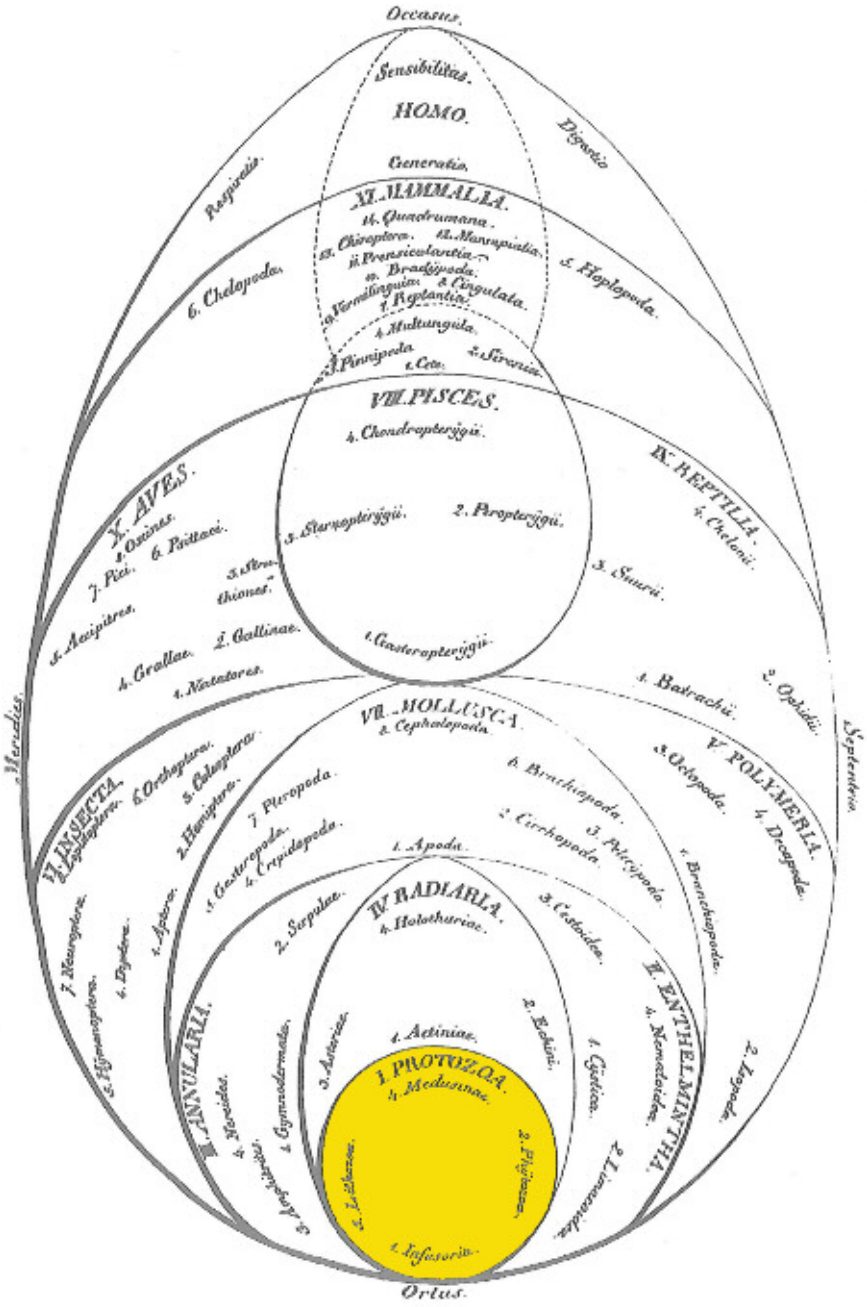
Culturing bias in the study of marine heterotrophic flagellates diversity



Javier del Campo

Departament de Biologia Marina i Oceanografia

Institut de Ciències del Mar

Tesi doctoral 2011

Departament de Microbiologia
Universitat de Barcelona
Programa de Microbiologia Ambiental i Biotecnologia 2009 - 2011

Culturing bias in the study of marine heterotrophic flagellates diversity

Biaix de cultiu en l'estudi de la diversitat dels flagel·lats heterotròfics marins

Memòria de la tesi presentada per obtenir el grau de
Doctor per la Universitat de Barcelona

Barcelona, maig de 2011

Javier del Campo García-Ramos
Departament de Biologia Marina i Oceanografia
Institut de Ciències del Mar - CSIC

Dr. Ramon Massana i Molera
Director

Dra. Rosa M. Pintó Solé
Tutora

Georg Augustus Goldfuss's phylogenetic tree published in 1820.
Adapted by Javier del Campo.

