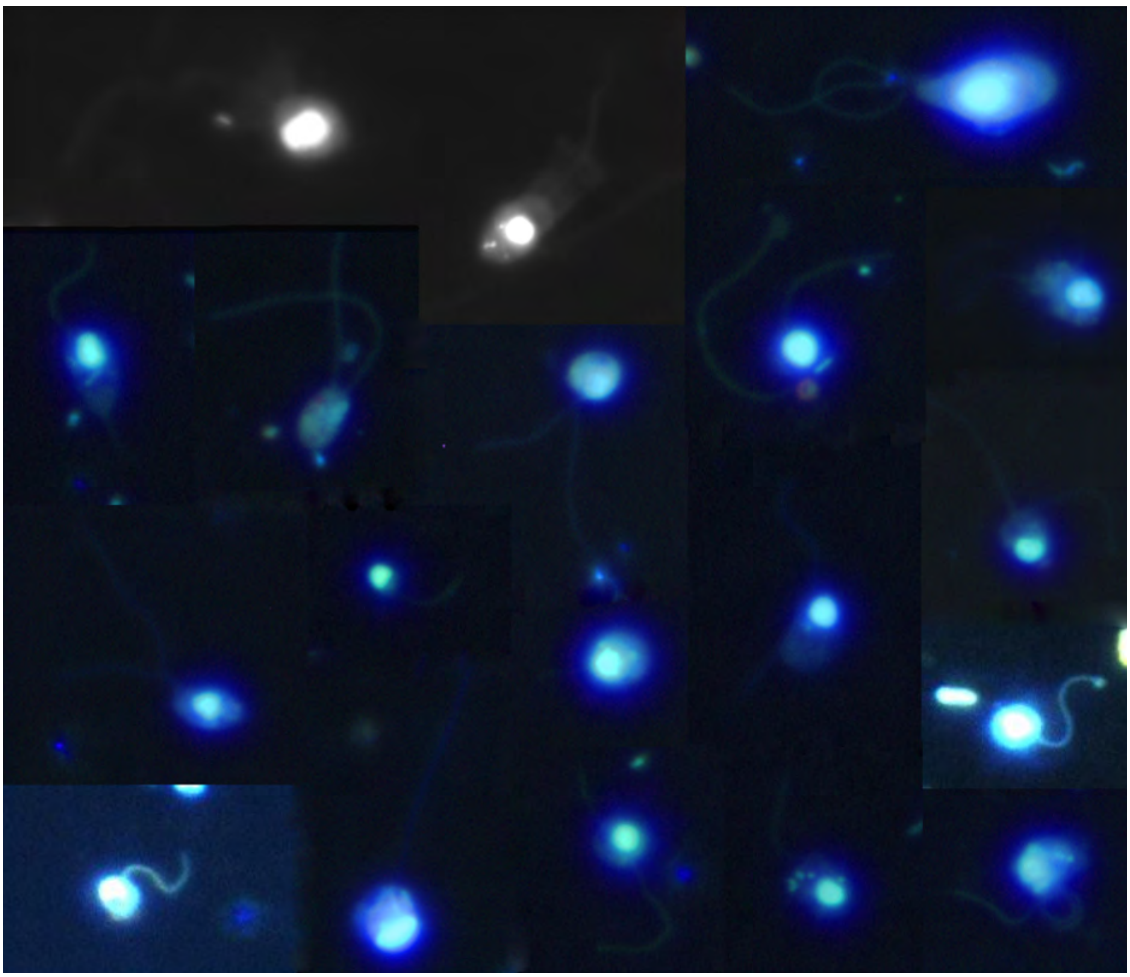


Figura I.9. Micrografías de epifluorescencia de flagelados heterotróficos teñidos con DAPI y observados bajo luz UV. Fotos cedidas por R. Massana e I. Forn.

Selección específica de genes de la subunidad pequeña del ribosoma

Durante los últimos años ha sido desarrollado un método sistemático de secuenciación de genes conservados de las subunidades ribosomales, abriendo de este modo diferentes rutas en la caracterización molecular de comunidades naturales microbianas (Amann et al 1995). Una dinámica general (figura I.10) consiste en la extracción de ácidos nucleicos, ya sea directa o tras un enriquecimiento específico de determinados organismos a partir de la muestra natural. Con la llegada de la reacción en cadena de la polimerasa (PCR) (White et al 1989), fragmentos del gen

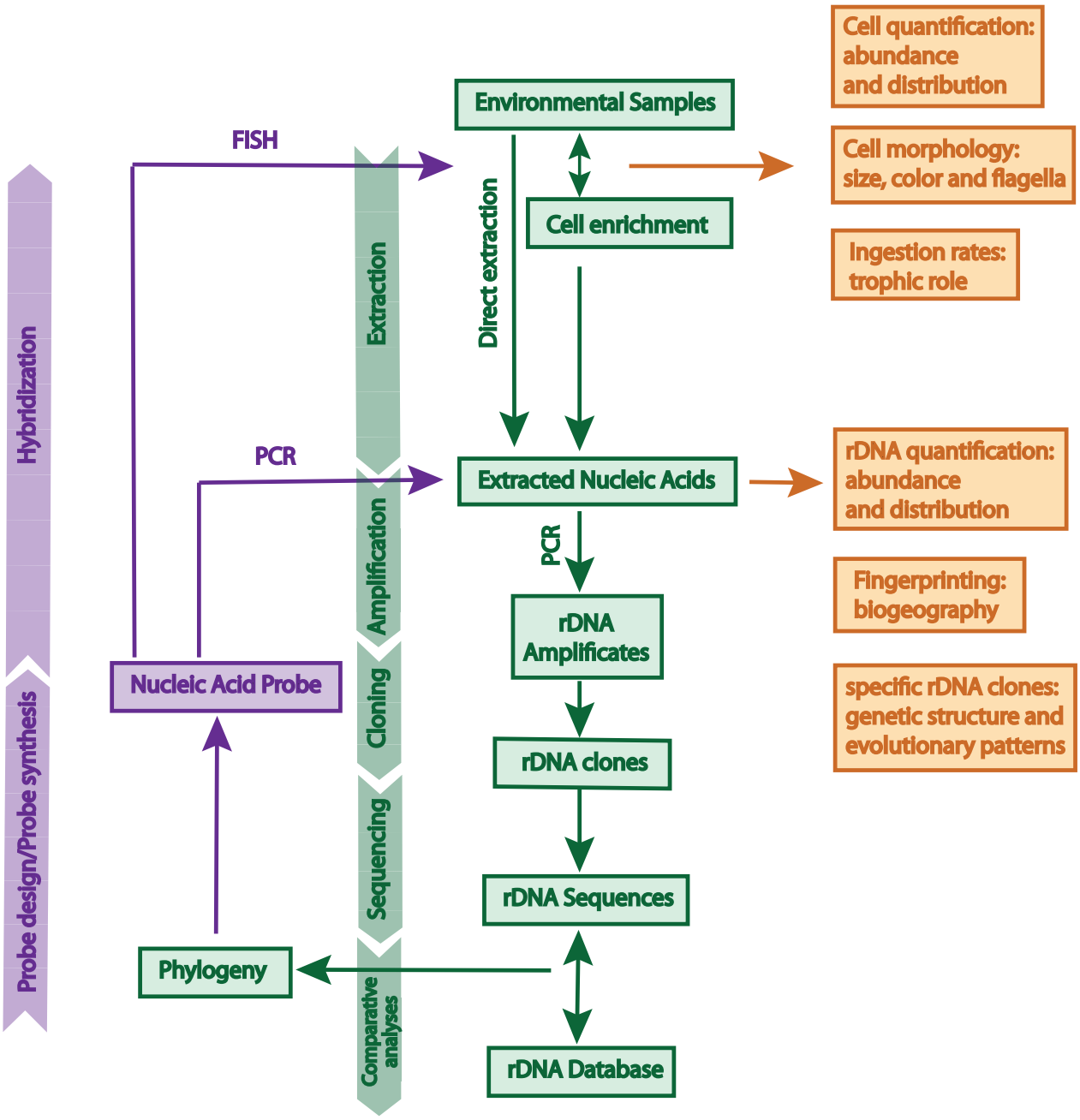


Figura I.10. Diagrama de flujo que muestra el camino usado en esta tesis para caracterizar las muestras ambientales analizando comparativamente distintas secuencias de ADNr (verde), las diferentes opciones usadas para marcar ADNr de linajes específicos (morado) y los resultados obtenidos (naranja). Figura adaptada de (Amann et al 1995).

del ARNr pudieron ser amplificados selectivamente y clonados a partir de la mezcla de ADNs ambientales. Una vez creadas las bibliotecas de clones, con los fragmentos de ADN definidos, se pueden entonces secuenciar. Estas secuencias son almacenadas en bases de datos públicas como el Genbank en el Centro Nacional de Información Biotecnológica (NCBI) (http://www.ncbi.nlm.nih.gov/genbank/), o en lugares específicos creados para secuencias de ADNr, como la base de datos Silva del software ARB (http://www.arb-silva.de/). Las bases de datos públicas, de cómodo acceso, facilitan el análisis comparativo con otras secuencias ambientales o de referencia. Esto nos permite ubicar y visualizar las secuencias ambientales en árboles filogenéticos y diseñar sondas específicas o conjuntos de iniciadores para un grupo concreto seleccionado. De esta manera se pueden detectar taxones microbianos específicos con un microscopio de epifluorescencia mediante la técnica de FISH, o amplificando su ADNr con el paso de la PCR requerido para muchas técnicas moleculares. El punto crítico consiste en conseguir sondas o iniciadores específicos de los microorganismos diana y en evitar los falsos positivos de las muestras ambientales. Por otra parte, a la hora de evaluar las comunidades naturales es importante considerar el número de posibles sesgos introducidos en cada paso realizado para la obtención de las secuencias de ADNr. Por ejemplo, el diferente número de copias del ADNr o la variabilidad en la eficiencia de amplificación de grupos en particular pueden sesgar severamente el producto de la PCR. Además, la PCR puede dar lugar a secuencias erróneas, no existentes en la muestra natural, debido a problemas de contaminación o a la formación de quimeras.

Para el estudio de los linajes específicos, la técnica de FISH (Pernthaler et al 2001) nos permite detectar los taxones microbianos diana, medirlos, cuantificarlos en comunidades naturales e incluso observar la ingestión de diferentes presas. Con la PCR podemos secuenciar sus genes para evaluar su diversidad y estructura genética, o comparar la diversidad entre taxones de diferentes muestras a través de las huellas moleculares. Cuando se utiliza la PCR cuantitativa, la ausencia de cultivos es un reto para la preparación del estándar de cuantificación. Típicamente, cuando el cultivo está disponible, el estándar permite relacionar el ADN cuantificado con el número de células. En esta tesis, dicho problema se resolvió utilizando plásmidos clonados con el inserto del ADNr 18S del taxón seleccionado. De esta manera se pudieron crear estándares con diluciones seriadas de las moléculas del ADNr 18S (figura I.11). Los valores obtenidos no pueden traducirse directamente en número de células, debido a la variabilidad que presentan los eucariotas en el número de copias del operón del ADNr, pero igualmente son de gran utilidad para la cuantificación.

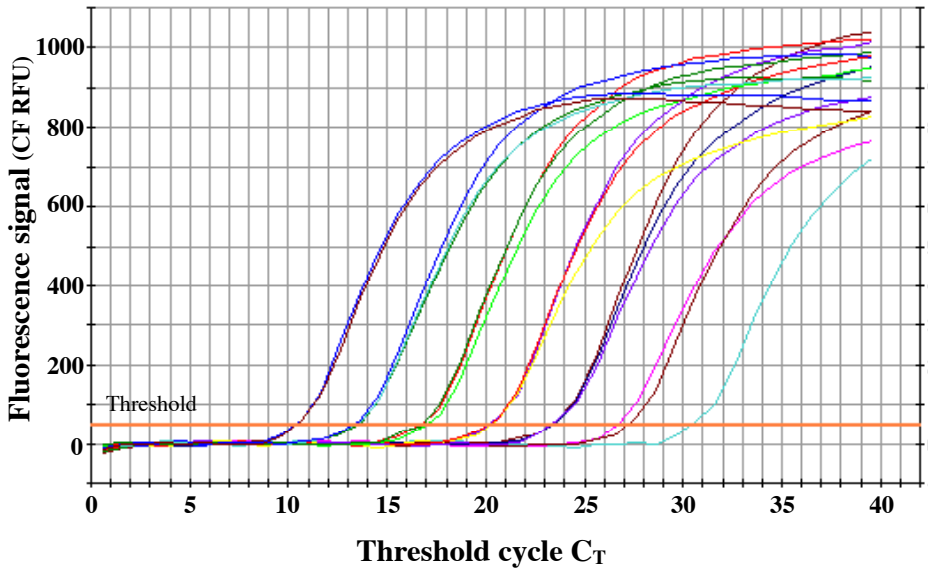


Figura I.11. Diluciones seriadas del plásmido (de 10^8 a 10^2 copias) usado como estándar de la PCR cuantitativa. El buen rendimiento del estándar se muestra por el alto coeficiente de correlación (0.999) y una eficiencia de la PCR cercana al 100% (99.2%).

Incubaciones de agua de mar no modificadas

El trabajo con cultivos puros ofrece numerosas ventajas, tales como la estimación de parámetros morfológicos y funcionales, la observación de la ultraestructura mediante el microscopio electrónico, la determinación de las respuestas numéricas y funcionales, del rango del tamaño de la comida, de la eficiencia en el crecimiento, de la temperatura óptima y de las respuestas de supervivencia. En ausencia de cultivos puros para la mayoría de los flagelados heterotróficos, un método que permite la obtención de alguno de estos parámetros es el uso de incubaciones de agua de mar no modificadas (Massana et al 2006a) para promover el crecimiento de ensamblajes mixtos de HF no cultivados. El proceso incluye una prefiltración a través de un filtro de 3 µm que elimina los grandes predadores seguida de una incubación en oscuridad para evitar el crecimiento de organismos fototróficos. Al inicio de la incubación, el número de bacterias aumentan alcanzando varias veces el número inicial de la población (figura I.12), a esto le sigue unos días más tarde un crecimiento en el número HF, que a su vez coincide con la disminución del número de bacterias. En el total de la muestra incubada los flagelados fotosintéticos y *Synechococcus* disminuyen continuamente en número debido a las condiciones de oscuridad de la incubación y quizás también por la depredación de los flagelados heterotróficos. De esta manera el conjunto de protistas cambia de un dominio de células fototróficas a dominio de células heterotróficas. En la mayoría de las muestras, este sencillo proceso da como resultado el enriquecimiento de varios grupos de MAST (Massana et al 2006a). El crecimiento de estos abundantes HF no cultivados se debe, probablemente, a que en las incubaciones no modificadas las bacterias se mantienen en la misma abundancia y tamaño que en las comunidades naturales (del Campo 2011). Este incremento de flagelados es moderado (10-100 veces), suficiente para medir sus tasas de crecimiento y proporcionar un material importante para las mediciones de la actividad. Aunque de corta duración, estos eventos ofrecen información

fenotípica y funcional interesante sobre los protistas no cultivados. Por lo tanto, las incubaciones de agua de mar no modificadas sirven para seleccionar los HF abundantes *in situ*, pero no para aislarlos en cultivo. Este sencillo enfoque se utilizó en varios capítulos de esta tesis.

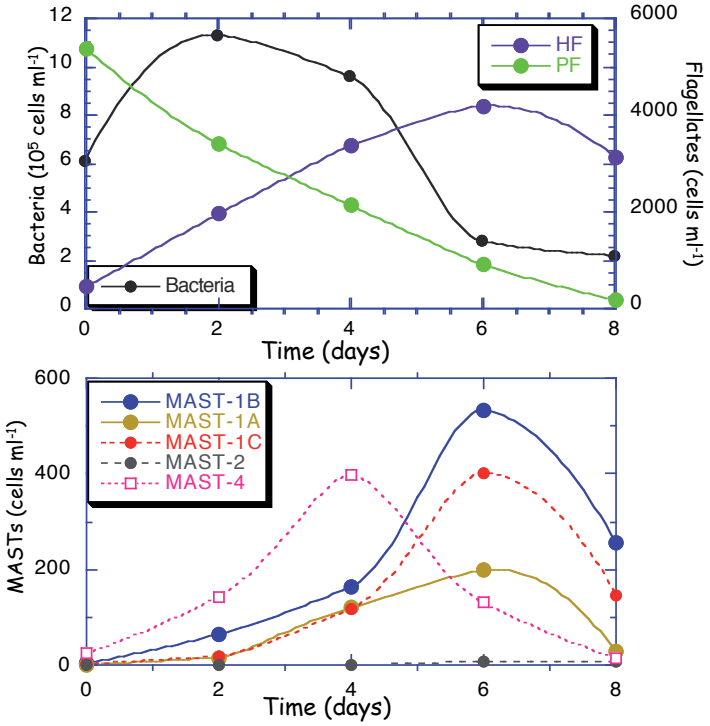


Figura I.12. Dinámica de los componentes microbianos en una incubación marina no modificada realizada con agua del Mar de Noruega. Panel superior: abundancia celular estimada por recuentos de DAPI para bacterias (negro), heterótrofos (morado) y flagelados fototróficos (verde). Panel inferior: abundancia celular estimada por FISH para cinco grupos de MAST. Figura tomada de (Massana et al 2006a).

Genomas individuales amplificados (SAGs)

Últimamente se han producido avances interesantes en el análisis de células aisladas. La actual capacidad de los citómetros de flujo para separar células individualmente, combinada con el uso de lysotraker, una sonda fluorescente que tiñe las vacuolas alimentarias (Rose et al 2004), está abriendo nuevos caminos en la ecología microbiana. Células microbianas individuales pueden entonces ser usadas como inóculos para el desarrollo de cultivos puros o como plantilla para la amplificación genómica total previa a la secuenciación (Yoon et al 2011). Este enfoque ha sido aplicado recientemente para la selección de HF a partir de comunidades marinas (Heywood et al 2011, Rose et al 2004). La evaluación de la diversidad de los protistas heterotróficos basada en la selección de células individuales, la amplificación total del genoma y la secuenciación de ADNr es mejor que la obtenida a través del estudio de las comunidades, ya que el número de copias del ADNr no es un problema. Además, los SAGs son la única forma de acceder a la información genómica de células no cultivadas y permiten analizar interacciones ecológicas (depredación, simbiosis) entre protistas y procariotas (Martínez-García et al 2012). Preparar y analizar SAGs del MAST-4 podría ser un enfoque prometedor para complementar el estudio presentado en esta tesis.

Objetivos y esquema de la tesis

El objetivo general de esta tesis fue el estudio de la ecología de un taxón relevante, el MAST-4, perteneciente a los flagelados heterotróficos no cultivados. Este protista es un picoeucariota importante, ampliamente distribuido en sistemas marinos, que representa una apreciable fracción del total de los flagelados heterotróficos. Además, tiene la ventaja de ser fácilmente enriquecido en incubaciones no modificadas y fácilmente detectado con herramientas moleculares.