Als meus pares
A la Gemma

Quan el senyor Mitsushige no era més que un nen, se li va demanar que llegís un passatge d'un llibre del monjo Kaion: ell es va dirigir als altres nens i als acòlits per dir-los:

"Us demano que us apropeu i escolteu. És molt difícil llegir quan amb prou feines tens algú que t'escolti". El sacerdot va quedar impressionat i els hi va dir als seus seguidors: "Aquest és l'esperit amb el que s'han de fer totes les coses".

Hagakure, Yamamoto Tsunetomo, 1716

Agraïments

Suposo que sent fidels a la història el primer agraïment hauria de ser per al Sr. Steven Spielberg per portar Jurassic Park al cinema i haver evitat d'aquesta manera que dediqués la meva vida a fer coses més profitoses com ser arquitecte i triés el camí de la recerca en biologia. Com que no el conec de res i com que començo a pensar que no em va fer cap favor no li agrairé res i ho deixaré així, com un apunt històric.

Així doncs, el meu primer agraïment és, com no podia ser de cap altra manera, per al Ramon Massana, les raons són tantes que fan de mal resumir. Gràcies per donar-me una oportunitat, gràcies per dedicar (bona) part del teu temps a discutir qualsevol tema per menor que fos (crec que ràpidament va adonar-se que discutir era lo meu), i sobretot gràcies per ensenyar-me com s'han de fer les coses en ciència: amb rigor, perseverança i passió.

A les qui han estat les meves companyes de fatigues durant la meva estada a l'Institut de Ciències del Mar, amb qui he pogut parlar de tot i que m'han ajudat en tot, a la Vane, la Clara i la Irene: mil gràcies. Hi ha algú més amb el que he passat hores i hores al microscopi intentant ensinistrar els flagel·lats heterotròfics més rebels, el Fabrice, que no només ha estat un gran company de laboratori sinó que també m'ha ensenyat a encarar la recerca amb audàcia i arriscant; a vegades només deixant de banda la seguretat i les pors es poden fer avançar les coses, gràcies a tu també.

Hi ha d'altre gent amb qui he tingut l'oportunitat de compartir dubtes i problemes al voltant d'aquests diminuts organismes que han monopolitzat els meus últims anys i al voltant de la vida del doctorand, que no és pas fàcil. Vull agrair a la Laure Guillou i el Mike Sieracki, per acollir-me als seus laboratoris, on vaig poder experimentar altres maneres de fer i de plantejar la recerca, i vaig conèixer un munt de gent interessant. També vull agrair a la Raquel, al Massimo, al Felix i al Ramiro el fet de compartir amb mi els seus coneixements i la seva feina, fent que l'estudi dels flagel·lats heterotròfics hagi esdevingut una rica i variada feina col·lectiva. Gràcies als meus incombustibles companys de despatx: Clara, Ben, Sergio i Martí, per fer que les hores passessin més ràpid (a vegades massa i tot). Un gran gràcies general a tota la gent de l'ICM, en especial a la gent del Departament de Biologia Marina i Oceanografia. Una gran família científica que ha estat capaç de crear a l'última planta de l'Institut un brou de cultiu perfecte per a la creativitat i la innovació.

Abans d'arribar a l'ICM però, vaig passar uns anys al Departament de Microbiologia de la Facultat de Biologia de la Universitat de Barcelona, no oblidaré pas el grup on se'm van obrir les portes del món de l'ecologia microbiana. Vull agrair al Ricard Guerrero per mostrar-me un món tant interessant com emocionant, una visió de la ciència que poca gent té a dia d'avui. Ell és un savi de la vella escola, això se troba a faltar i em considero afortunat per haver compartit amb ell inquietuds i projectes. I no trobaré paraules adequades per donar les gràcies a la Laura Villanueva, sense el seu suport i guiatge no hagués pogut tirar endavant la feina que vaig desenvolupar al Departament i que en últim terme em va conduir a l'ICM, sempre serà la "jefa". Aquells anys a micro no haguessin estat pas el mateix sense companys com la Cristina, l'Oscar, el Quim o l'Unai, per mi l'ànima del departament; sense un estudiant de pràctiques de nom Albert que em va donar un cop de mà, i diversos ensurts al lab, durant els meus últims mesos allà i que ara és un doctorand de profit al CEAB o sense el Rafa, veí del Departament de Bioquímica, reusenc orgullós, casteller i gran amic.

Apartant-nos del món de la recerca no puc (ni vull) deixar de fer esment a aquells que sempre m'han fet costat, a les verdes i a les madures, la família. Gràcies a la família Moncunill (i rodalies), al Josep i a la Mariona, per esdevenir uns segons pares per mi, i que juntament amb la Mariona i el Silver han fet de les sobretaules de diumenge la millor assessoria sociolaboral que hom hagi pogut tenir. A la Picu, la Mari i l'Esther (les meves "àvies" maternes) i al Julian i la Manuela (els meus avis paterns), amb la seva confiança i el suport incondicional que sempre em van mostrar, em van impulsar fins aquí des del mateix moment que vaig ser capaç de fer 4 passes (i abans i tot). A l'Enrique, per exercir de germà gran sempre que ho he necessitat. Als meus pares, suport moral, financer, principal club de fans de la meva persona i, sense cap mena de dubte, model i inspiració per anar per aquests mons de Déu durant aquests anys, i els que vindran. Gràcies per creure en mi i per inculcar-me els valors de l'esforç i del respecte per l'estudi i per la feina.

Per últim vull destacar que bona part d'aquesta tesi ha estat possible gràcies a la Gemma, sense els seus consells, la seva paciència i la seva comprensió això de cap manera hagués arribat a bon port (i no són paraules buides, realment no hagués arribat a bon port). Gràcies per infinites correccions i suggeriments tant en articles com en aquesta mateixa memòria, gràcies pels consells assenyats al llarg de tot el doctorat i gràcies per omplir de bons moments totes aquelles estones en que ens hem pogut permetre el luxe de no dedicar-nos ni pensar en ciència.

Se que em deixo gent, però tinc el ferm objectiu de fer només una pàgina d'agraïments, tot i que hi ha masses persones a les que donar les gràcies. Doncs el periple d'aquesta tesi m'ha permès créixer com a científic i com a persona, gràcies a tots aquells amb qui he compartit el trajecte fins aquí. Amb alguns ens veurem sovint amb d'altres mai més, però sigui com sigui l'experiència i el record romandran per sempre.

Contents

Culturing bias in the study of marine heterotrophic flagellates diversity

Introduction to heterotrophic flagellates

350 years of natural history of protozoa

Protists (unicellular eukaryotes) are major forces driving our planet ecosystems and are an outstanding reservoir of biological diversity (genes, molecules, metabolic pathways, and cellular processes) yet to be discovered (Falkowski et al. 2008) accounting for most eukaryotic diversity. They are also main actors in macro- and microevolutionary processes for life on Earth (Cavalier-Smith and Chao 2006). We have centered our research interests on protozoa. The original concept of protozoa referred to unicellular organisms with an animal like behaviour (for example motility) and we know today that this concept does not have any evolutionary or taxonomic meaning. Instead, protozoa can be regarded as a functional and operative classification to target heterotrophic protists. Within protozoans, heterotrophic flagellates are the ones that posses one or more whip-like organelles called flagella, that are used both for motility and predation. Our knowledge on these organisms, initially considered as tiny animals, has been outlined and improved during the last 350 years (Leadbeater 2000), and relevant scientists have participated in this collaborative adventure (Fig. 1).