Los flagelados heterotróficos (HF) se cuantifican habitualmente por microscopía de epifluorescencia, después de una tinción con DAPI (Porter y Feig 1980), pero esto muestra pocas características morfológicas, por lo que en general continúan sin identificar. Con la aparición de los estudios moleculares se diseñaron sondas de oligonucleótidos para la detección de varios linajes de los MAST (Massana et al 2002, Massana et al 2006b) y se han usado para su identificación por la técnica de FISH. Estudios adicionales sobre su distribución y abundancia revelaron que estaban distribuidos globalmente y que un solo grupo, el MAST-4, contribuía aproximadamente con el 9% de los HF de la superficie de los sistemas marinos (Massana et al 2006b). Hasta el momento, el MAST-4 ha sido cuantificado por FISH. Esta es una técnica muy fiable pero requiere mucho tiempo para procesar la gran cantidad de muestras generadas durante las campañas oceanográficas. El primer objetivo de esta tesis **(capítulo 1)** fue desarrollar una técnica rápida y sensible para evaluar la abundancia y distribución del flagelado heterotrófico no cultivado MAST-4, basada en la detección de los genes del ARNr 18S mediante la cuantificación de la reacción en cadena de la polimerasa a tiempo real (Q-PCR).

La depredación bacteriana es muy importante en los ecosistemas acuáticos y la llevan a cabo principalmente pequeños protistas flagelados de hasta 5 μm de diámetro (Sherr y Sherr 2002). Controla la abundancia de las bacterias en un amplio rango de condiciones de los ecosistemas, canaliza el carbono orgánico a niveles tróficos superiores y libera nutrientes inorgánicos que a menudo están controlando la producción primaria (Jürgens y Massana 2008, Pernthaler 2005). El segundo objetivo de esta tesis fue el estudio de las tasas de depredación y las preferencias de presa **(capítulo 2)** junto con la respuesta funcional **(capítulo 3)** de HF no cultivados propios de las comunidades naturales, incluyendo el MAST-4. Esta parte se basa en la estimación de la actividad de alimentación de predadores específicos detectados por FISH después de experimentos de ingestión a corto plazo, con presas en concentraciones traza, las cuales se hizo un posterior recuento en el interior de las vacuolas alimenticias de los protistas.

Los microbios tiene un papel vital para el funcionamiento de la biosfera (Falkowski et al 2008), aunque todavía nos encontramos alejados de tener estimaciones aceptables de su diversidad. Por otra parte, no está claro cómo la diversidad microbiana se distribuye en el espacio y el tiempo

y cómo las categorías de la diversidad se traducen en el significado ecológico de las interacciones y procesos. Los protistas marinos de pequeño tamaño, los picoeucariotas, son algunos de los microbios menos explorados con gran importancia ecológica (Massana 2011). El tercer objetivo de esta tesis **(Capítulo 4)** fue comprender la estructura genética y los patrones de evolución del picoeucariota MAST-4. Se basa en la secuenciación de un fragmento largo del operón del ADNr y en la investigación de las estructuras secundarias del ITS (espaciador interno transcrito) para explorar los posibles límites de entrecruzamiento entre tipos relacionados.

La biogeografía es el estudio de la distribución en el espacio y el tiempo. La evidencia actual confirma que la selección ambiental es fundamental para la variación espacial en la diversidad microbiana (Martiny et al 2006). La siguiente frontera es averiguar si estos patrones están también influenciados por barreras geográficas que facilitan la evolución y la diversificación. En los últimos años se obtuvieron resultados contradictorios, sin una imagen universal emergente, en parte porque la respuesta puede depender de cada situación analizada en particular. El último objetivo de esta tesis **(capítulo 5)** fue estudiar la biogeografía de los protistas marinos utilizando el MAST-4 como modelo. La estructura de la comunidad y su distribución fue determinada mediante la combinación del análisis automatizado del espaciador interno ribosómico (ARISA) (Fisher y Triplett 1999) y las bibliotecas de genes 18S-ITS1.

El esquema de los diferentes temas estudiados es el siguiente:

• *Objetivo 1: Abundancia y distribución*

Capítulo 1 "Distribución del protista no cultivado MAST-4 en el Océano Índico, el Pasaje de Drake y el Mar Mediterráneo, calculada mediante la PCR cuantitativa en tiempo real"

Desarrollamos un protocolo de Q-PCR para determinar rápidamente la abundancia de este grupo utilizando ADN medioambiental. Se diseñó un conjunto de iniciadores, que seleccionaban los genes del ARNr 18S del MAST-4, se optimizó el protocolo de Q-PCR y se calibró utilizando un plásmido con la secuencia diana como inserto. La Q-PCR fue utilizada para cuantificar el MAST-4 a lo largo de tres transectos, uno longitudinal en el Océano Índico, otro latitudinal en el Pasaje de Drake y el último de costa a mar abierto en el Mar Mediterráneo. También se realizó un estudio temporal en una estación costera del Mar Mediterráneo.

• *Objetivo 2: Papel trófico*

Capítulo 2 "Tasas de depredación y diversidad funcional en flagelados heterotróficos no cultivados"

En este capítulo medimos las tasas de depredación de los protistas no cultivados procedentes de comunidades naturales (detectados por FISH) e investigamos su preferencia por diferentes presas bacterianas en experimentos de ingestión a corto plazo. Esto incluía bacterias marcadas

fluorescentemente (FLB) y dos cepas de las familias Rhodobacteraceae *y* Flavobacteriaceae. Estas dos últimas, con varios tamaños celulares, eran ofrecidas vivas y detectadas después de la ingestión por la técnica del CARD-FISH. Obtuvimos tasas de ingestión de los flagelados MAST-4 y MAST-1C.

Capítulo 3 "Repuestas funcionales de tres taxones de flagelados heterotróficos procedentes de comunidades naturales"

En este otro apartado se determinó la respuesta funcional (tasa máxima de ingestión y constante de semi-saturación) de tres taxones de flagelados heterotróficos (MAST-4, "*Candidatus* Minorisa minuta" y *Paraphysomonas* sp.) y de la comunidad total de ensamblajes naturales mezclados. Utilizamos bacterias marcadas fluorescentemente añadidas en diferentes concentraciones (de 10^5 a 10^7 cél ml^{-1}) y las contamos dentro de las vacuolas alimenticias de los protistas.

• *Objetivo 3: Estructura genética y patrones de evolución*

Capítulo 4 "Baja diversificación evolutiva en un abundante y extensamente distribuido protista no cultivado (MAST-4)"

En este estudio se investigó la diversidad del MAST-4, con el objetivo de evaluar sus límites y estructura. Obtuvimos secuencias de ADNr (por pirosecuenciación y clones con una amplia cobertura del operón del ADNr) y las completamos con las secuencias del GenBank. También evaluamos las regiones conservadas de las estructuras secundarias del ITS1 y el ITS2 para delinear las diferentes especies biológicas.

• *Objetivo 4: Biogeografía*

Capítulo 5 "Biogeografía del picoeucariota marino no cultivado MAST-4: la temperatura dirige los patrones evolutivos"

Por último, estudiamos la biogeografía del MAST-4 combinando la técnica de identificación genética de ARISA con bibliotecas genéticas de la región del ITS1. Este trabajo aborda esta pregunta mediante el estudio de la tendencia espacial y temporal de comunidades de MAST-4 y sus factores ambientales asociados.

Síntesis de Resultados y Discusión General

El principal objetivo de esta tesis fue estudiar la ecología del estramenópilo marino no cultivado de extensa distribución MAST-4, utilizándolo como un modelo de los flagelados marinos heterotróficos. Se han abordado diferentes temas, incluyendo el alcance de su distribución, la abundancia, el papel trófico, la estructura genética, los patrones evolutivos y la biogeografía. En la siguiente discusión evaluaré los resultados obtenidos en cada uno de estos temas por separado. En general, esta tesis confirma el importante papel del taxón MAST-4 en los ambientes marinos planctónicos.

Abundancia y distribución

Para calcular la abundancia y distribución del flagelado heterotrófico no cultivado MAST-4 en grandes colecciones de muestras medioambientales presentamos un protocolo robusto de Q-PCR (PCR cuantitativa) utilizado en una rápida cuantificación de las moléculas de ADNr **(capítulo 1)**. Una dificultad, cuando se trata de organismos no cultivados, es que se cuantifican moléculas de interés en lugar de células. Para convertir la abundancia de las moléculas a células hace falta conocer el número de copias del gen. El número de copias del operón del ADNr del MAST-4 se estimó comparando las señales de Q-PCR con las de FISH (hibridación fluorescente *in situ*) en las mismas muestras y resultó alrededor de 30. Este número relativamente bajo es consistente con su pequeño tamaño, 2-3 µm de diámetro (Massana et al, 2006a) y encaja perfectamente dentro de la relación descrita para 18 cepas de fitoplancton (Zhu et al, 2005) (figura S.1). Sin embargo el número de copias no resultó el mismo en los diferentes conjuntos de muestras analizadas. Consideramos que

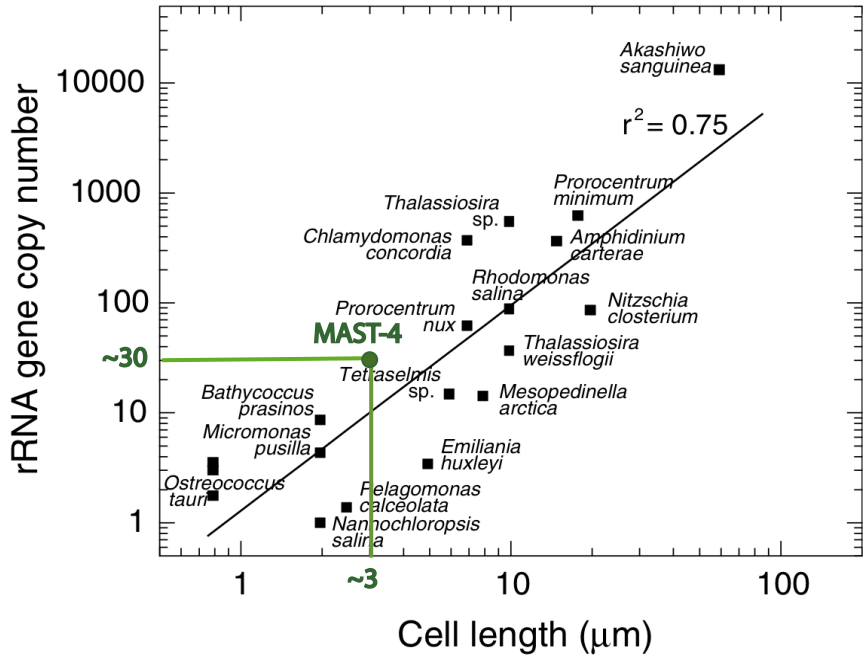


Figura S.1 Correlación entre el número de copias del ADNr estimado por Q-PCR y la longitud de la célula en 18 cepas de fitoplancton (Zhu et al 2005). La posición estimada del MAST-4 está incluida en la figura.

la principal causa de esta mayor variabilidad y menor número de copias del ADNr en las muestras ambientales es que no fueron recogidas para ser procesadas cuantitativamente. En efecto, el grado en el que este protocolo da abundancias absolutas depende del grado de precisión durante la recogida de la muestra y la extracción de ADN. Sin embargo, una explicación alternativa para esta señal ruidosa en las muestras ambientales podría ser que están compuestas por distintos linajes del MAST-4, mientras que las incubaciones podrían estar seleccionando un solo ribotipo. Como se ha visto en el **capítulo 4**, el grupo MAST-4 se compone en efecto por varias especies, las cuales podrían tener diferente número de copias en un rango similar y aparecer mezcladas en las muestras ambientales. Además, la incubación de Blanes parece seleccionar un único ribotipo.

Teniendo en cuenta las restricciones de nuestras muestras ambientales, está claro que la aproximación que ofrece la Q-PCR **(capítulo 1)** es útil desde un punto de vista global para los patrones de distribución del MAST-4 (presencia y predominancia) y para realizar una estima mínima de la abundancia del ADNr. Esto se aplicó a una amplia colección de muestras a diferentes profundidades de diferentes campañas oceanográficas, incluyendo un transecto del Sur de África a Australia atravesando el Giro del Océano Índico, una sección de costa a mar abierto de la parte norte del Giro del Oeste del Alborán y un transecto a través del Frente Polar en el Pasaje de Drake, desde aguas del Atlántico Sur hasta aguas Antárticas (figura 3, **capítulo 1**). En los estudios del Océano Índico y del Mar de Alborán se detectó un patrón similar, con abundancias mayores en las muestras de la costa que en mar abierto y presentes en toda la capa fótica, siendo casi dos veces más abundantes en el DCM (profundidad del máximo de clorofila) que en la superficie. El MAST-4 presentó abundancias muy bajas en la zona afótica superior y prácticamente no fue detectado a mayor profundidad, lo cual concuerda con el hecho de que los flagelados heterotróficos (HF) y su alimento bacteriano son escasos en las aguas mesopelágicas (Fukuda et al, 2007; Tanaka y Rassoulzadegan, 2002). La temperatura del agua de mar parece ser el segundo factor en la distribución del MAST-4. Por lo tanto, las moléculas del MAST-4 solo se detectaron en las dos estaciones más al norte del Pasaje del Drake, las que ofrecían temperaturas más templadas. En las otras estaciones, con temperaturas por debajo de los 5ºC, no se detectó ninguna molécula diana. De esta manera, el MAST-4 parecía estar excluido en aguas por debajo de 5ºC, una característica intrigante compartida con otros microorganismos como las picocianobacterias (Partensky et al, 1999).

Aparte de las tres campañas oceanográficas, también se analizó un muestreo mensual por un periodo de seis años en el Observatorio Microbiológico de la Bahía de Blanes (figura 4, **capítulo 1**). No se observó un patrón estacional claro en la abundancia del MAST-4 en cambio se observó la aparición de cambios abruptos en fechas consecutivas. Una razón podría ser una variación del MAST-4 a corta escala de tiempo, con lo cual el muestreo mensual no sería suficiente para describir adecuadamente la variación temporal. Otra alternativa, como se propone en (Piwosz y Pernthaler

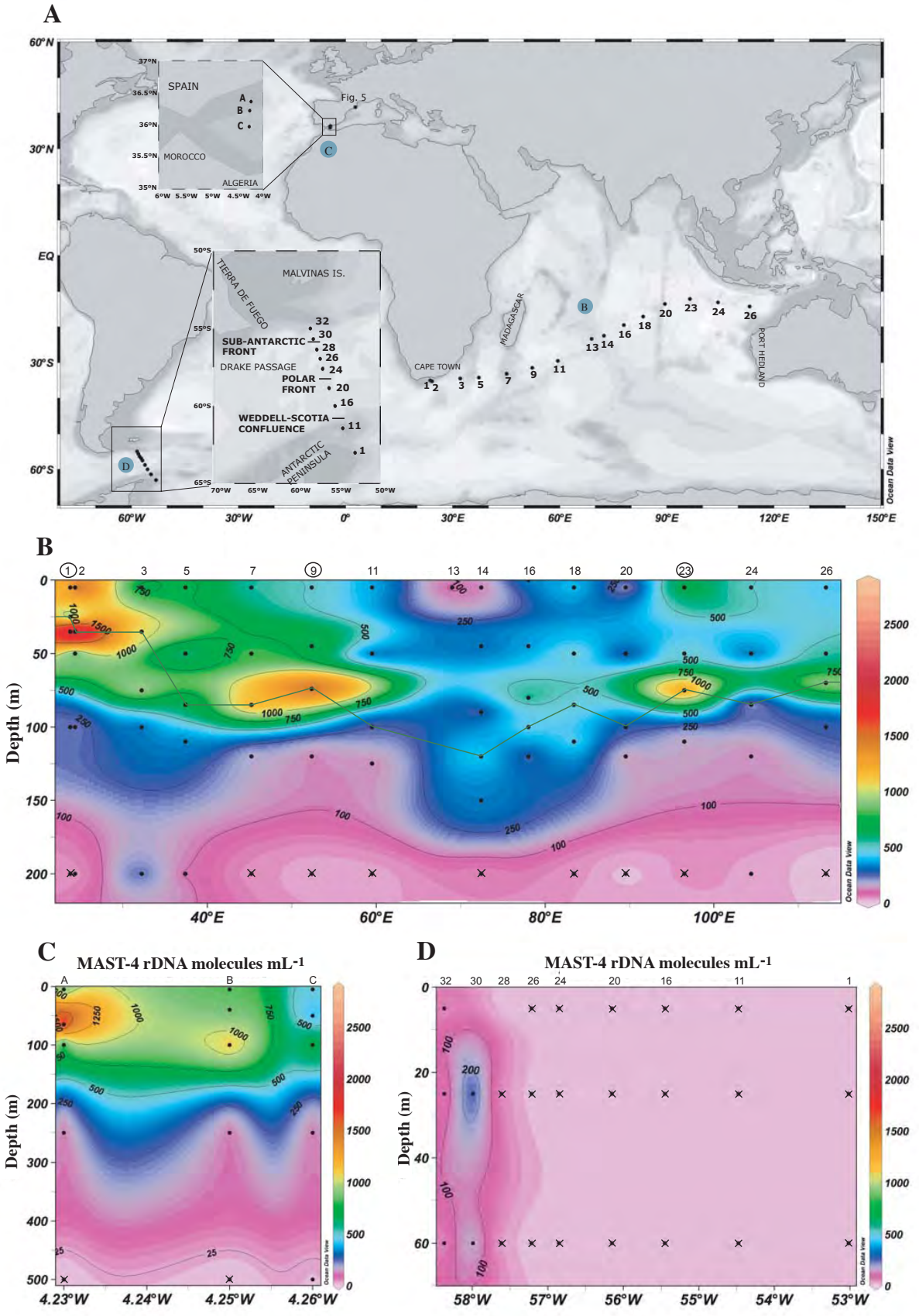


Figura 3, capítulo 1. Visión general de los sistemas marinos investigados (A) y abundancia del MAST-4 (ADNr 18S moléculas ml^{-1}) en tres de ellos a varias profundidades: (B) transecto en el Océano Índico. La línea verde marca el DCM (profundidad del máximo de clorofila) y las estaciones analizadas por FISH aparecen rodeadas por un círculo. (C) Transecto en el Mar de Alborán. (D) Transecto en el Pasaje del Drake. Las muestras con una cruz en B, C y D indican ausencia de amplificación.

2010), es que la aparente carencia de estacionalidad puede derivar de una sucesión de linajes diferentes. En el **capítulo 4** se determinó que el grupo MAST-4 estaba compuesto por al menos 5 especies biológicas diferentes. Y en el **capítulo 5** se observó que la composición de la comunidad del MAST-4 en la Bahía de Blanes dependía de la temperatura de la muestra. Por lo tanto, los diferentes componentes del MAST-4 se sucedían en esta estación costera a lo largo del ciclo termal estacional y esto podría causar la falta de patrones estacionales cuando se analiza el grupo al completo. Desenmarañar las dinámicas particulares de estos linajes es un reto para estudios futuros.

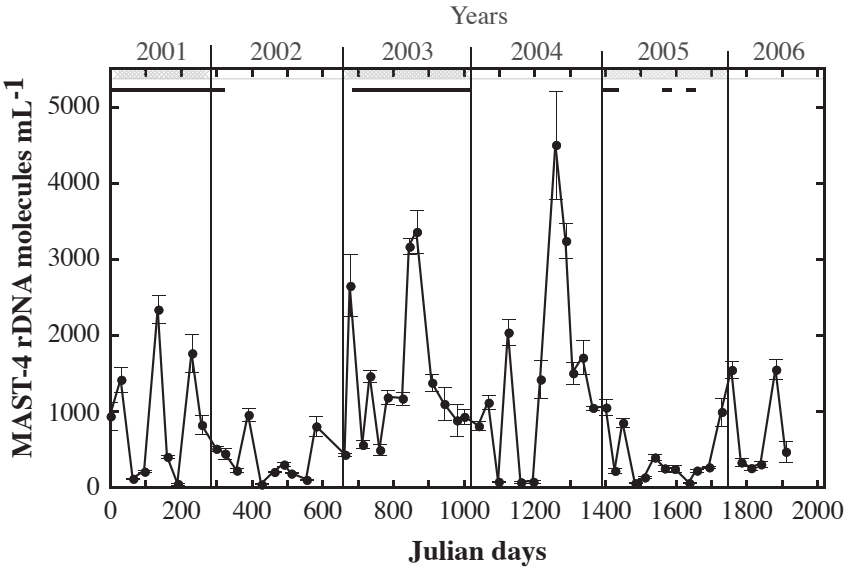


Figura 4, capítulo 1. Abundancia (media y error estándar) de las moléculas del MAST-4 en la Bahía de Blanes durante seis años de estudio con un muestreo mensual. Las líneas negras en la parte superior de la figura marcan los periodos donde se obtuvieron datos con FISH.