Protists were first described by the merchant draper from Delft (Holland), and father of microbiology, Antoine van Leeuwenhoek. Using his primitive homemade microscopes, during the second half of the XVIIth century he was able to observe protists, among many other microorganisms. Some of the descriptions made by Leeuwenhoek correspond actually to flagellates, probably *Monas* spp. or *Bodo* spp. (van Leeuwenhoek 1677). But was Otto Friedrich Müller, a Danish marine invertebrate zoologist, the first to formally describe several species of flagellates (Müller 1773 and 1786) under the nomenclature system created by Linnaeus. Müller was also responsible for the seminal taxonomy of heterotrophic protists. In 1817 Georg Augustus Goldfuss introduced the term Protozoa, but without a proper definition or explanation (Goldfuss 1817). According to Goldfuss, protozoa included polyps, medusa, infusoria and phytozoa (animal-like plants and plant-like animals).

In the XIXth century under the influence of cell theory by Schelieden (1838) and Schwan (1839) different naturalists recognized and described the unicellularity of several eukaryotic microbes. Among them, Meyer and Dujarin were the pioneers studying protozoa under the influence of the cellular paradigm and Barry and von Siebold systematized, explained and tidied all this knowledge.

Figure 1. Great names in protozoology: a) Antoine van Leeuwenhoek, b) Otto Friedrich Müller, c) Georg Augustus Goldfuss, d) Ferdinand Julius Cohn, e) Otto Bütschli, f) Thomas Cavalier-Smith.

Ferdinand Julius Cohn is recognized as the responsible of the term Flagellates (Cohn 1853), to refer to some flagellated protozoa. He also originated one of the biggest teasers in protist studies when confirmed that the green coloration of some flagellates was due to chlorophyll and that these organisms had the ability to photosynthesize. This finding leaded to a situation were both "plants" and "animals" were included inside protists, which were then classified separately in the plant and the animal kingdom.

After the publication of the Origin of Species by Charles Darwin (1859) several works were born under the prism of evolution setting the basis for protist taxonomy, as we know it nowadays. Big names of biology such as Haeckel, Jans-Clark and Kent contributed with their studies on the relationship between sponges and choanoflagellates. But there is one name that must be highlighted: Otto Bütschli, the great architect of protozoan systematics. He developed the first exhaustive systematics, based mainly in locomotive traits, classifying all known species among five classes: *Sarcodina* (amoebae), *Mastigophora* (flagellates), *Infusoria* (ciliates), *Sporozoa* (apicomplexa) and *Radiolaria* (Bütschli 1880/9). His classification influenced protist systematics during the first half of the XXth century.

The seventies represented the end of Bütschli's systematics. Even being very convenient it didn't reflect real evolutive relationships between organisms. The introduction of electron microscopy in protistology studies and the easiest access to molecular data clearly contributed to that end. One of the most influent protistologists from the 70s to the end of the XXth century was the Englishman Thomas Cavalier-Smith, who had done a meticulous job on systematics and taxonomy of protists based on molecular markers, mainly 18S rDNA.The last change of paradigm occurred recently, during the first years of the XXIst century with the introduction of environmental molecular surveys of protist communities, opening the doors of the omics age to protistology. In 2001 appeared the first culture independent studies where environmental 18S rDNA sequences were used to describe protist diversity in marine water column samples and to improve known phylogenies (López-García et al. 2001 and Moon-van der Staay et al. 2001). These seminal studies highlighted how much *in situ* protist diversity was still unknown.

This is the point where we stand now, having the huge responsibility to integrate this new knowledge with the invaluable heritage left to us by the great scientists such as Cavalier-Smith, Bütschli or Müller. We know now that protists are present in all branches of the eukaryotic tree of life and represent the most widespread life form within eukaryotes (Fig. 2). Excepting archaeplastida, all supergroups also contain protozoans.