La abundancia promedio del MAST-4 en el Océano Índico y el Mar Mediterráneo, sistemas con temperaturas medias comparables (16-24°C), eran similares, alrededor de 800 moléculas ADNr ml^{-1} (tabla 1, **capítulo 1**). El MAST-4 no aparecía en las frías aguas Antárticas y se presentaba en bajas proporciones al norte del frente del Atlántico Sur. Prácticamente todas las muestras de aguas epipelágicas (de superficie a 120 m) con temperaturas superiores a 5°C contienen moléculas de MAST-4. Por otra parte, en el **capítulo 1** se añadieron conteos adicionales de FISH y el promedio de la abundancia ya conocida de 131 cél ml^{-1} (Massana et al 2006b) fue apoyado con una media de 100 cél ml^{-1} en la superficie y 150 cél ml^{-1} en el DCM. Esta presencia amplia y sistemática de MAST-4 en los sistemas planctónicos marinos es similar a la descrita en algunas bacterias marinas como el SAR11 (Morris et al, 2002), *Roseobacter* (Selje et al, 2004) y *Prochlorococcus* (Partensky et al, 1999). Sin embargo, debemos tener en cuenta la diversidad filogenética del MAST-4. Por lo tanto, a pesar de que tienen el mismo aspecto por FISH (Massana et al, 2006a), incluyen linajes distintos con adaptaciones ecológicas diferentes y complementarias **(capítulo 4** y **5)** que podrían explicar esta amplia distribución, como se ha propuesto para otros picoeucariotas (Rodríguez et al, 2005).

Tabla 1, capítulo 1. Número de moléculas y células cuantificadas por QPCR y FISH respectivamente provenientes de los diferentes lugares analizados en este estudio. Relación con la temperatura.

	18S rDNA molecule number			Cell number			Temperature (ºC)	
	Mean	Range	n	Mean	Range	n	Mean	Range
INDIAN OCEAN								
Above DCM	504	52 -2484	28	98	82 - 118	5	24	21 - 24
DCM	864	91 - 2845	14	155	79 - 262	3	22	19 - 25
Upper aphotic (200m)	48	0 - 285	13	0	0	3	16	8 - 20
ALBORAN SEA								
Surface (5m)	706	469 - 826	3				16.5	16.5
Subsurface (50-100m)	954	320 - 2205	6				15	14 – 16.5
Upper aphotic (250m)	38	21 - 48	3					
DRAKE PASSAGE								
Above SAF (5-100m)	150	31 - 351	6				5.7	5.5 – 5.8
Below SAF (5-60m)	0	0	20				1.3	(-1) – 4.4
BLANES BAY								
Surface (5m)	926	34 - 4500	62	91	18 - 244	26	17.6	11 - 26

DCM, Deep Chlorophyll Maximum
SAF Sub Antarctic Front

Papel trófico

Para estudiar el papel trófico de los HF no cultivados que habitan los ambientes naturales, presentamos una aproximación basada en la estima de la actividad de alimentación de predadores específicos detectados mediante FISH después de experimentos de ingestión a corto plazo con diferentes presas. Como presas se incluyeron las frecuentemente utilizadas bacterias marcadas con fluorescencia (FLB) **(capítulo 2** y **3)** y dos cepas bacterianas de la familia Rhodobacteraceae (MED479) y Flavobacteriaceae (MED134), con varios tamaños celulares. Estas dos últimas fueron añadidas vivas y detectadas después de su ingestión por un segundo paso con FISH **(capítulo 2)**. El objetivo de esta parte fue determinar las tasas de depredación y la preferencia de presas del MAST-4.

Las tasas de depredación del MAST-4 *in situ* y en muestras incubadas fueron comparables (**capítulo 2**). Esto sugiere que las células del MAST-4 no estaban estimuladas artificialmente por la incubación y que la aparición de tasas de ingestión más altas en la muestra incubada fue debida a la mayor abundancia de presas. Además, se vieron diferencias importantes en las tasas de depredación con el uso de distintas presas. Las FLBs dieron las tasas de ingestión más bajas (1 bacteria predador^{-1} h^{-1}) y las MED134 las tasas más altas (3 bacteria predador^{-1} h^{-1}). Estas diferencias no se pudieron explicar por el tamaño celular de las presas, ya que las tasas variaban enormemente utilizando presas de volumen similar. Un patrón claro y estadísticamente significativo surgió al relacionar las tasas de depredación con el porcentaje de células vivas en las presas añadidas (figura 3, **capítulo 2**). Por lo tanto, el MAST-4 parecía preferir las bacterias en un buen estado fisiológico,

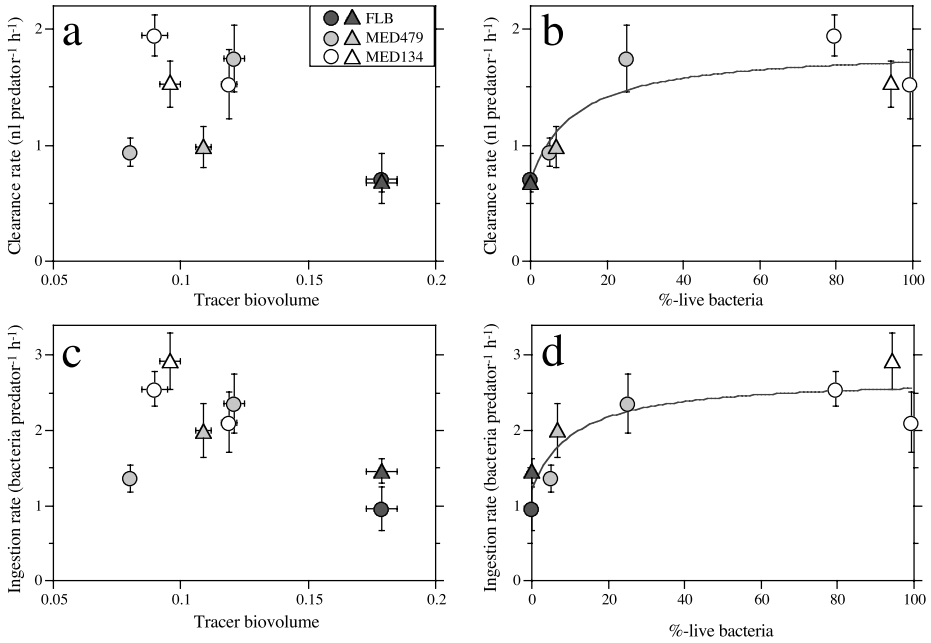


Figura 3, capítulo 2. Tasas de aclarado (a, b) y tasas de ingestión (c, d) de las células del grupo 4 de los estramenópilos marinos (MAST-4) representadas en función del biovolumen (a, c) y del porcentaje de ácidos nucleicos de doble tinción (NADS) positivos que indican las células "vivas" (b, d) del marcador bacteriano utilizado en cada uno de los ocho experimentos independientes (los círculos representan las muestras *in situ*; los triángulos representan las muestras incubadas). FLB se refiere a bacterias marcadas con fluorescencia, MED479 a la cepa de Roseobacter y MED134 a la cepa de flavobacterias. En (b) y (d) se utilizó un ajuste hiperbólico (ecuación de Michaelis–Menten con una constante inicial). Las barras representan el error estándar.

dando unas tasas de depredación 2-3 veces superiores para las bacterias vivas sobre las FLB muertas. Estos resultados encajaban bien con la visión general de que cuando se utilizaban bacterias vivas como marcador en los experimentos de ingestión, se obtenían tasas de ingestión superiores a las de las FLB (Boenigk et al 2001, Landry et al 1991), con una preferencia por el consumo de las bacterias en crecimiento sobre las hambrientas (González et al 1993). Un caso extremo de selección negativa se vio en los experimentos de células MED134 muertas por calor, que no eran ingeridas en absoluto. Por último, aparte de las diferencias relacionadas con la viabilidad celular, no se observó ninguna otra diferencia en las tasas de depredación por el uso de dos cepas bacterianas diferentes. Por lo tanto, las células de MED479 y MED134 se ingerían en partes iguales por el MAST-4, por lo que las supuestas diferencias importantes en la filogenia y estrategia de vida entre estas dos cepas no determinaban la preferencia de la presa.

Las tasas específicas de depredación del MAST-4 medidas en el **capítulo 2** son comparables con las tasas medidas *in situ* en otros estudios ambientales (Unrein et al 2007, Vaqué et al 1994), pero mucho menores que la mayoría de estimas en cepas cultivadas. Sin embargo, estas comparaciones están afectadas por el valor real de la abundancia de presa durante las estimaciones de la depredación, el cual se sabe que influye drásticamente en la medida de las tasas, dando la forma

de la función de la respuesta funcional (Holling 1959). La respuesta funcional y numérica (relación de las tasas de crecimiento con la abundancia de presas) ha sido estudiada repetidamente en cepas cultivadas de flagelados heterotróficos. Sin embargo, a nuestro entender, no existen estudios que analicen la respuesta funcional de flagelados no cultivados. Para su estudio, hemos proporcionado diferentes abundancias de FLB, llegando a ser éstas en muchas de las botellas la principal presa. Posteriormente, los experimentos de ingestión a corto plazo se analizaron con recuentos de FISH de los depredadores en las comunidades mezcladas **(capítulo 3)**. Para minimizar el problema de las bajas densidades de predadores realizamos los experimentos utilizando incubaciones de agua de mar no modificadas, de las que se conoce que promueven el crecimiento del MAST-4. Por primera vez se determinó la respuesta funcional (tasa de ingestión máxima y constante de semi-saturación, Ks) de un flagelado heterotrófico no cultivado. La tasa máxima de ingestión era comparable con las estimas previas de depredación (figura S.2) y el hallazgo más notable fue que la Ks del MAST-4 era 8.7 10^5 presas ml^{-1} (figura 3, **capítulo 3**). Por lo tanto, este flagelado heterotrófico abundante

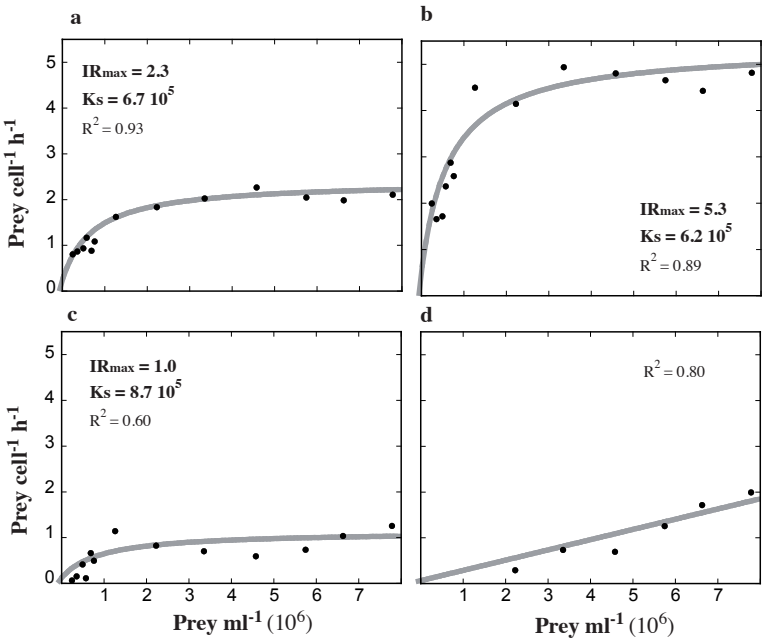


Figura 3, capítulo 3. Respuestas funcionales (relación de la tasa de ingestión con la abundancia de presa) de las comunidades naturales de flagelados heterotróficos (a), *Minorisa minuta* candidatus (b), células del MAST-4 no cultivado (c) y *Paraphysomonas* sp. (d). Ks = constante de semi-saturación (presa ml^{-1}) y IR_{max}= tasa de ingestión máxima (presa cél.$^{-1}$ h^{-1}).

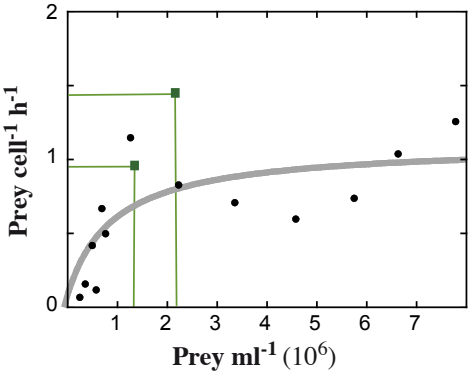


Figura S.2. Respuesta funcional del MAST-4 utilizando FLBs (**capítulo 3**), junto con las tasas de depredación superpuestas calculadas en el **capítulo 2**.

y ampliamente distribuido parece bien adaptado a la abundancia bacteriana de los ambientes planctónicos marinos, típicamente alrededor de 10^6 bacteria ml^{-1} (Fuhrman y Hagström 2008). Por lo contrario, la Ks de los HF cultivados es frecuentemente al menos un orden de magnitud superior, con un rango de 1.1 a 45 10^6 bacteria ml^{-1} (tabla 2, **capítulo 3**). Estas Ks elevadas son los valores esperados de organismos que crecen eficientemente en medios ricos y al mismo tiempo establecen una limitación obvia para su desarrollo en las concentraciones bacterianas habitualmente bajas *in situ*.

Flagellate cultures	Prey	μ_{max}	IR_{max}	Ks	Ref. [a]
Actinomonas mirabilis	*Pseudomonas sp.*	0.250		1.4	*1*
Bodo designis	*B1*	0.160		3.4	*2*
	Aeromonas sp.	0.120		8.8	*3*
Ciliophrys infusionum	*B1*	0.045		45.0	*2*
Codosiga gracilius	*B1*	0.052		9.7	*2*
Diaphanoeca grandis	*Pseudomonas sp.*	0.120		2.4	*4*
Jakoba libera	*Aeromonas sp.*	0.080		5.3	*3*
	B1	0.036		5.4	*2*
Monosiga sp.	*Pseudomonas sp.*	0.170		13.5	*5*
Ochromonas sp.	*Pseudomonas sp.*	0.190		19.0	*5*
Paraphysomonas vestita	*Pseudomonas sp.*	0.230		14.9	*5*
P. Imperforata	*B1*	0.210		1.1	*2*
	Aeromonas sp.	0.120		4.4	*3*
	Vibrio sp.	0.220		13.0	*6*
Pleuromonas jaculans	*Pseudomonas sp.*	0.160		38.6	*5*
Pseudobodo tremulans	*Pseudomonas sp.*	0.150		8.4	*5*
Stephanoeca diplocostata	*Pseudomonas sp.*	0.076		6.8	*7*
	B1	0.035		2.3	*2*
Cafeteria roenbergensis	*Photobacterium angustum*	0.260		5.8	*8*
	Vibrio vuinificus	0.210		2.7	*8*
	Sphingopyxis alaskensis	0.240		7.4	*8*
Cafeteria sp	*Mixed bacterial communities*	0.041		8.7	*9*
	Flavobacterium sp.	0.041		9.1	*9*
	Alteromonas sp.	0.040		9.2	*9*
	Pseudomonas sp.	0.040		9.1	*9*
Jakoba libera	*Mixed bacterial communities*	0.024		5.1	*9*
	Flavobacterium sp.	0.032		3.7	*9*
	Alteromonas sp.	0.004		9.5	*9*
	Pseudomonas sp.	0.041		1.4	*9*
Poterioochromonas malhamensis	*Polinucleobacter*	0.042		18.2	*10*
	Listonella pelagia	0.071		1.5	*10*
Spumella sp.	*Polinucleobacter*	0.083		20.5	*10*
		0.096		22.0	*11*
	Listonella pelagia	0.100		1.2	*10*
		0.117		1.2	*11*
	Bodo	0.161		2.4	*12*
		0.183		3.9	*12*
			72.8	7.4	*12*
			64.6	3.5	*12*

[a] References: *1*) Fenchel 1982b, *2*) Eccleston-Parry and Leadbeater 1994, *3*) Hammond 1991, *4*) Andersen 1989, *5*) Fenchel 1982a, *6*) Edwards 1989, *7*) Geider and Leadbeater 1988, *8*) Anderson et al 2011, *9*) Mohapatra and Fukami 2004, *10*) Boenigk et al 2006, *11*) Pfandl and Boenigk 2006, *12*) Jürgens 1995

Tabla 2, capítulo 3. Comparación de las tasas de crecimiento específicas máximas (μ_{max}, h^{-1}), tasas de ingestión máximas (IR_{max}, presa cél.$^{-1}$ h^{-1}) y constante de semi-saturación (Ks, 10^6 presa ml^{-1}) para varias especies de flagelados cultivados y con distintos tipos de presas.

En el **capítulo 2** también comparamos las tasas de depredación del MAST-4 junto con otros flagelados no cultivados, el MAST-1C y el total de eucariotas. Se mostraron claramente diferencias funcionales entre ambos taxones, aunque todavía se desconocen los mecanismos subyacentes de estas diferencias. Al comparar las tasas de depredación utilizando FLBs, el MAST-4 apareció menos activo que la comunidad total de eucariotas en contraste con el MAST-1C que aparecía más activo (figura 4, **capítulo 2**). Este resultado concordaba con el tamaño mayor del MAST-1C frente al MAST-4. Respecto a la viabilidad de las presas utilizadas, el patrón observado para el conjunto de los eucariotas resultó similar al del MAST-4, dando las mayores tasas el MED134 frente a las FLBs. Por el contrario, el MAST-1C se desviaba del resultado general, con la obtención de las tasas más altas en la depredación de las FLBs y con una menor ingestión de las bacterias vivas (la MED134 no fue ingerida en absoluto). Es evidente que las preferencias alimentarias de MAST-4 y MAST-1C eran diferentes y este último no se comportaba como el conjunto típico de bacterívoros. Una explicación razonable podría ser que el límite del tamaño óptimo de presa para el MAST-1C se encuentra fuera del rango de tamaño de bacterias utilizadas. La cepa MED134, es la bacteria más pequeña analizada y podría encontrarse fuera del rango de tamaño comestible para este predador y escapar de su depredación. El hecho de que el MAST-1C pueda estar adaptado a una alimentación de bacterias más grandes que las del MAST-4 es consistente también con su mayor tamaño, siguiendo la relación establecida entre tamaño del predador y de la presa (Fenchel 1987). En general, está aceptado que el tamaño de la presa es el factor principal en la vulnerabilidad de la depredación (Gonzalez et al 1990), con fuertes límites de tamaño fuera de los cuales las presas no pueden ser ingeridas (Fenchel 1987, Jürgens y Matz 2002). Es entonces, al encontrarse todas las presas dentro del tamaño comestible, cuando otros factores entran en juego con un menor impacto. Por ejemplo, el MAST-4 prefería las bacterias

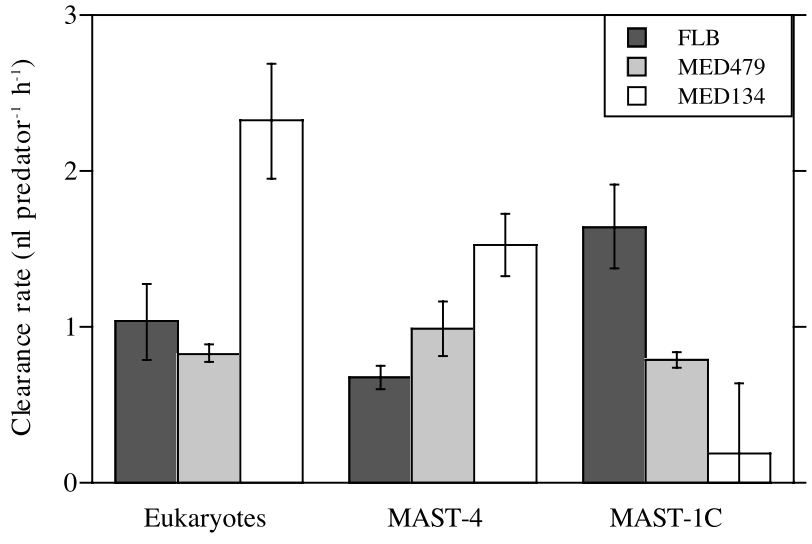


Figura 4, capítulo 2. Tasas de aclarado del conjunto de los eucariotas (en su mayoría células de flagelados heterotróficos (HF)), de los estramenópilos marinos específicos pertenecientes al grupo 4 (MAST-4) y del taxón MAST-1C, calculadas en la muestra incubada y estimadas con tres diferentes presas bacterianas diana: bacterias marcadas con fluorescencia (FLB), MED479 (cepa de Roseobacter) y MED134 (cepa de flavobacterias). Las barras representan los errores estándares.

vivas en un buen estado fisiológico, aunque también se alimentaba de FLB muertas en una tercera parte de la tasa máxima. Nuestros datos no revelaban diferentes comportamientos de alimentación relacionados con la afiliación filogenética de las bacterias analizadas.