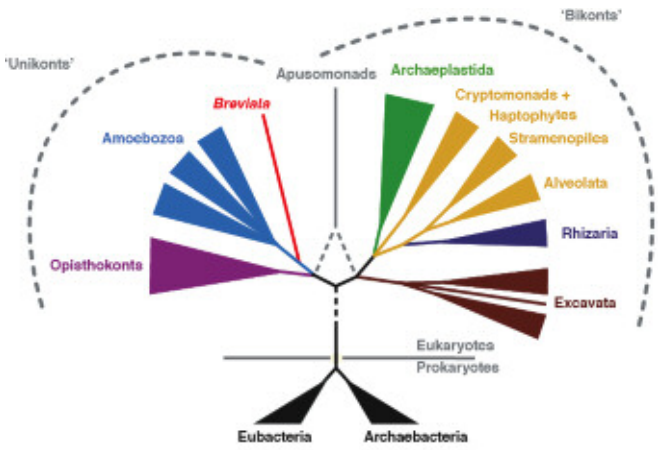


Figure 2. The eukaryotic tree of life. Adapted from Roger and Simpson 2009.

The role of heterotrophic flagellates

Within protozoa, it is well known that heterotrophic flagellates (Fig. 3) play a key role in marine food webs as already highlighted in the seminal paper presenting the microbial loop concept (Azam et al. 1983; see also Fenchel 1988). These minute organisms are distributed in planktonic environments at concentrations between 10^2 and 10^5 cells ml^{-1}, representing 10-30% of microbial eukaryotes in upper marine waters (Jürgens and Massana 2008). They are main actors in global biogeochemical cycles as bacterial grazers, trophic linkers and nutrient remineralizers (Sherr and Sherr 2002) and constitute a diverse assemblage of poorly identified species (Arndt et al 2000, Vaulot et al. 2002). Heterotrophic flagellates are often phagotrophs that graze and control the abundance of prokaryotes and very small eukaryotes (Pernthaler 2005) but also may include dispersal stages of parasites of other marine organisms (Guillou et al. 2008). This central role in marine ecosystems has been translated into a great interest in maintaining these organisms under controlled conditions in the laboratory, to study grazing rates, growth rates, prey preferences, ultrastructure, genomics or transcriptomics. Cultures have been essential for physiological and phylogenetic studies but the ecological relevance of cultured strains is not clear. So, do the cultured heterotrophic flagellates represent those that dominate in natural environments?

The gold standard for environmental diversity

Over the last decade, 18S rDNA clone libraries have been considered as the gold standard approach for studying protist diversity in the environment (Epstein and López-García 2008, Massana and Pedrós-Alió 2008). Environmental sequences highlight the dominant members of natural assemblages and may reveal new and unexpected lineages. These investigations, mostly performed on the picoplanktonic size fraction (0.8 to 3 µm), have unveiled novel high rank groups such as the so-called MALV, marine

alveolates (López-García et al. 2001), MAST, marine stramenopiles (Massana et al. 2004), and picobiliphytes (Not et al. 2007), many of which have become cornerstone taxa that often dominate the community in terms of clonal abundance. This newfound diversity has significantly altered our description of marine microbial food webs and the evolution of eukaryotes.

There is little doubt that molecular surveys offer an improved view of *in situ* diversity for very small protists as compared with previous strategies, essentially based on microscopical inspections or isolation in pure cultures (Caron et al. 2004). However, they do not give the definitive answer. Eukaryotic microbial diversity assessed by means of environmental 18S rDNA sequences have generated broadly similar composition patterns in the different studies done so far, with dominance of nonphotosynthetic groups, including tiny parasites and grazers. In contrast, epifluorescence microscopy typically reveals a dominance of photosynthetic or mixotrophic cells over heterotrophic cells in the oceans (Jürgens and Massana 2008). This suggests that 18S rDNA clone libraries also significantly bias protist diversity.

Several technical limitations inherent to culture-independent explorations of microbial diversity have been highlighted (Wintzingerode et al. 1997; Moeseneder et al. 2005).

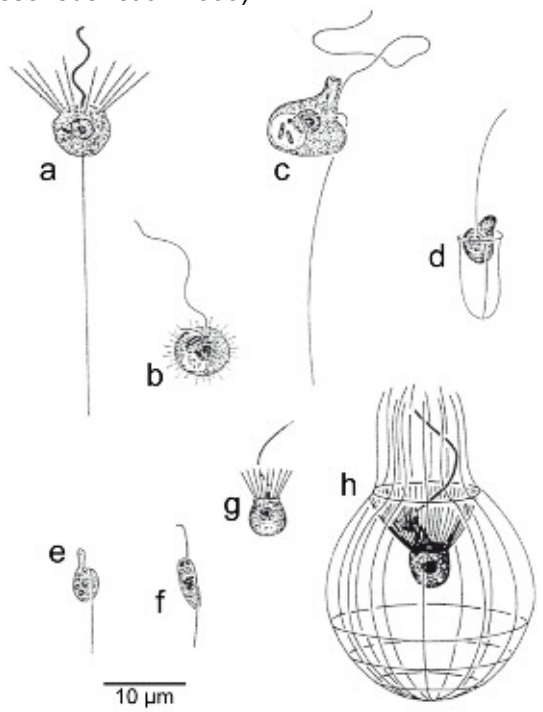


Figure 3. Some of the heterotrophic flagellates considered abundant in marine plankton according to culture dependant studies: a) *Pteridomonas*, b) *Paraphysomonas*, c) *Pseudobodo*, d) *Bicoeca*, e) *Rynchomonas*, f) *Bodo*, g) *Monosiga*, h) *Diaphaoneca*. Adapted from Fenchel 1986.

Particularly relevant among them are (Suzuki and Giovannoni 1996), primer selectivity, varying rDNA operon copy numbers (Zhu et al. 2005), and the existence of pseudogenes (Thornhill et al. 2007) or extracellular DNA (Paul et al. 1990.). A promising alternative that does not require PCR steps is the metagenomic approach, based on direct cloning and shotgun sequencing of environmental DNA (Rusch et al. 2007). With respect to eukaryotic microbes, phylogenetic information present in metagenomic libraries has thus far received very little attention (Piganeau et al. 2008). Another approach is to target directly the 18S rRNA (the ribosomes themselves) as a proxy for both diversity and metabolic activity of cells (Poulsen et al. 1993, Stoeck et al. 2007). This avoids considering differences in rDNA copy number and the interference of dissolved DNA

Culturing bias

Known protist diversity and biology is generally based on cultivated strains, on which ultrastructural, physiological and molecular studies have been performed. Cultivated heterotrophic flagellates belong to many different taxonomic groups, being represented in all eukaryotic supergroups (being archaeplastida the only exception). In aquatic samples, chrysophytes, choanoflagellates and bicosoecids are the most commonly isolated groups (Leipe et al. 1994, Andersen et al. 1999; Cavalier-Smith and Chao 2006) and were proposed to account for a significant fraction of heterotrophic flagellates (Arndt et al. 2000; Patterson and Lee 2000). In top of that, environmental molecular surveys have unveiled entirely novel lineages that have refused cultivation so far. Thus, fundamental evolutionary and ecological insights might have passed unnoticed due to our inability to culture relevant species. So, morphological and genomic information that can be obtained from cultures are missing.

Nowadays, culture bias definitely remains as one of the most critical challenges faced by scientists aiming to achieve a full understanding of the ecological role of microbes (Giovannoni et al. 2007) and is currently a bottleneck in ecosystem studies (Giovannoni and Stingl 2007, Raes and Bork 2008). Environmental DNA surveys demonstrate the extent to which culturing efforts poorly capture *in situ* microbial diversity (Pedrós-Alió 2006) (Fig. 4). It is estimated that as little as 0.1 to 1% of bacterial and protist cells can be easily cultured (Amann et al. 1995, Caron et al. 1989). Ironically, the most represented taxa in the environment refuse culturing while most of the strains represented in culture are very scarce in the environment (Massana et al. 2004).

Bacterivorous flagellates have been invariably cultivated using rich media composed of seawater supplemented with cereal grains or yeast extract that promote the growth of large bacteria at superior densities than *in situ* abundances.