Por otra parte, los diferentes taxones dentro del conjunto de la comunidad presentaron diferentes respuestas funcionales (figura 3, **capítulo 3**), por lo tanto tenían diferentes nichos ecológicos, quizás incluso con diferentes estrategias de filtración y preferencias de presa. "*Candidatus* Minorisa minuta", un flagelado cultivado recientemente utilizando técnicas que imitaban las condiciones naturales (del Campo 2011), resultó estar bien adaptado a la baja abundancia de presas siendo muy eficiente en la ingestión de las FLBs y el filtrado de agua. Conjuntamente con el MAST-4, ambos taxones aparecen como bacterívoros típicos en las comunidades naturales, siendo sus respuestas funcionales comparables con las determinadas para el total de los eucariotas. No obstante, *Minorisa* parecía ser más voraz que MAST-4, con una tasa máxima de ingestión cinco veces mayor, lo cual probablemente explica las altas tasas de crecimiento medidas de *Minorisa* frente a las de MAST-4 (1.56 versus 0.62 d^{-1}). Una de las razones de la baja tasa de depredación del MAST-4, vista en los **capítulos 2** y **3,** podría deberse a una selección negativa de las FLBs muertas por calor, prefiriendo bacterias vivas en un buen estado fisiológico, aunque lo más probable es que esto sea también lo que ocurre para la mayoría de los taxones de HF en las comunidades naturales (Fu et al 2003, Landry et al 1991). Por último, tuvimos la suerte de poder encontrar en los experimentos de depredación un flagelado heterotrófico típico de los cultivos, *Paraphysomonas* spp., que no alcanzó la saturación de la tasa de ingestión incluso con las mayores concentraciones de presa utilizadas (casi 10^7). Con estos resultados deberíamos esperar una Ks alta, similar a la de los flagelados cultivados. Las cepas de *Paraphysomonas* se obtienen comúnmente a partir de cultivos donde se alimentan con bacterias grandes a densidades muy altas por lo que no es de extrañar que no sean abundantes en el mar.

Nuestros estudios revelan importantes diferencias funcionales entre diferentes protistas no cultivados y podrían ser un buen punto de partida para desenmarañar la complejidad de las redes tróficas microbianas. Diferentes taxones de flagelados parecen tener diferentes tasas de depredación y preferencias de presa, por lo tanto, la diversidad funcional observada aquí proporciona un significado ecológico para la enorme diversidad filogenética de protistas heterotróficos marinos (Vaulot et al 2002). Por ejemplo, datos recientes obtenidos a través de la secuenciación de células aisladas (Martínez-García et al 2012) sugieren que el MAST-4 podría estar adaptado a la depredación de una presa específica, como la *Pelagibacter ubique*. Además, nuestros resultados asientan las bases de las diferencias fundamentales entre depredadores bacterianos cultivados y no cultivados y podrían explicar el por qué de que los bacterívoros marinos más predominantes no se consigan cultivar mediante los métodos clásicos. Los HF naturales daban una Ks de 6-9 10^5 presas ml^{-1}, menor

que la de los tradicionales flagelados cultivados, por lo que se ven bien adaptados a las abundancias bacterianas típicas de los ambientes marinos planctónicos.

Estructura filogenética y patrones evolutivos

El tamaño de la población total del MAST-4 se estima en unas 10^{24} células. La manera en que esta enorme cantidad de células se agrupa en diferentes linajes se puede evaluar mediante la estructura filogenética de las secuencias de ADNr existentes. A su vez, esta estructura puede ser utilizada para inducir los patrones evolutivos del grupo. Para una mejor comprensión de la estructura filogenética del MAST-4 y de su significado biológico, en el **capítulo 4** se secuenció un fragmento largo del operón del ADNr, incluyendo la región altamente variable del ITS (espaciador interno transcrito) y el comienzo del 28S. También se recopilaron todas las secuencias públicas disponibles del ADNr 18S del MAST-4 y se incluyeron nuevas secuencias obtenidas por pirosecuenciación de la región V4 del ADNr 18S. Además en el **capítulo 5** incrementamos diez veces el número de secuencias dentro de este grupo, englobando la región V9 del ADNr 18S y la región ITS1.

Estableciendo el número de los principales clados

El árbol filogenético de las distintas secuencias del ADNr 18S de MAST-4 recuperadas de una búsqueda exhaustiva en GenBank reveló solo cinco clados principales (figura 2, **capítulo 4**), mostrando cada uno una baja divergencia entre las secuencias (distancia genética no corregida por debajo de 0,012). Las pirosecuencias, recuperadas de un amplio análisis ambiental, añadían más de un orden de magnitud en número de secuencias y confirmaban la baja diversidad del MAST-4, ya que todas las secuencias 454 se afiliaban a los cinco clados conocidos, sin ningún clado adicional. La única secuencia divergente del GenBank, tampoco concordaba con ninguna pirosecuencia nueva y podría ser un pseudogén (Thornhill et al 2007). El árbol del ADNr 18S para los nuevos clones secuenciados en el **capítulo 4** concordaba con los mismos clados. El árbol del ADNr 28S mostraba los mismos cinco clados que el árbol del ADNr 18S, aunque estos aparecían con una mejor resolución y separados con mayores distancias filogenéticas. Finalmente, el árbol basado en el final del ADNr 18S y la región variable del ITS1 revelaba los cinco mismos clados observados previamente, algunas veces organizados en subclados separados (figura 2, **capítulo 5**). Por lo tanto, a pesar de la presencia de un gran número de células de MAST-4 en los océanos y del análisis de cientos a miles de secuencias ambientales, su diversidad está estructurada en tan solo cinco clados principales.

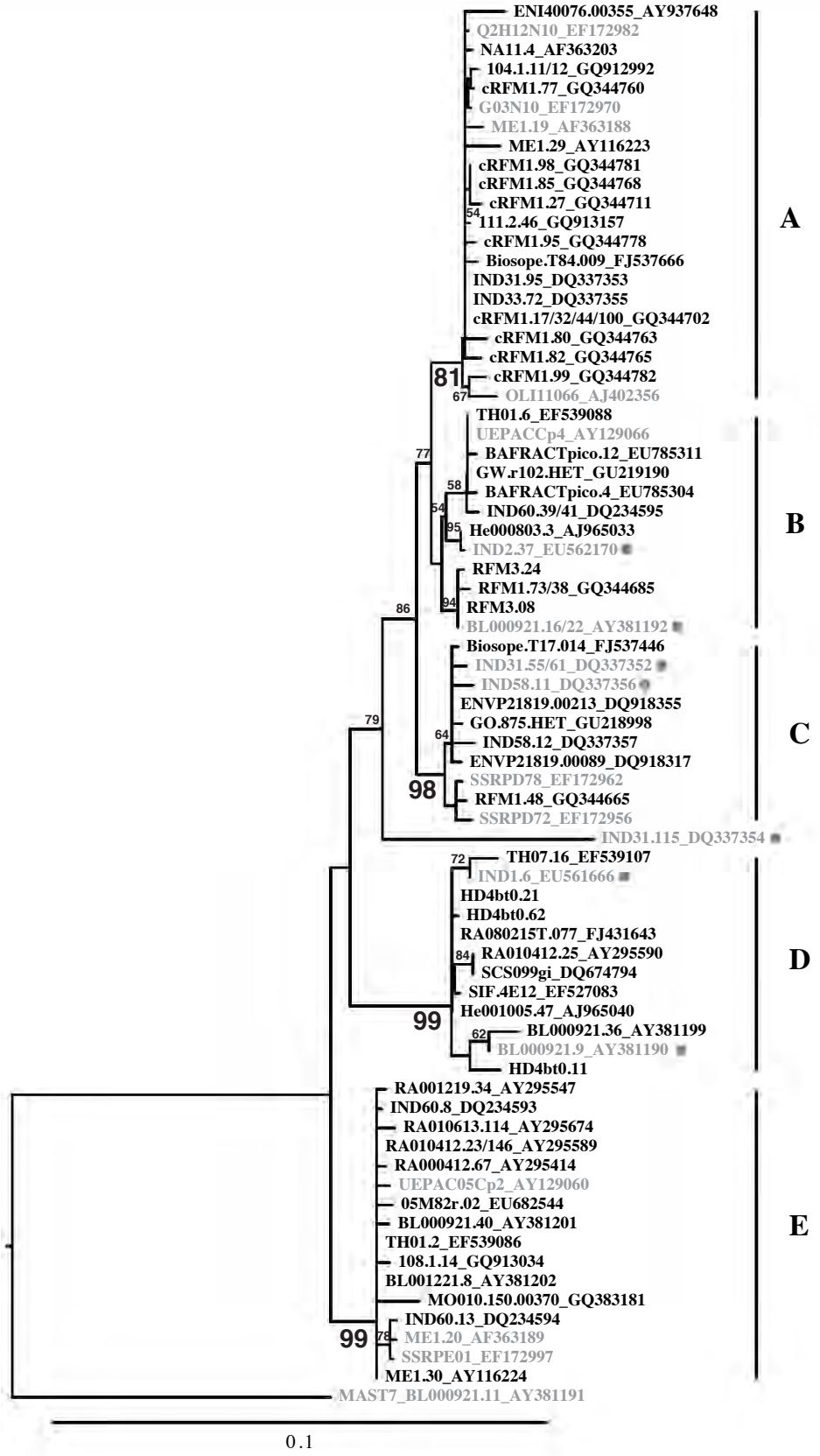


Figura 2, capítulo 4. Árbol filogenético de máxima probabilidad de secuencias completas (gris) y parciales (negro) de ADNr 18S afiliado al MAST-4, procedentes del GenBank. Se muestran los cinco clados principales (nombrados del A al E) y su apoyo "bootstrap". Los ADNr 18S secuenciados completamente en este estudio están marcados con un punto gris.

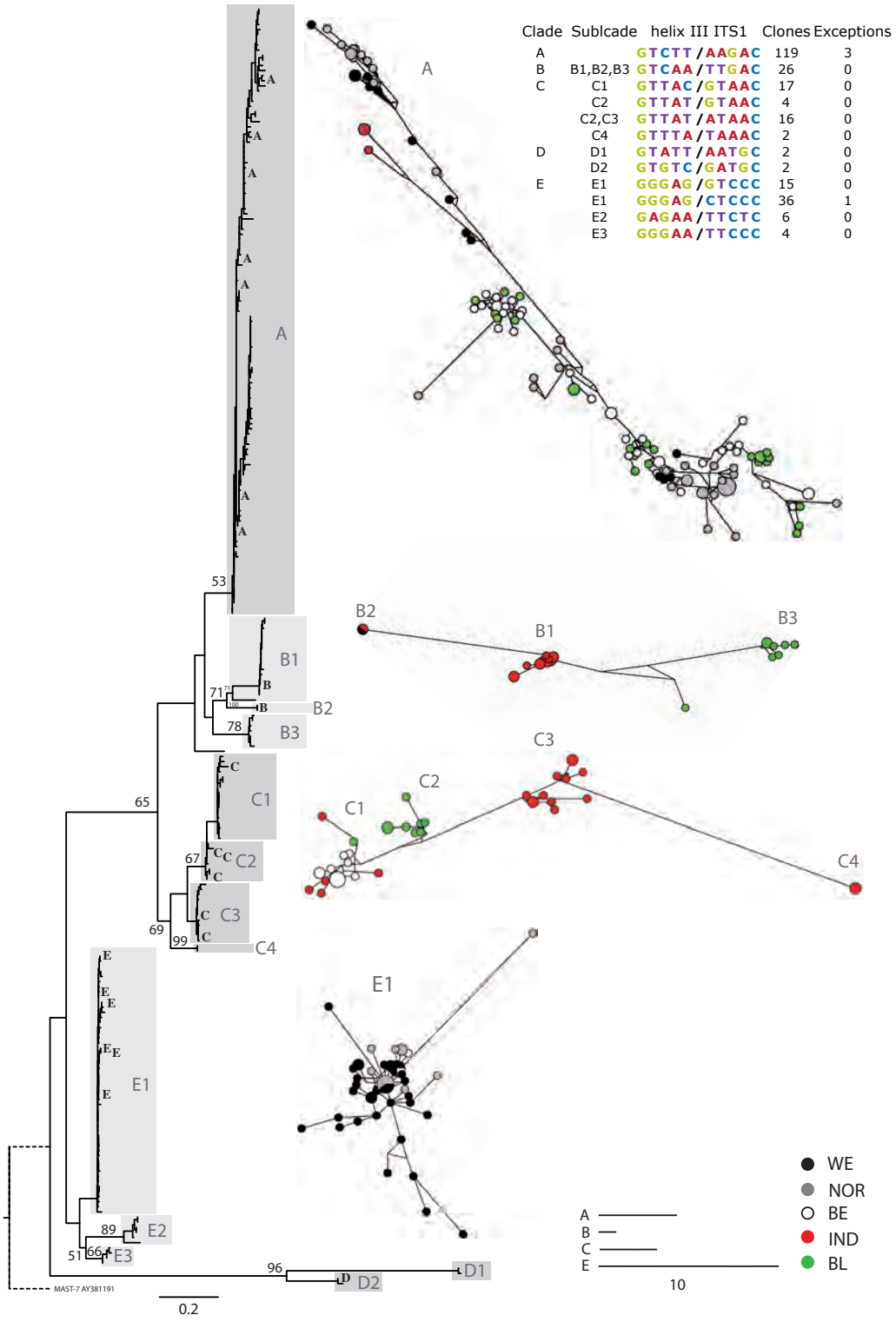


Figura 2, capítulo 5. Árbol filogenético de máxima probabilidad del MAST-4 construido a partir de 228 secuencias nuevas (287 pares de bases del ADNr 18S más el ITS1 completo) y 22 clones de un estudio previo (indicadas con letras). Los clados y subclados están separados por áreas grises. Se muestras los valores de "bootstrap" por encima del 50%. La barra de escala indica 0.2 sustituciones por posición. A la derecha del árbol se encuentras las redes de unión media para cada clado, resaltando las diferencias del ribotipo entre secuencias (los colores diferentes se refieren a las distintas bibliotecas). Las barras de escala indican 10 cambios de base entre ribotipos (considerar las diferentes escalas para cada clado). Arriba a la derecha aparece una lista de las secuencias para el tallo conservado de la hélice III (derivado de la estructura secundaria del ITS1).

Número hipotético de especies

Para la delimitación teórica de las diferentes especies biológicas en el MAST-4 se han utilizado diferentes enfoques **(capítulo 4 y 5)**. En primer lugar, se consideró la estructura secundaria del ITS2 (figura 5, **capítulo 4**), ya que se ha demostrado que las cepas que presentan al menos un cambio de compensación de base (CBC) en los nucleótidos conservados de las hélices II (5 pares de bases) y III (18-30 posiciones continuas) pertenecían a diferentes especies biológicas (Coleman 2003, Coleman 2007, Coleman 2009). La presencia de una hemiCBC (un lado) permitiría un grado débil de reproducción (Coleman 2009). Por ejemplo, se vio que dos cepas de *Pseudo-nitzschia* diferenciadas por tres hemiCBCs producían cigotos pero nunca daban descendencia viable, por lo tanto se consideraban como especies separadas (Amato et al 2007). Un segundo criterio fue la inspección de regiones análogas en la estructura secundaria del ITS1**.** Mientras que el ITS2 se ha investigado extensamente, el ITS1 todavía carece de un modelo universal de estructura secundaria. Para nuestras secuencias de ITS1 establecimos una estructura secundaria común con tres hélices y encontramos una región con un comportamiento similar al ITS2 respecto a la delimitación de especies. Esto consistía en cuatro pares de bases en el tallo III que aparecían conservadas dentro de los clados definidos pero diferían en CBCs entre ellos. Con la expansión al quinto par de bases aparecían polimorfismos dentro de alguno de los clados (figura 6, **capítulo 4**). En tercer lugar contrastamos la topología obtenida de los árboles de las regiones del ITS1 e ITS2 (figura 4B, **capítulo 4**). Planteamos la hipótesis de que los grupos que se habían diversificado lo suficiente para formar especies diferentes presentarían topologías coherentes en los árboles del ITS1 y del ITS2, debido a la ausencia de recombinación entre estos marcadores. En el caso contrario, los grupos que todavía formaran una sola especie deberían presentar topologías incongruentes debido a los eventos recientes de recombinación. Básicamente este es el principio de concordancia-discordancia utilizado para el reconocimiento filogenético de las especies (Taylor et al 2000), que puede aplicarse fácilmente utilizando estos espaciadores de rápida evolución (Coleman 2007, Mullineux y Hausner 2009). En cuarto lugar, comparamos la divergencia de la región ITS entre los distintos clados del MAST-4 con la observada en otras especies (tablas 1 y 2, **capítulo 4**). Por último, utilizamos herramientas de genética de poblaciones en la delimitación de las especies: redes de unión-media (MJ) (Bandelt et al 1999) que teorizan filogenias intraespecíficas y visualizan potenciales caminos evolutivos alternativos (figura 2, **capítulo 5**) y el índice de fijación (Fst) que valora la diferenciación genética entre poblaciones (figura 3, **capítulo 5**).

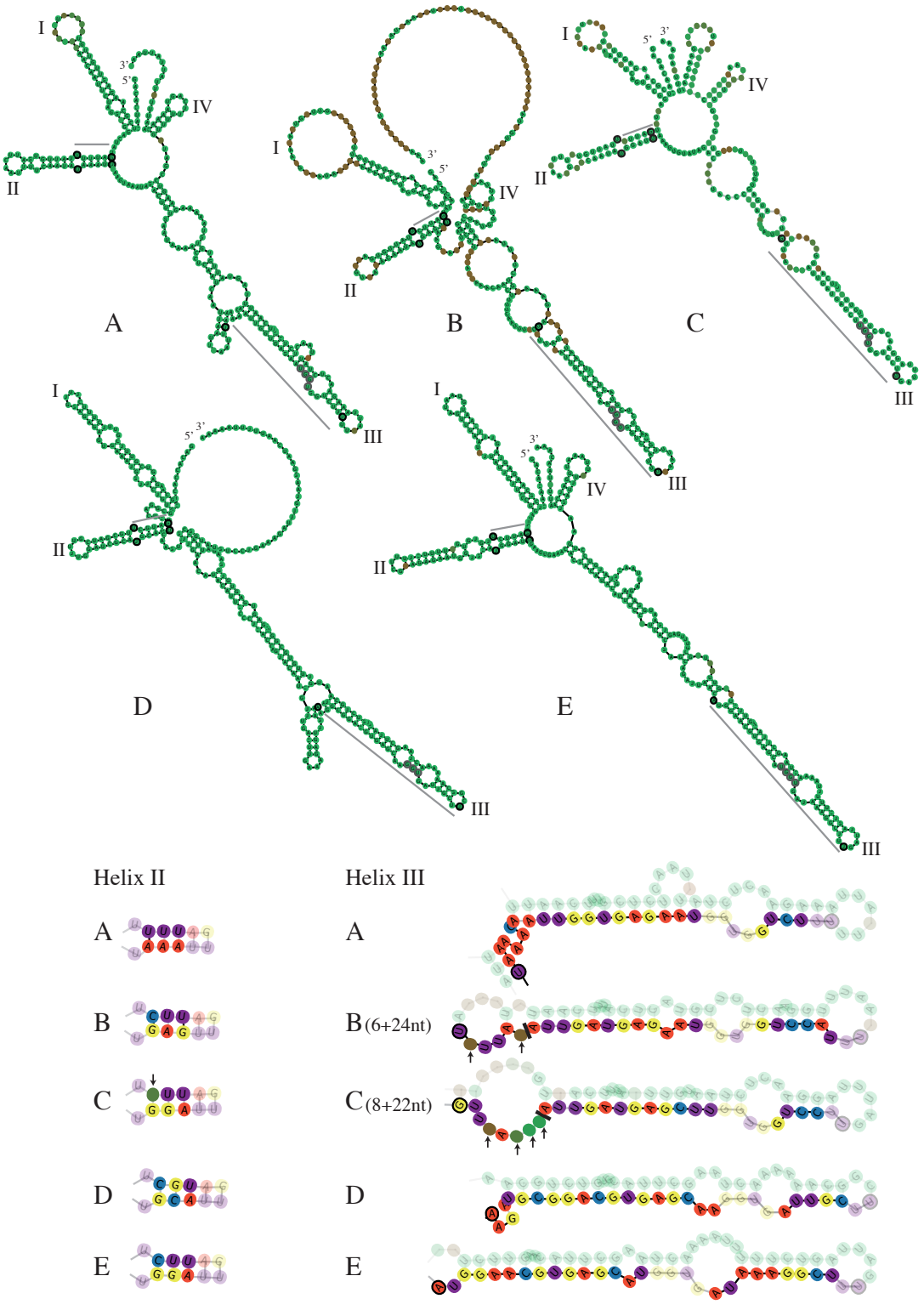


Figura 5, capítulo 4. Estructura secundaria consenso del ITS2 para cada uno de los cinco clados de MAST-4. Muestra tres hélices principales (de la I a la III). Los pares de bases altamente conservadas dentro de cada clado se muestran en verde; las posiciones variables aparecen en marrón. Los nucleótidos con círculos grises representan el motivo UGGU. El detalle de las hélices II y III está representado en la parte de abajo de la figura, resaltando en colores intensos los nucleótidos conservados dentro de cada clado y que difieren entre clados. Los nucleótidos conservados en todas las secuencias aparecen en colores tenues. Las líneas negras de la hélice III delimitan la mayor longitud (si esta es menor que 30 nucleótidos) de las regiones conservadas hasta la aparición de un polimorfismo. Las flechas apuntan a los polimorfismos (en las regiones conservadas) dentro de cada clado.

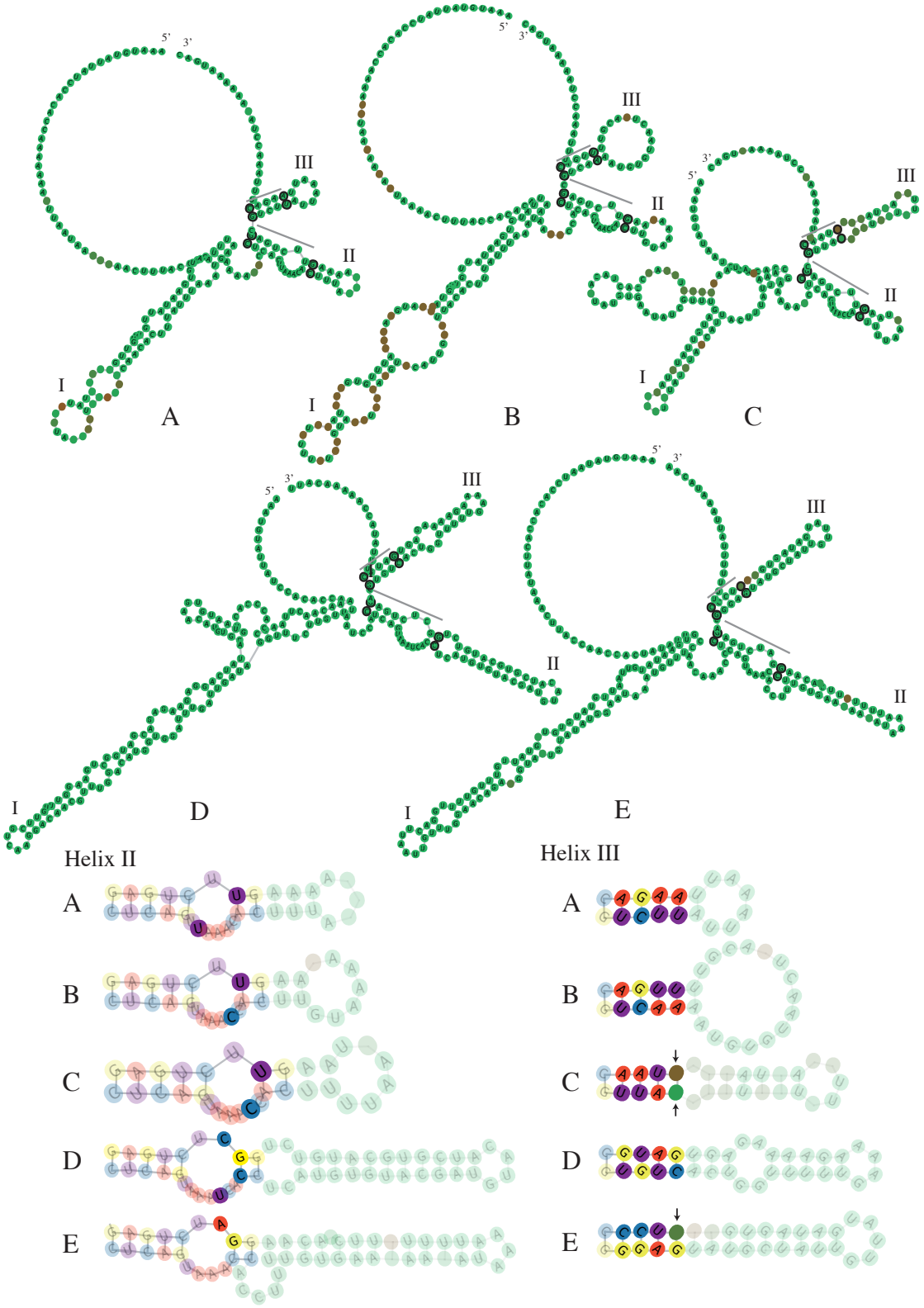


Figura 6, capítulo 4. Estructura secundaria consenso del ITS1 para cada uno de los cinco clados del MAST-4. Muestra tres hélices principales (de la I a la III). Los pares de bases altamente conservadas dentro de cada clado se muestran en verde; las posiciones variables aparecen en marrón. El detalle de las hélices II y III está representado en la parte de abajo de la figura, resaltando en colores intensos los nucleótidos conservados dentro de cada clado y que difieren entre clados. Los nucleótidos conservados en todas las secuencias aparecen en colores tenues. Las flechas apuntan a los polimorfismos (en las regiones conservadas) dentro de cada clado.

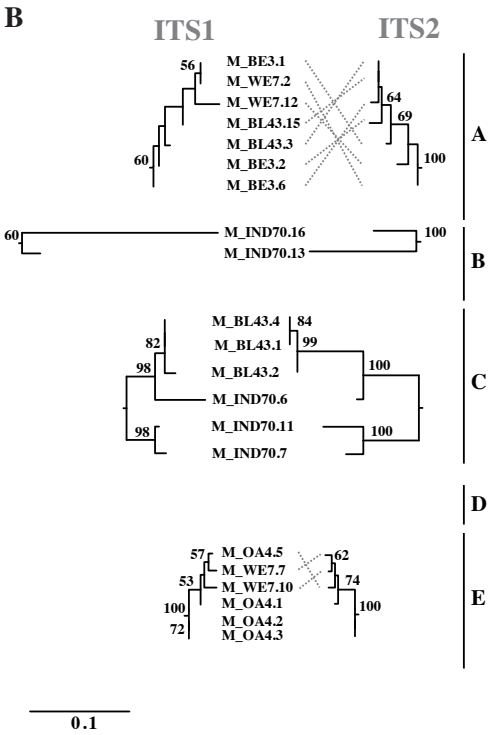


Figura 4B, capítulo 4. Árbol filogenético de máxima probabilidad de las secuencias del ITS1 e ITS2 contrastadas provenientes de 22 clones del MAST-4. La barra de escala se refiere a todos los árboles e indica sustituciones por posición. Se muestran los valores de "bootstrap" por encima de 50.

Tabla 1, capítulo 4. Divergencia de secuencias (distancia p no corregida) dentro de y entre clados del MAST-4. Los valores representan el promedio (mínimo-máximo).

	Dataset SSU-complete		Dataset SSU-LSU						
CLADE	n	**18S rDNA**	n	**18S rDNA**	**5.8S rDNA**	**28S rDNA**	**ITS1**	**ITS2**	**ITS**
A	4	0.007 (0.002-0.012)	7	0.001 (0-0.002)	0	0.005 (0.001-0.010)	0.044 (0.005-0.071)	0.036 (0.004-0.057)	0.031 (0.011-0.050)
B	3	0.008 (0.008-0.009)	2	0.001	0	0.021	0.204	0.197	0.163
C	4	0.008 (0.004-0.011)	6	0.005 (0-0.010)	0	0.020 (0-0.033)	0.104 (0.005-0.156)	0.121 (0-0.195)	0.078 (0.002-0.117)
D	2	0.011	1						
E	3	0.006 (0.002-0.008)	6	0	0	0.003 (0-0.006)	0.020 (0.004-0.032)	0.021 (0-0.031)	0.016 (0.010-0.013)
Interclade	16	0.030 (0.011-0.044)	22	0.027 (0.010-0.047)	0.024 (0-0.049)	0.161 (0.074-0.218)	0.420 (0.200-0.559)	0.418 (0.298-0.562)	0.270 (0.185-0.416)

Tabla 2, capítulo 4. Diferentes ejemplos de divergencia de secuencias del ITS (distancia p no corregida; representada como mínimo-máximo) entre cepas o especies relacionadas de cultivos de eucariotas.

		Intraclonal			Intraspecies			Interspecies			
	Species	ITS	ITS1	ITS2	ITS	ITS1	ITS2	ITS	ITS1	ITS2	Ref.[a]
Diatoms	*E. bilunaris*	0.000-0.052			0.000-0.123						1
	E. bilunaris	0.000-0.043			0.000-0.044						2
	P. multistriata				0.006	0.010	0.006				3
	P. pungens	0.000-0.070			0.000-0.044	0.000-0.050	0.000-0.064				4
	P. seriata and *P. australis*								0.036	0.027	5
	P. decipiens and *P. dolorosa*				0.000–0.005			0.105–0.108			6
	P. delicatissima and *P. decipiens*				0.000–0.049			0.075–0.090			
	P. dolorosa and *P. delicatissima*				0.000–0.002			0.129–0.151			
	Several species (5.8S+ITS2)				0.000-0.070			0.110-0.260			7
Dinoflagellates	*Symbiodinium*	0.006-0.061	0.009-0.043	0.010-0.124							8
	Several species	0.000-0.017	0.000-0.034	0.000-0.026	0.000-0.021	0.000-0.040	0.000-0.021	0.042-0.577	0.038-0.734	0.020-0.732	9
	Several species							0.000–0.014			10
	P. limbatum and *P. willei*					0.000-0.099	0.000-0.111		0.551-0.566	0.432–0.463	11
	Scrippsiella trochoidea				0.002–0.015						12
Ciliates	*Halteria grandinella*				0.001-0.082						13
Mollusca	*Haliotis*					0.000-0.049	0.000-0.044		0.380-0.590	0.380-0.480	14
Copepod	Several species						0.000-0.008			0.002-0.034	15
Magnoliophyta	Several species								0.000-0.480	0.000-0.440	16
Averages		0.001-0.049	0.005-0.039	0.005-0.075	0.000-0.042	0.002-0.050	0.001-0.042	0.077-0.200	0.201-0.481	0.144-0.363	

[a] References: 1) Vanormelingen et al. 2008, 2) Vanormelingen et al. 2007, 3) D'Alelio et al. 2008, 4) Casteleyn et al. 2008, 5) Fehling et al. 2004, 6) Lundholm et al. 2006, 7) Moniz, Kaczmarska 2009, 8) Thornhill, Lajeunesse, Santos 2007, 9) Litaker et al. 2007, 10) Logares et al. 2008, 11) Kim et al. 2004, 12) Montresor et al. 2003, 13) Katz et al. 2005, 14) Coleman, Vacquier 2002, 15) Goetze 2003, 16) Goertzen 2003

Hemos combinado estos cinco criterios para investigar la correspondencia entre clados filogenéticos y especies biológicas dentro del MAST-4. Las estructuras secundarias del ITS1 y del ITS2 se diferenciaban entre los clados mediante CBCs en las regiones conservadas, indicando que cada clado es una especie biológica diferente. Entonces, buscamos dentro de cada clado en particular. El clado A parecía estar compuesto por una única especie. No mostraba polimorfismos en las regiones críticas de los ITSs, las topologías de los árboles eran incongruentes y la divergencia de las secuencias muy baja. Además, presentaba una red MJ muy mezclada y un flujo génico alto entre los diferentes lugares. El clado B presentaba solo un motivo en las regiones conservadas de los ITSs, aunque el fragmento conservado de la hélice III del ITS2 era corto. La divergencia de las secuencias era alta, formando tres subclados filogenéticos y exhibiendo una red MJ con una estructura fuertemente separada. Por estas razones este clado podría incluir dos o tres especies. El clado C tenía una hemiCBC entre varios clones en la hélice II del ITS2 y un fragmento conservado de la hélice III corto, también presentaba cuatro motivos distintos en la hélice III del ITS1. La topología del árbol revelaba tres subclados y la red MJ presentaba una fuerte población estructurada entre los lugares. Además, la divergencia dentro del clado era similar a la divergencia media mínima entre especies. Por lo tanto, concluimos que el clado C debería incluir tres o cuatro especies. El clado D estaba escasamente muestreado y las cuatro secuencias disponibles sugerían dos especies diferentes. El clado E presentaba cuatro motivos para el ITS1, dos de ellos formando linajes bien diferenciados y los otros dos incluidos en el subclado E1, con tan solo una hemiCBC en el par de base más variable, el quinto. Las secuencias dentro de este subclado E1 no presentaban polimorfismos en las regiones conservadas del ITS2, tenían una baja divergencia de las secuencias, topologías más o menos incongruentes de los árboles del ITS y poblaciones bien mezcladas entre puntos distantes con un bajo Fst, por lo tanto este subclado parecía formar una única especie. En general, el clado E podría estar formado por tres especies. A pesar de la enorme cantidad de células del MAST-4 en los océanos, la evidencia actual después de la inspección de 250 secuencias derivadas de 5 lugares diferentes indica un máximo de 13 especies biológicas diferentes.

Baja diversificación evolutiva

La diversidad del MAST-4 se estructura en tan solo cinco clados principales, cada uno representando al menos una especie biológica y con la evidencia actual de un máximo de 13 especies **(capítulo 4** y **5)**. Este linaje se presenta como un grupo discreto bien apoyado en las filogenias del ADNr 18S, donde las secuencias emparentadas más cercanas son similares solo en un 91%. Además, la máxima divergencia entre secuencias del ADNr 18S **(capítulo 4)** dentro del MAST-4 es de 0,044, un valor muy bajo en comparación con otros grupos de protistas (Pernice M, comunicación personal). En conjunto, la divergencia genética del MAST-4 es sorprendentemente baja para un organismo tan abundante y extendido, indicando una baja diversificación evolutiva

dentro del grupo. Esto apunta a una divergencia evolutiva muy reciente con una dispersión global o un fuerte filtro ambiental que penaliza cualquier desviación del diseño óptimo de la célula. Un escenario similar de una baja diversidad y una distribución cosmopolita parece existir en otros picoeucariotas, como la prasinofícea *Micromonas* (Slapeta et al 2006).

Biogeografía

El MAST-4 aparecía presente en prácticamente todas las muestras de las aguas epipelágicas con temperaturas superiores a 5°C **(capítulo 1)**. No obstante, se ha determinado, con las secuencias recuperadas provenientes de diferentes lugares oceánicos, que el grupo MAST-4 incluye varias especies **(capítulo 4)**. La pregunta sobre la distribución de cada una de estas especies aún queda por esclarecer y el definir distribuciones específicas particulares podría también dar algunas pistas sobre la diferenciación ecotípica. El muestreo de la diversidad del MAST-4 se basó en marcadores moleculares fácilmente amplificados de muestras naturales. Esta aproximación, inicialmente basada en el ADNr 18S, también utilizó la región altamente variable del ITS1, que proporcionaba una mayor resolución taxonómica para los estudios de diversidad y biogeografía (Brown y Fuhrman 2005). La región ITS1 se ha utilizado para el estudio de variaciones de poblaciones entre- e intraespecíficas (Orsini et al 2004) o para comparar las agrupaciones a través del análisis de huella genética ARISA (análisis automatizado del espaciador interno ribosómico) (Hubbard et al 2008). En el **capítulo 5**, para estudiar la biogeografía del MAST-4, combinamos ARISA y bibliotecas de clones 18S-ITS1 y examinamos tanto los cambios espaciales y temporales de la comunidad como los factores ambientales asociados. Se confirmó que la región ITS es un buen marcador para el estudio de la biogeografía microbiana.

El análisis de los agrupamientos de MAST-4, utilizando la identificación génica del ARISA, mostró que la temperatura era el principal factor que influenciaba en los patrones de distribución (figura 6, **capítulo 5**). Esto se había observado también en otros microbios marinos como el *Prochlorococcus* (Martiny et al 2009). Los tests de Mantel entre distancias genéticas y geográficas daban correlaciones débiles, indicando que la distancia espacial no era importante para explicar la composición de la muestra. Por el contrario, la correlación comparando temperatura en lugar de distancia geográfica fue alta. En un segundo análisis, la distribución de las muestras en un espacio de 2 dimensiones (gráfico NMDS) no revelaba ninguna tendencia para las diferentes campañas (localización geográfica), en cambio sugería una clara agrupación de acuerdo con la temperatura de la muestra. Muestras alejadas que compartían la misma temperatura podían tener diversidad similar de MAST-4, mientras muestras con distintas temperaturas eran muy diferentes. Otras pruebas estadísticas confirmaron que la temperatura era el factor que explicaba la mayoría de la

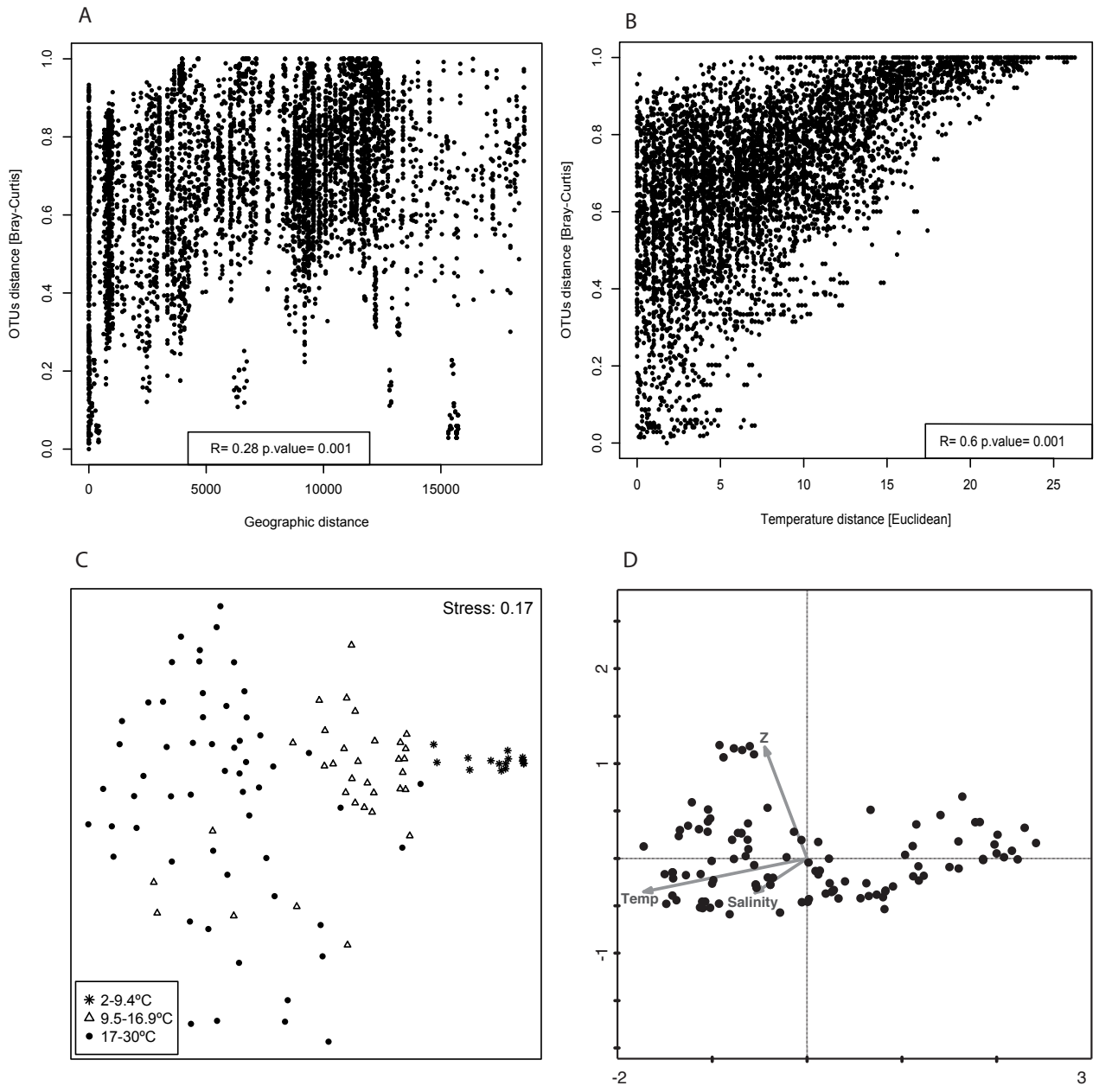


Figura 6, capítulo 5. Análisis estadísticos del conjunto de datos de la identificación genética por ARISA. (**a**) Test de Mantel que compara la composición del MAST-4 (distancia de OTUs) y la distancia geográfica. (**b**) Test de Mantel que compara la composición del MAST-4 y la distancia ambiental (temperatura). (**c**) Diagrama del escalado multidimensional no métrico (NMDS) utilizando la desemejanza Bray-Curtis de la tabla de OTUs (matriz de distancia de los OTUs). Los distintos símbolos representan los grupos de temperatura. (**d**) Diagrama del análisis de correspondencia constreñida (CCA) de la tabla de OTUs constreñida para tres de los factores más importantes: temperatura (Temp), salinidad (Sal) y profundidad del muestreo (Z). Las flechas representan la dirección y magnitud del gradiente de los factores ambientales.

varianza (CCA y PERMANOVA) y que los grupos de temperatura eran diferentes significativamente entre ellos (ANOSIM). Este patrón global fue visto en un estudio temporal de un mismo lugar (con un ciclo termal de 12-24°C) (figura 8, **capítulo 5**). Muestras de distintos años, agrupadas de nuevo por temperatura, confirmaban la importancia de este factor estructural en una escala de tiempo.

Además, el análisis del Fst reveló un gran flujo génico (dentro de los clados A y E1) entre muestras distantes con temperatura similar. Otros factores abióticos que se conoce que son importantes en definir la composición de la comunidad microbiana, como la salinidad (Logares et al 2009) y la profundidad de muestreo (Winter et al 2008), tomaron importancia dentro de cada grupo de temperatura definida. Por lo tanto, la distribución espacial del MAST-4 parecía estar principalmente controlada por factores ambientales contemporáneos con bajo o nulo grado de provincianismo.

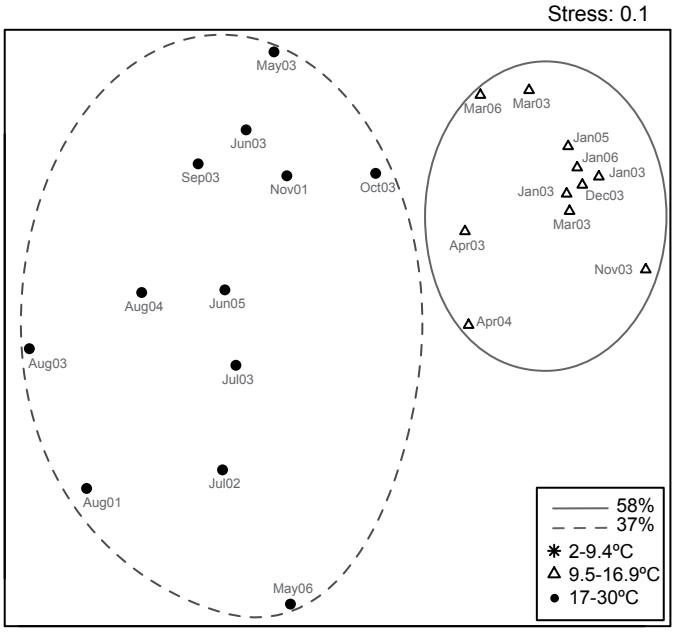


Figura 8, capítulo 5. Análisis estadístico NMDS de la huella génica ARISA en un estudio temporal de la Bahía de Blanes. Las muestras están representadas con los símbolos para la temperatura igual que en la figura 6 del capítulo 5 y con la fecha de muestreo (mes-año). Los puntos englobados por la línea discontinua se agrupan a un 37% de semejanza y los puntos englobados por la línea continua se agrupan a un 58% de semejanza en un dendrograma no mostrado.

Un análisis más detallado del clado A, que probablemente constituye una sola especie, reveló que estaba ampliamente distribuido (aparecía en las cinco bibliotecas de clones) con una buena mezcla (figura 2, **capítulo 5**). Las cinco poblaciones examinadas de este clado exhibían un alto flujo génico (figura 3, **capítulo 5**) y las muestras de diferentes lugares con la misma temperatura mostraban poblaciones completamente mixtas en la red MJ con estimas bajas de mutación por ribotipo. Un escenario similar ocurrió dentro del subclado E1, probablemente también formado por una única especie. Este subclado tiene incluso una menor proporción de mutaciones por ribotipo. Por el contrario, los clados B y C, están compuestos probablemente por más de una especie, presentan cierta estructuración espacial, con subclados que aparecen en solo una biblioteca (lugar) y con muchas más mutaciones por ribotipo entre subclados. Esta enigmática característica, que indica cierta restricción en la dispersión para los clados B y C, podría también deberse a un escaso muestreo y merece más atención en análisis futuros. Si en futuros resultados se confirman las limitaciones de dispersión de los clados B y C, éstas serían comparables a las encontradas en las cosmopolitas diatomeas del plancton marino *Pseudo-nitzschia pungens* (Casteleyn et al 2010).

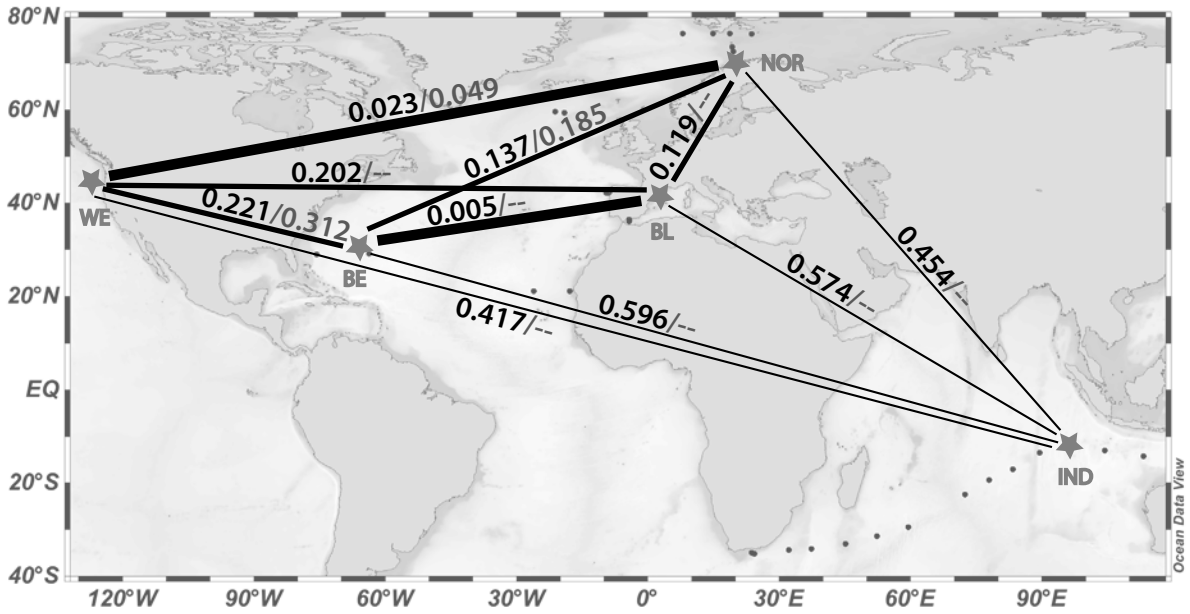


Figura 3, capítulo 5. Flujo génico propuesto entre poblaciones de los clados A y E1 pertenecientes a los cinco lugares de las bibliotecas de clones. La intensidad del flujo se indica por el grosor de la línea. Por encima de las líneas se muestran los valores de Fst para el clado A (primer valor) y el E1 (segundo valor, cuando está disponible).

Se encontraron cinco OTUs (picos de ARISA) que proporcionaban las diferencias entre muestras frías, templadas y cálidas, resaltando linajes específicos del MAST-4 que parecían adaptarse a los diferentes regímenes de temperatura. Por lo tanto, el clado B y uno o posiblemente ambos clados A y C, parecían ser característicos de aguas templadas y cálidas, mientras el clado E1, representado por el OTU 589, era el único capaz de habitar las aguas frías. La estructura genética del MAST-4 con diferentes linajes, algunos ubicuos en los océanos y con propiedades ecológicas particulares, se asemeja a la de otros picoeucariotas marinos. De este modo, diferentes linajes de *Ostreococcus* (Rodríguez et al 2005) y *Synechococcus* (Ahlgren y Rocap 2006) están adaptados a diferentes niveles de luz, mientras también ha sido descrito un clado de *Micromonas pusilla* adaptado a temperaturas frías (Lovejoy et al 2006). Se ha propuesto que esta diferenciación de ecotipos puede explicar parcialmente el éxito de los picoeucariotas, permitiéndoles explotar el espectro completo de la variabilidad de hábitats.

El protista no cultivado de vida libre MAST-4 es muy pequeño y abundante, por lo tanto posee las características para una distribución a nivel mundial (Finlay 2002). Además vive en ambientes marinos, donde el viento, las olas y las corrientes producen eventos de mezcla que facilitan la dispersión. Sin embargo, no observamos todos los clados de MAST-4 en todos los lugares, sino que se vieron patrones biogeográficos. Esto hace hincapié en la importancia de la segunda parte de un

conocido principio, "todo está en todas partes, pero el ambiente selecciona". Es razonable plantear la hipótesis de que el grupo MAST-4 tiene una capacidad de dispersión enorme y puede llegar a todas partes dentro del ambiente marino. Por ejemplo, existe un registro de una secuencia de MAST-4 en el Océano Ártico (Lovejoy y Potvin 2011), que muestra su potencial para llegar hasta tan alta latitud, probablemente arrastrado por las corrientes costeras de las aguas del Pacífico. Suponemos entonces que diferentes organismos se asentarán y crecerán dependiendo de las condiciones medioambientales, dando distintos patrones de comunidad. Esto es coherente con alguna clase de diferenciación ecológica entre los tipos relacionados, como se ha descrito para las especies de *Skeletonema* (Kooistra et al 2008). A pesar de que los microorganismos pueden dispersarse a través de todos los hábitats adecuados, adaptaciones locales eventualmente reducirían el flujo génico y promoverían la especiación (Medlin 2007). Como conclusión, el extendido protista MAST-4 es un modelo ideal para el estudio de la biogeografía microbiana, no mostrando ninguna limitación en la dispersión en sistemas marinos junto con una fuerte selección ambiental.

Conclusiones

1) El MAST-4 es un componente estructural de la parte fótica de las aguas marinas templadas. Está presente en casi todas las muestras de las aguas epipelágicas (desde la superficie hasta 120 m) con temperaturas por encima de ≈5°C.

2) La abundancia promedio de MAST-4 en sistemas de temperaturas cálidas (16-24°C) resultó similar, en un rango de 100 a 150 cél ml^{-1}. En temperaturas más frías era menos abundante.

3) Las bacterias utilizadas en esta tesis (de 0,07 a 0,18 μm^3) tenían un tamaño celular dentro del rango comestible para el MAST-4. Este organismo por lo general come de 1 a 3 bacterias por hora, por lo que se puede definir como un predador no muy voraz. Además, el MAST-4 parece que prefiere bacterias que están en buen estado fisiológico, dando unas tasas de depredación 2-3 veces mayores con las bacterias vivas que con las FLB muertas.

4) Los flagelados heterotróficos naturales tenían una Ks menor que los flagelados tradicionales de cultivos, por lo que se deduce que están bien adaptados a las abundancias bacterianas típicas de los ambientes marinos planctónicos. En concreto, el MAST-4 presentaba una Ks de 8,7 10^5 bacterias ml^{-1}.

5) La diversidad funcional observada entre el MAST-4 y otros flagelados heterotróficos (como el MAST-1C, "*Candidatus* M. minuta" y *Paraphysomonas*) da un significado ecológico a la alta diversidad filogenética de los protistas heterotróficos marinos, con taxones diferentes adaptados a distintos nichos ecológicos.

6) A pesar de la presencia de un alto número de células del MAST-4 en los océanos su diversidad está estructurada en tan solo cinco clados principales, cada uno representando al menos una especie biológica. La evidencia actual indica un máximo de 13 especies separadas.

7) La divergencia genética del MAST-4 resultó sorprendentemente baja para un organismo tan abundante y extendido, indicando una bajísima diversificación evolutiva, lo que apunta, bien a una divergencia evolutiva muy reciente o a un fuerte filtro medioambiental que penalice cualquier desviación del diseño óptimo celular.

8) No se observaron barreras geográficas para la dispersión de los clados más representados del MAST-4, mientras que se vio que la temperatura era el factor principal que influenciaba en los patrones de distribución.

9) El ambiente parecía hacer una selección taxonómica, puesto que diferentes clados del MAST-4, con supuestas adaptaciones fisiológicas, se establecían en ambientes con diferentes condiciones. En particular, el clado E1, representado por el OTU 589, resultó ser el único aclimatado a las aguas frías. Esta diferenciación de ecotipos podría explicar en parte el éxito de los picoeucariotas, permitiéndoles explotar todo el espectro de la variabilidad del hábitat.

Perspectivas de futuro

1) A lo largo de esta tesis hemos visto la importancia del linaje MAST-4. Por lo tanto existe una necesidad urgente de conseguirlo en cultivo. Creemos que esto es factible, especialmente teniendo en cuenta la ventaja de tener nuevos datos recogidos sobre este microbio no cultivado, como sus hábitos alimenticios, las limitaciones ambientales y las sondas optimizadas para una rápida detección molecular.

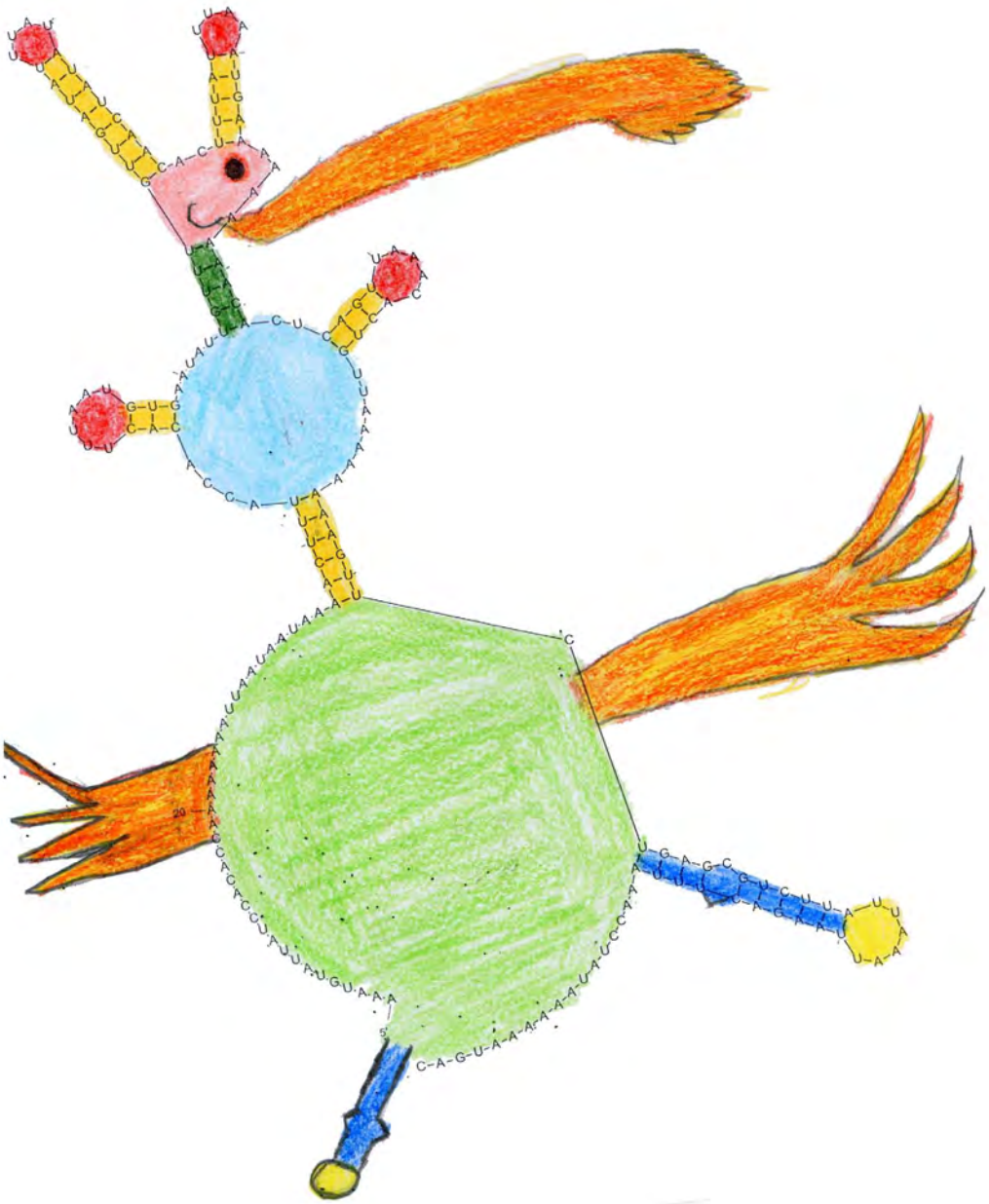
2) Los clados A y E proporcionan una visión general de los patrones biogeográficos, que se rige por la temperatura en lugar de la distancia geográfica. Sin embargo, los otros clados muestran resultados menos claros, debido probablemente a una falta de muestreo. Sería muy interesante completar este estudio añadiendo muchas más secuencias de diferentes lugares. Esto podría hacerse explotando los iniciadores y protocolos de PCR diseñados aquí para su uso en la secuenciación masiva, como por ejemplo la pirosecuenciación. Un incremento en el número de secuencias perfeccionaría la genética de poblaciones de este linaje de protistas.

3) De manera similar a las regiones conservadas del ITS2, hemos propuesto una región conservada para la delimitación de especies biológicas en la estructura secundaria del ITS1. Sería interesante corroborarlo con experimentos de reproducción.

4) La genómica de amplificación de aislados es un enfoque prometedor para obtener el contenido genómico de estos microorganismos no cultivados y para identificar interacciones biológicas específicas.

5) Sería interesante integrar el MAST-4 en las redes tróficas microbianas y conocer la cantidad de carbono que están procesando y, todavía más intrigante, cuáles son sus factores de mortalidad (probablemente la depredación y quizás la infección viral).

General References



Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR *et al* (2005). The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists. *J Eukaryot Microbiol* **52:** 399-451.

Ahlgren NA, Rocap G (2006). Culture Isolation and Culture-Independent Clone Libraries Reveal New Marine Synechococcus Ecotypes with Distinctive Light and N Physiologies. *Appl Environ Microbiol* **72:** 7193-7204.

Amann R, Ludwig W, Schleifer K (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143-169.

Amaral-Zettler LA, Mccliment EA, Ducklow HW, Huse SM (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE* **4:** e6372.

Amato A, Kooistra W, Levialdi Ghiron J, Mann D, Pröschold T, Montresor M (2007). Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* **158:** 193-207.

Andersen RA, Sounders GW, Paskind MP, Sexton JP (1993). Ultrastructure And 18S rRNA Gene Sequence For Pelagomonas Calceolata Gen. Et Sp. Nov. And The Description Of A New Algal Class, The Pelagophyceae Classis Nov. *J Phycol* **29:** 701-715.

Archibald JM, Keeling PJ (2005). On the origin and evolution of plastids. *In: Microbial Evolution: Concepts and Controversies* Sapp, J (ed) Oxford University Press, Oxford**:** 238-260.

Atkinson D, Ciotti BJ, Montagnes DJS (2003). Protists decrease in size linearly with temperature: ca. 2.5% ºC^{-1}. *P Roy Soc Lond B Bio* **270:** 2605.

Baldauf SL (2003). The deep roots of eukaryotes. *Science* **300:** 1703-1706.

Baldauf SL (2008). An overview of the phylogeny and diversity of eukaryotes. *J Syst Evol* **46:** 263-273.

Bandelt H, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16:** 37.

Boenigk J, Jost S, Stoeck T, Garstecki T (2001). The problematic nature of fluorescently labeled bacteria (FLB) in Spumella feeding experiments-an explanation by using video microscopy. *Arch Hydrobiol* **152:** 329-338.

Brown M, Fuhrman J (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41:** 15-23.

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.

Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS *et al* (2007). Phylogenomics Reshuffles the Eukaryotic Supergroups. *PLoS ONE* **2:** e790.

Burki F, Inagaki Y, Brate J, Archibald JM, Keeling PJ, Cavalier-Smith T *et al* (2009). Large-Scale Phylogenomic Analyses Reveal That Two Enigmatic Protist Lineages, Telonemia and Centroheliozoa, Are Related to Photosynthetic Chromalveolates. *Genome Biol Evol* **2009:** 231-238.

Casteleyn G, Leliaert F, Backeljau T, Debeer A, Kotaki Y, Rhodes L *et al* (2010). Limits to gene flow in a cosmopolitan marine planktonic diatom. *Proc Natl Acad Sci USA* **107:** 12952-12957.

Cavalier-Smith T, Allsopp MTEP, Chao EE (1994). Thraustochytrids are chromists, not Fungi: 18s rRNA signatures of Heterokonta. *Phil Trans R Soc Lond B* **346:** 387-397.

Cavalier-Smith T (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Micr* **52:** 297-354.

Cavalier-Smith T, Chao EE-Y (2003). Phylogeny of Choanozoa, Apusozoa, and Other Protozoa and Early Eukaryote Megaevolution. *J Mol Evol* **56:** 540-563.

Coleman A (2003). ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *TRENDS Genet* **19:** 370-375.

Coleman A (2007). Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res*.

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

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:** 1219-1232.

del Campo J (2011). Culturing bias in the study of marine heterotrophic flagellates diversity. *PhD thesis University of Barcelona*.

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

Falkowski PG, Fenchel T, Delong EF (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320:** 1034-1039.

Fehling J, Stoecker DK, Baldauf SL (2007). Photosynthesis and the eukaryote tree of life. In: Falkowski PG, Knoll AH eds *Evolution of primary producers in the sea* New York: Academic Press**:** 75-107.

Fenchel T (1987). Ecology of Protozoa. The Biology of Free-living Phagotrophic Protists. *Brock/ Springer Series in Contemporary Bioscience* **47 figs.:** 197pp.

Fenchel T (1988). Marine plankton food chains. *Annu Rev Ecol Syst* **19:** 19-38.

Finlay BJ (2002). Global Dispersal of Free-Living Microbial Eukaryote Species. *Science* **296:** 1061-1063.

Fisher MM, Triplett EW (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65:** 4630-4636.

Fu Y, O'Kelly C, Sieracki M, Distel DL (2003). Protistan grazing analysis by flow cytometry using prey labeled by in vivo expression of fluorescent proteins. *Appl Environ Microbiol* **69:** 6848-6855.

Fuhrman JA, Hagström A (2008). Bacteria and archaeal community structure and its patterns. In Kirchman DL (ed) *Microbial Ecology of the Oceans* Wiley Series**:** 45-80.

Fukuda H, Sohrin R, Nagata T, Koike I (2007). Size distribution and biomass of nanoflagellates in meso-and bathypelagic layers of the subarctic Pacific. *Aquat Microb Ecol* **46:** 203-207.

Fuller NJ, Campbell C, Allen DJ, Pitt FD, Zwirglmaier K, Gall FL *et al* (2006). Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using PCR biased towards marine algal plastids. *Aquat Microb Ecol* **43:** 79-93.

González J, Sherr EB, Sherr BF (1990). Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. *Appl Environ Microbiol* **56:** 583-589.

González JM, Sherr EB, Sherr BF (1993). Differential feeding by marine flagellates on growing versus starving, and on motile versus nonmotile, bacterial prey. *Mar Ecol Prog Ser* **102:** 257-267.

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D (2007). Phylogenomic Analysis Supports the Monophyly of Cryptophytes and Haptophytes and the Association of Rhizaria with Chromalveolates. *Mol Biol Evol* **24:** 1702-1713.

Head I, Saunders J, Pickup R (1998). Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb Ecol* **35:** 1-21.

Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R (2011). Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J* **5:** 674-684.

Holling CS (1959). Some Characteristics of Simple Types of Predation and Parasitism. *Can Entomol* **91:** 385-398.

Honda D, Yokochi T, Nakahara T, Raghukumar S, Nakagiri A, Schaumann K *et al* (1999). Molecular phylogeny of labyrinthulids and thraustochytrids based on the sequencing of 18S ribosomal RNA gene. *J Eukaryot Microbiol* **46:** 637-647.

Hubbard K, Rocap G, Armbrust E (2008). Inter- and intraspecific community structure within the diatom genus Pseudo-nitzschia (Bacillariophyceae). *J Phycol* **44:** 637-649.

Jürgens K, Matz C (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *A Van Leeuw* **81:** 413-434.

Jürgens K, Massana R (2008). Protist grazing on marine bacterioplankton. In: D.L. Kirchman (ed.) *Microbial ecology of the Oceans. 2nd edn***:** 383-441.

Kooistra WHCF, Sarno D, Balzano S, Gu H, Andersen RA, Zingone A (2008). Global Diversity and Biogeography of Skeletonema Species (Bacillariophyta). *Protist* **159:** 177-193.

Landry M, Lehner-Fournier J, Sundstrom J, Fagerness V, Selph K (1991). Discrimination between living and heat-killed prey by a marine zooflagellate, Paraphysomonas vestita (Stokes). *J Exp Mar Biol Ecol* **146:** 139-151.

Lee WJ, Patterson DJ (1998). Diversity and Geographic Distribution of Free-Living Heterotrophic Flagellates - Analysis by PRIMER. *Protist* **149:** 229-244.

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

Lim E, Dennett M, Caron D (1999). The ecology of Paraphysomonas imperforata based on studies employing oligonucleotide probe identification in coastal water samples and enrichment cultures. *Limnol Oceanogr* **44:** 37-51.

Lim EL (1996). Molecular identification of nanoplanktonic protists based on small subunit ribosomal RNA gene sequences for ecological studies. *J Eukaryot Microbiol* **43:** 101-106.

Logares R, Bråte 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, Pedrós-Alió C, Moreira D (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409:** 603-607.

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

Lovejoy C, Potvin M (2011). Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. *J Plankton Res* **3:** 431-444.

Margulis L, Chapman M, Guerrero R, Hall J (2006). The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proc Natl Acad Sci USA* **103:** 13080-13085.

Martínez-García M, Brazel D, Poulton N, B Swan, M Gómez, D Masland *et al* (2012). Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. *J ISME* **6:** 703-707.

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

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

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

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

Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A *et al* (2004b). Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol* **70:** 3528-3534.

Massana R, Guillou L, Terrado R, Forn I, Pedrós-Alió C (2006a). Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol* **45:** 171-180.

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

Massana R (2011). Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65:** 1-47.

Medlin L (2007). If everything is everywhere, do they share a common gene pool? *Gene***:** 1-4.

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.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806-810.

Mullineux T, Hausner G (2009). Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera Grosmannia and Leptographium. *Fungal Genet Biol* **46:** 855-867.

Not F, Romari K, Lovejoy C, Massana R, Toebe K, Vaulot D *et al* (2007). Picobiliphytes: A marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315:** 252-254.

Okamoto N, Inouye I (2005). A secondary symbiosis in progress? *Science* **310:** 287.

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.

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

Patterson DJ, Larsen J (1991). The Biology of Free-Living Heterotrophic Flagellates. *Clarendon Press, Oxford*.

Patterson DJ, Zöffel M (1991). Heterotrophic flagellates of uncertain taxonomic position. In D J Patterson and J Larsen (eds), *The Biology of Free-Living Heterotrophic Flagellates Clarendon Press***:** 427-475.

Pernthaler J, Glöckner F, Schönhuber W, Amann R (2001). Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. *Meth Microb* **30:** 207-226.

Pernthaler J (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* **3:** 537-546.

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

Poirazidis K, Schindler S, Kakalis E, Ruiz C, Bakaloudis D, Scandolara C *et al* (2011). Population estimates for the diverse raptor assemblage of Dadia Natinal Park, Greece. *Ardeola* **58:** 3-17.

Porter K, Feig Y (1980). The use of DAPI for identifying and counting aquatic microflora. *Limnol Oceanogr* **25:** 943-948.

Richards TA, Vepritskiy AA, Gouliamova DE, Nierzwicki-Bauer SA (2005). The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. *Environ Microbiol* **7:** 1413-1425.

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

Rose JM, Caron DA, Sieracki ME, Poulton N (2004). Counting heterotrophic nanoplanktonic protists in cultures and aquatic communities by flow cytometry. *Aquat Microb Ecol* **34:** 263-277.

Scheckenbach F, Wylezih C, Weitere M, Hausmann K, Arndt H (2005). Molecular identity of strains of heterotrophic flagellates isolated from surface waters and deep-sea sediments of the South Atlantic based on SSU rDNA. *Aquat Microb Ecol* **38:** 239-247.

Schindler S, Vasilakis D, Poirazidis K (2006). Error Assessment of a Telemetry System for Eurasian Black Vultures (Aegypius monachus). *International Conference on Sustainable Management and Development of Mountainous and Island Areas***:** 305-314.

Selje N, Simon M, Brinkhoff T (2004). A newly discovered Roseobacter cluster in temperate and polar oceans. *Nature* **427:** 445-448.

Shalchian-Tabrizi K, Eikrem W, Klaveness D, Vaulot D, Minge MA, Le Gall F *et al* (2006). Telonemia, a new protist phylum with affinity to chromist lineages. *P Roy Soc Lond B Bio* **273:** 1833-1842.

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

Shi XL, Lepère C, Scanlan DJ, Vaulot D (2011). Plastid 16S rRNA Gene Diversity among Eukaryotic Picophytoplankton Sorted by Flow Cytometry from the South Pacific Ocean. *PLoS ONE* **6:** e18979.

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.

Stechmann A, Hamblin K, Pérez-Brocal V, Gaston D, Richmond GS, van der Giezen M *et al* (2008). Organelles in Blastocystis that blur the distinction between mitochondria and hydrogenosomes. *Curr Biol* **18:** 580-585.

Stoeck T, Hayward B, Taylor GT, Varela R, Epstein SS (2006). A Multiple PCR-primer Approach to Access the Microeukaryotic Diversity in Environmental Samples. *Protist* **157:** 31-43.

Tanaka T, Rassoulzadegan F (2002). Full-depth profile (0-2000 m) of bacteria, heterotrophic nanoflagellates and ciliates in the NW Mediterranean Sea: Vertical partitioning of microbial trophic structures. *Deep-Sea Res Pt II* **49:** 2093-2107.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS *et al* (2000). Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol* **31:** 21-32.

Thornhill DJ, Lajeunesse TC, Santos SR (2007). Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. *Mol Ecol* **16:** 5326-5340.

Unrein F, Massana R, Alonso-Sáez L, Gasol J (2007). Significant year-round effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal system. *Limnol Oceanogr* **52:** 456-469.

Vaqué D, Gasol JM, Marrasé C (1994). Grazing rates on bacteria: the significance of methodology and ecological factors. *Mar Ecol Prog Ser* **109:** 263-274.

Vasilakis D, Poirazidis K, Elorriaga J (2005). Breeding season range use of a Eurasian Black Vulture

(Aegypius monachus) population in the Dadia National Park and the adjacent regions. *Proceedings of the International Congress on the Conservation and Management of Vulture Populations***:** 127-137.

Vasilakis D, Poirazidis K, Elorriaga J (2008). Range use of a Eurasian black vulture (Aegypius monachus) population in the Dadia-Lefkimi-Soufli National Park and the adjacent areas, Thrace, NE Greece. *J Nat Hist* **42:** 355-373.

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

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

Vigil P, Countway P, Rose J, Lonsdale D, Gobler C, Caron D (2009). Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. *Aquat Microb Ecol* **54:** 83-100.

White TJ, Arnheim N, Erlich HA (1989). The polymerase chain reaction. *Trends Genet* **5:** 185-189.

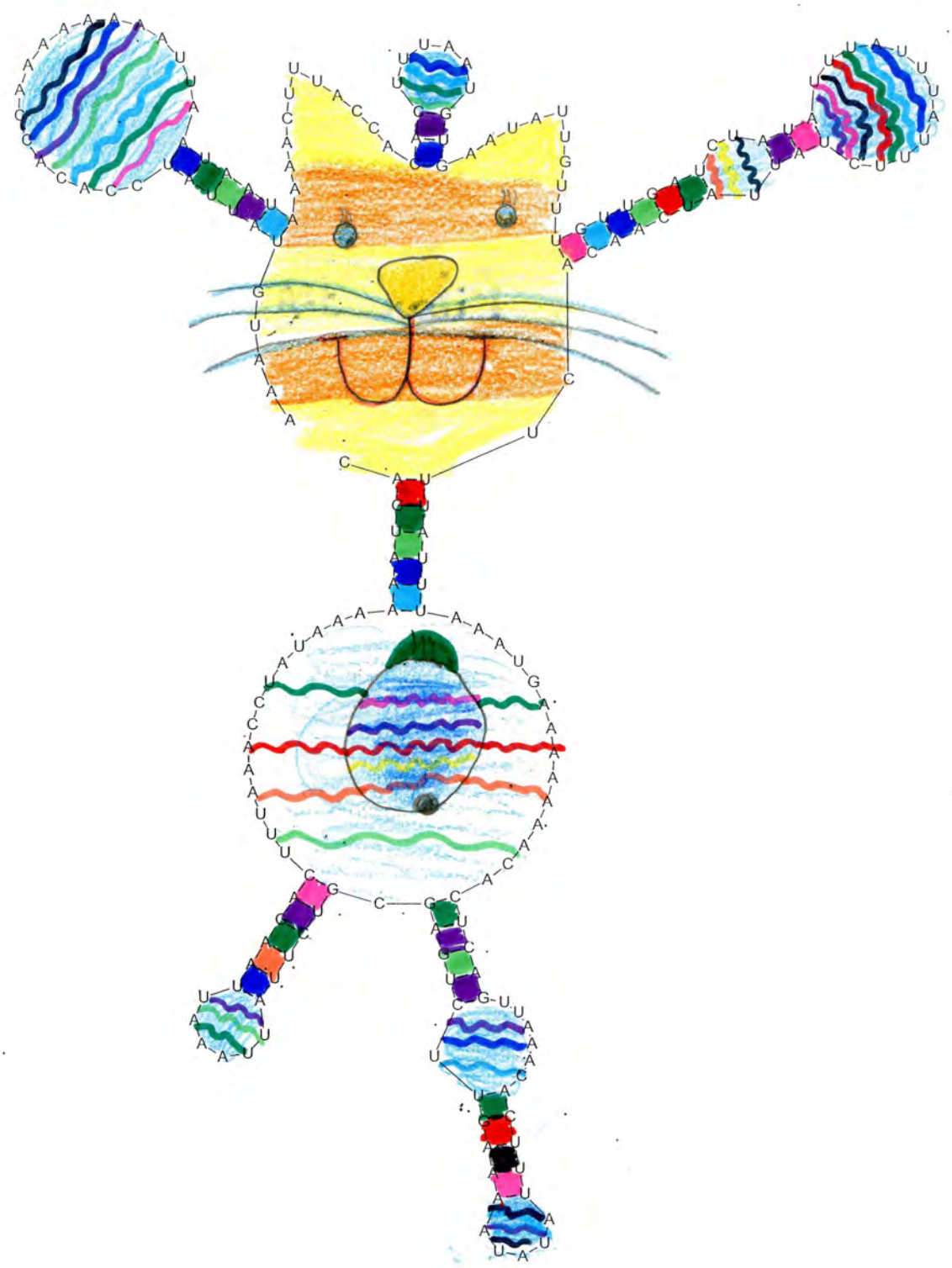
Winter C, Moeseneder MM, Herndl GJ, Weinbauer MG (2008). Relationship of Geographic Distance, Depth, Temperature, and Viruses with Prokaryotic Communities in the Eastern Tropical Atlantic Ocean. *Microb Ecol* **56:** 383-389.

Woese CR (1987). Bacterial evolution. *Microbiol Rev* **51:** 221-271.

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

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

Acknowledgements



"Señores viajeros, bienvenidos a bordo del tren con destino Barcelona-Sants. Durante el recorrido les anunciaremos la llegada a las estaciones con parada..."

De nuevo en el tren, arrastro sueño y cansancio de tantas noches en vela con las últimas correcciones y retoques que parecen interminables; pero también alegría, vuelvo de visitar mi ciudad natal con energía y motivación suficiente para acabar el doctorado (eso que llaman Doctor Philosophiae). Y es ahora, en este momento de meditación, en este pequeño oasis de relajación de la recta final, cuando al mirar por la ventanilla y ver el paisaje pasar, pienso en todas estas personas que están o han estado a mi lado, cada una compartiendo, contribuyendo y enriqueciendo este recorrido, sin las cuales alcanzar el destino final no hubiera sido posible. Y es por esto que me gustaría agradecerles a todas y cada una de ellas su aportación en este viaje. Será difícil no dejarme a alguna en el tintero, pero eso querrá decir que realmente ha sido un gran viaje repleto de gente y emociones. Por lo tanto a todas estas personas, aquí mencionadas o ausentes... GRACIAS!

La primera gran decisión en este viaje fue elegir el destino, y ahí tengo que agradecer a mi director de tesis, Ramon Massana, que siempre me supo transmitir su gran motivación y pasión por los flagelados heterotróficos no cultivados, unos tales estramenópilos marinos. Desde el primer día, la primera entrevista, allá en su despacho, en el último rincón perdido del ICM, tras sus gafas de pasta, me explicó su proyecto y consiguió que eligiera el rumbo al mundo de los MAST. Un viaje que realmente ha merecido la pena y para el cual siempre me ha estado apoyando, enseñando, encarrilando cuando me desviaba demasiado de mi camino y confiando en que llegara a la siguiente estación cuando me perdía de vista. Gracias por toda tu ayuda en este último empujón para conseguir acabar. Ramon, moltíssimes gràcies!

Conseguí el billete, un billete para cuatro años de viaje, el cual al principio parece interminable, pero al final te encuentras haciendo autostop. En este momento debería nombrar a mi familia. A mis padres, porque sé que siempre les cuesta verme partir, pero aún así apoyan mis decisiones y siempre puedo contar con ellos, para lo bueno o para lo malo. Por sus sabios consejos que, aunque a veces desobedezca, siempre me resultan muy útiles para formarme y aprender a "vivir la vida" con alegría y motivación. A mis hermanos: Francis, al que nunca le harás enfadar, un sabio en las nubes que nos va enseñando el camino; y Mario, tan filosófico él y siempre dispuesto a escuchar y ayudar. A Olga, por ser tan observadora, discreta y detallista. A "la yaya", por su buen sentido del humor, sus consejos realmente prácticos para la vida y su preocupación por nuestro bienestar: si comemos, pasamos frío... A mi abuela María, por esas comidas tan exquisitas que nos preparabas y las labores tan minuciosas que hacías que me han hecho aprender a cuidar los detalles. A mis tíos y primos: Javier, Conchi, Javier, Rocío, Mario, Marta, Álvaro, Adrián y, en especial, a mi tía María Jesús, porque sé que te hubiera encantado poder ver esta tesis acabada, porque siempre me has valorado tanto y porque te echo de menos. A mi sobrino Víctor, que desde que naciste fuiste una gran alegría y que tus carcajadas contagiosas por las cosas más banales son la terapia perfecta para reponerse del doctorado en cualquier momento. Porque es un placer verte crecer.

Te subes al tren, las puertas se cierran y comienza a moverse. Buscas tu asiento, confiando en que no te afecte el overbooking y es entonces cuando te encuentras rodeado de gente que con el tiempo irás conociendo. Gracias a mis compañeros de vagón, a los que siempre se han sentado cerca de mí, porque han sido muy importantes en este viaje. A Ero, por esos paseos curativos de vuelta a casa después de una larga jornada laboral, por tu bondad y comprensión, porque hemos viajado codo con codo durante mucho tiempo. A Montse, pel teu doble sentit de l'humor, per aquelles converses divertides en moments distesos, per la teva passió. Y al resto de compañeros de despacho que han hecho de nuestra pequeña cueva un lugar más habitable: a Eli Sañé, por tu generosidad, a Ruth, a Sesé, a la reciente incorporada Clara R, llegando en el momento oportuno. A Eli Sá, por nuestros días eli-raki, por tus sabios consejos, por esas conversaciones de "regreso" que solo el cierre del metro consigue cortar, por tu ayuda y esfuerzo en las correcciones finales. A Bea, por tu energía y vitalidad, por lo fácil y divertido que es viajar contigo y porque esta histeria del final siempre se lleva mejor compartida. A Massimo, comenzaste el viaje más tarde, pero me has apoyado tanto que tengo la sensación de que siempre estuviste a nuestro lado. Gracias por esos cafés al sol y porque eres un artista. A Lorena, mi asesora de imagen principal y la iniciadora en el descubrimiento del mundo chileno-argentino, gracias por ayudarme a crear nuestro hogar y por tantos buenos momentos compartidos; tenemos un viaje pendiente. A Pedro, por tus historias fascinantes y tu conocimiento de los peligros de cualquier parte del mundo. A Juanlu por tu apoyo en las crisis computacionales y tu punto de vista masculino.

Primera parada: Canarias. ¿Qué mejor que conocer la universidad a la que estarás vinculado y sus alrededores? Otra manera de enfocar los habitualmente pesados cursos de doctorado. Gracias a todos los profesores del programa de Oceanografía de la ULPGC y a todo el personal que está facilitándonos todos estos tediosos trámites del depósito.

Coincides con viajeros que llegan a su fin, que no puedes compartir demasiado con ellos porque se van, pero que te enseñan, te animan, te ayudan. Gracias a Laura A, Imelda, Maria V, Itziar L, Andrés, Dacha. Y con el paso del tiempo suben los nuevos, culpables en parte de no querer abandonar nunca el tren. Gracias a Ana Mari y Fran, por ese arte que os caracteriza. A Guillem, por tus enseñanzas en "R". A Amanda, esas habitas del mosquito nos esperan. A Caterina R, la sucesión Rodríguez en el mundo de los MAST, los dejo en buenas manos. A Sarah-Jeanne, a Roy, a Fran A. En el momento de estirar las piernas, encuentras a compañeros de viaje que aunque en vagones más alejados siempre aparecen por los pasillos, en la cafetería, con conversaciones, comentarios, ideas, informaciones... siempre de gran utilidad. Gracias a Julia, Rodrigo A, Javier, Bego, Rachele, Juancho, Arancha, Clara R, Cristina R, Sara Z, Pati, Mireia, Martí, Cristina D, Ana G, Elena, María P, Vero, Eva F, Estela, Silvia Anglés, Albert R, Sonia, Rodrigo M y Gemma V. A Dan, por ese cochinillo pendiente. A Mariona y Thomas L, partícipes también de esta histeria colectiva del final. ¡Qué ganas de celebrar!

Siguiente parada: Warnemünde. Un pueblecito perdido bañado por el Báltico, al Norte de Alemania. El mejor lugar de trabajo, meditación y reencuentro con uno mismo sobre todo si te toca visitarlo en los meses de invierno. Gracias Fer, por tu compañía y por tus enseñanzas para un principiante de no dejar que

la ciencia absorba por completo tu vida personal. Ahora cocino, cultivo, bailo y me divierto entre paper y paper. Thanks Klaus Jürgens, to hosting me in your lab; Matthias Labrenz, to teach me so well about the Q-PCR and to all the people in the lab. Thanks to the "family", Rosa, Jonas, Alar, you made it an amazing time!

Compartiendo viaje, en algún que otro momento del recorrido aparecen los viajeros con experiencia, Postdocs, "Ramones y Cajales", jóvenes investigadores... Ya han superado ese viaje inicial de inexpertos y siempre he aprendido mucho de ellos, de sus sabios consejos. Gracias a Fabrice, por ese mano a mano en los experimentos de grazing, a Ramiro, por culturizarme en la estadística y otras cosas útiles de la vida, a Silvia Acinas, sí sí, la del Nature!, que lujo me pareció poder hablar contigo el primer día que te encontré, gracias a ti y a Isabel por vuestros grandes consejos y buenos ratos, a Hugo, por esas discusiones filosóficas del comedor, a Bea D, José M, Bibiana, Marta S, Ivo, Thomas P, Evaristo y Katherina.

Tercera parada: el Ártico. Una increíble oportunidad en mi vida. Gracias Ramon y Carles por hacer que fuera posible. Y cómo no, agradecer a todos los compañeros de campaña y a la tripulación, a Ramon T, Stephane, Loïc, Stelly, Myriam, Eve, Sebastian... a los jefes de campaña Tim y Jody y a los cocineros, por esos suculentos gâteaux personalizados.

En este viaje es importante no olvidarme de los componentes principales, los que hacen que todo pueda funcionar, que han creado un entramado de vías de tal dimensión que uno pasaría la vida entera recorriéndolas todas. Gracias a Carles Pedrós, Cèlia Marrasé, Pep Gasol, Rafel Simó, Montse Sala, Dolors Vaqué, Marta Ribes, Marta Estrada, Elisa Berdalet, Albert Calbet, Enric Sainz, Miquel Alcaraz, Miquel Latasa, Emilio Casamayor, José González, gracias porque siempre estáis dispuestos a recibirme para aclarar cualquier tipo de duda, por vuestra cercanía, así es más fácil empaparse de tanta sabiduría. Y sin duda, también de gran importancia, los técnicos que hacen posible la construcción y el buen funcionamiento del sistema. Gracias a Irene, por tu eficacia y tu gran ayuda con el FISH, porque trabajar contigo hace que todo sea fácil y divertido. A Vanessa y a Clara C, por toda vuestra ayuda siempre conservando la calma y el buen humor. A Paula, a Eli Sá, ¿cultivaré contigo?. E indispensable mencionar en este mantenimiento, a Xavi Leal, Alex Mir, Conchita, y todo el servicio informático, de administración y limpieza, que acuden en llamada de auxilio en los momentos críticos. También agradecer al servicio de seguridad, a Xavi, por preocuparse de cuándo me tiene que abrir la puerta y por recordarme que hay que salir del despacho cuando se me hace tarde.

Última parada antes de la llegada al destino final: Seattle. Gracias Marcela por tu gran acogida, por tu kit de supervivencia-recién-llegada y por esos buenos ratos que hemos pasado juntas. Thanks Gabrielle for hosting me in your lab and for teach me and discuss with me all the ITS world. I admire your big ability to change your mind so fast from one to other student work. Thanks to all the people in the lab, April and Kate, for worry about me. Thanks to Billy, Clara, Marcela, Rika, Eric, Gavin, Karen, Min for sharing with me the Seattle life. To the people in the house, Nivin, Scott, Liang, or the adopted Rossking. Thanks Eric,

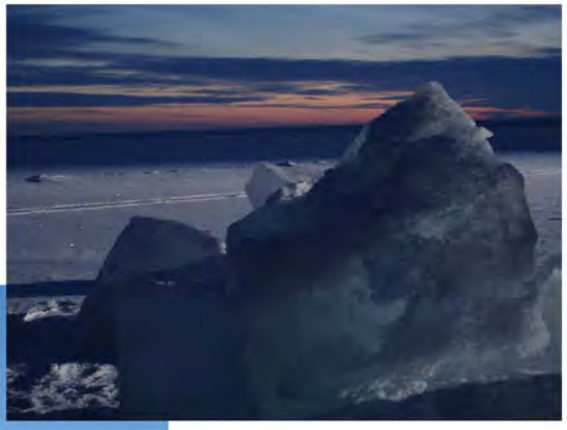
for providing me with essential tools in Seattle, as a bike or a beer glass, for driving me to so amazing places and music concerts and for all your help, affection and good moments shared.

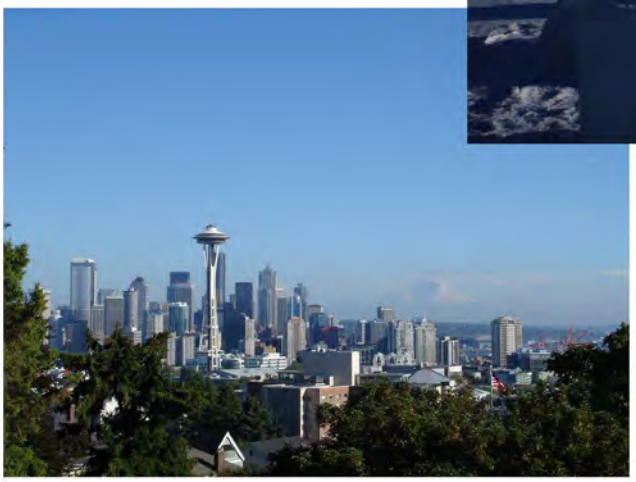
Gracias a toda la gente que, aunque esté más lejos, siempre ha estado siguiendo mi viaje y apoyándome. A los de Logroño, casilla de salida por la cual siempre es necesario pasar de vez en cuando. A los de León, donde me formé como Bióloga en muy buena compañía. Gracias a Esther M, Elvira, Manolo, Aitor, Raúl, Sofía, Mazu, Javi B, José, Bea, Sergio, Ana, Eva, Javi, Marcos, Cristina. ¿Esta noche Laurel?. A Henar, Reyes, Natalie, Mary, Lidia, Elvira, Esther R, Itziar F, Ana I, por esas quedadas que seguimos haciendo año tras año y que me encantan. A Gelo, Iván, Irene, Ricardo, Alberto. A Tito, siempre es un placer oírte cantar, gracias por tus esfuerzos en comprender qué era eso de los estramenópilos. A mis vecinos, casi familia y los "del monte", seguidores de mis crónicas de viaje. A todos con los que trabajé en Grecia, meditación previa al comienzo del doctorado, a Javi, Stefan, Carlos, Myriam, Lise, Dora, Dimitris, Poira...

Entre viaje y viaje siempre necesitas momentos de evasión, de recuperación de energía, gracias a todo el equipo Bollywood, a Pinky, por tus grandes coreografías y a las Dilses: Aglaia, Sara, Silvia, Bea, Eli, Olga, Rebekka, Jenny, Selma, Mont, Valerie y Marta; por tantos ensayos y actuaciones, por esos nervios y risas por detrás del escenario. Y también necesitas un hogar reponedor, gracias a todos los habitantes del piso de Av Gaudi, a Alba, ¡qué buenos tiempos aquellos!, a Marco, Simó, Gloria, Naiara, Patricia y Antonio, ¡cómo cambia la vida!, a Lorena, a Gonzalo ¡qué buenas esas empanadas Argentinas! y al dúo chileno: Alejandro y Sergio, gracias por ayudarme y comprenderme en estos momentos de "locura extrema".

Y después de todo, el viaje llegará a su fin, solventando atascos, retrasos, altas velocidades, temporales, pero lleno de alegrías, buenas experiencias y grandes recuerdos. Porque sin vosotros no hubiera sido posible.

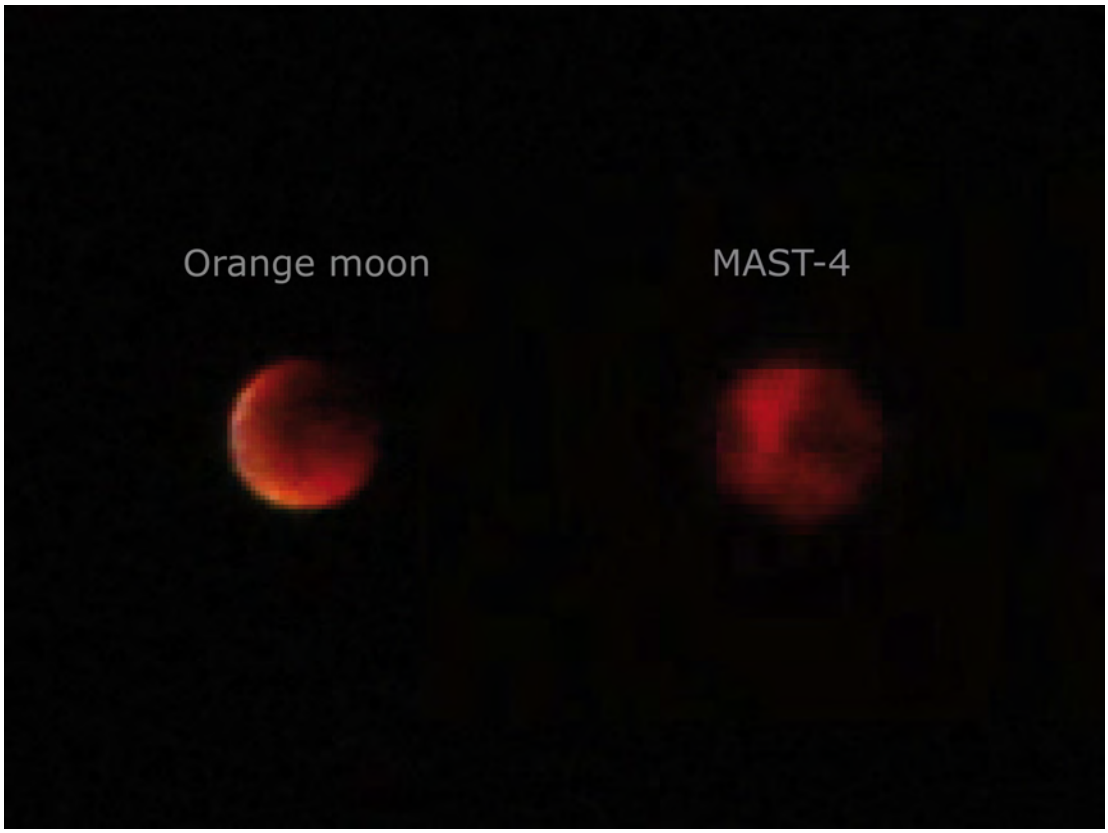
¡GRACIAS!
GRÀCIES!
THANK YOU!
MERCIE!
DANKE SHÖN!
DHANYAVAAD!





“Nature, [...], cares nothing for appearances, except in so far as they are useful to any being.”

The origin of species. Charles Darwin



Las ilustraciones de los capítulos han sido dibujadas, a partir de pruebas de estructuras secundarias del ITS2, por niños menores de 10 años y ancianos mayores de 90: Adela, Adrián, Álvaro, Martina, Alejandra, Sofía, Lucía y los alumnos de 3ºA de primaria del colegio Villa y tierra de Saldaña (Julia, Pilar, Jorge, Carla, Javi, Ibrahim y Vero).