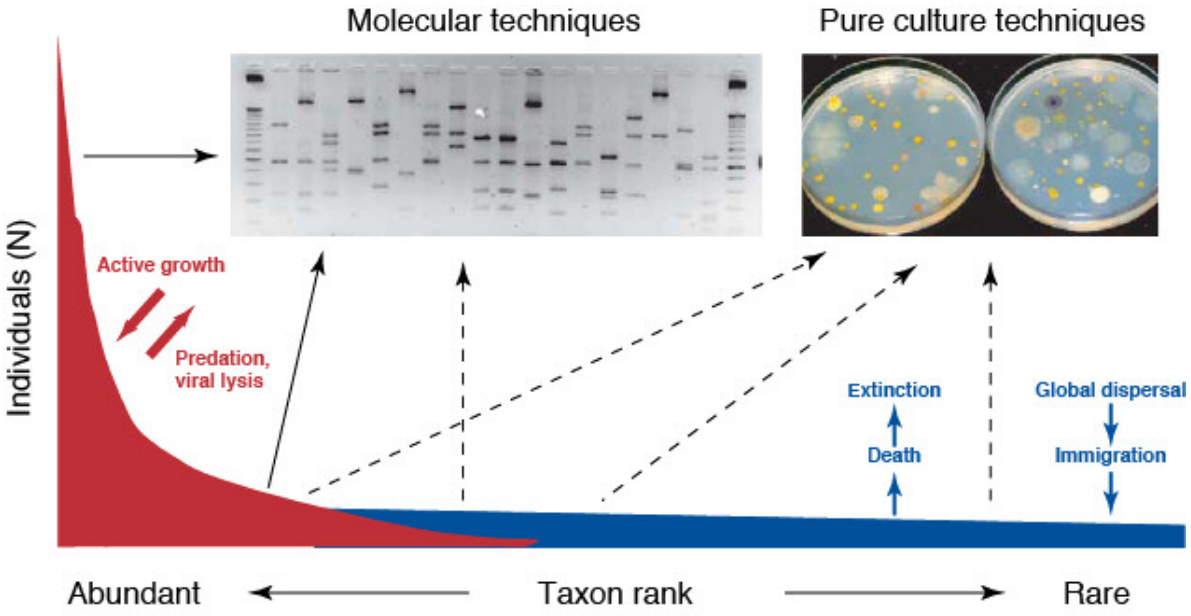


Figure 4. Plots of number of individuals of the different taxa in an assemblage, ranked according to their respective abundance. The total curve represents biodiversity and is postulated to be composed of two sections. The red section represent the abundant taxa that constitute the diversity, and are expected to have an active role in carbon and energy flow in a given ecosystem. The blue section of the curve corresponds to rare taxa, which survive in the ecosystem at low abundance, perhaps as resting stages or spores. Adapted from Pedrós-Alió 2006.

This strategy retrieves mainly the same pools of species such as *Cafeteria* spp., *Paraphysomonas* spp., or *Bodo* spp. (Fenchel 1982; Arndt et al. 2000, Scheckenbach et al. 2005), which are considered to be generally rare in the marine plankton (Jürgens and Massana 2008). Abundant taxa identified by molecular surveys still remain uncultured (e.g. bacterivorous MASTs clades, Massana et al. 2006a). Although culturing bias and organic matter effect on enrichments is a well-known (and always controversial) topic in protistology, there are no published studies where the effect of the added substrates on the microbial community is comprehensively analyzed. Previous studies addressing the protists dynamics in microbial amended or unamended incubations (Lim et al. 1999, Countway et al. 2005, Massana et al. 2006b) focused on the evolution and properties of the incubated community along time and were not designed to face the culturing bias conundrum.

Culture bias can be overcome by using original culturing strategies, as demonstrated for *Pelagibacter ubique* and marine crenarchaea (Rappé et al. 2002, Könneke et al. 2005), both initially detected through environmental molecular surveys and later identified as ecologically relevant taxa. *Pelagibacter ubique* was brought into culture by mimicking oligotrophic conditions and marine crenarchaea was cultured in media amended with ammonia once molecular data revealed they were ammonia oxidizers. Similar culturing efforts have seldom been applied to marine protists, even though culture bias is perceived as a major limitation to investigate further the functional role and ecological significance of photosynthetic protists (Vaulot et al. 2008), being particularly severe for the heterotrophic ones which depends on organic food source for growth (Jürgens and Massana 2008).

Objectives

Our main goal in this study was to overcome the culture bias in marine heterotrophic flagellates. This main aim could be structured in three general aspects. The first was to determine the importance and representativity of cultured flagellates in environmental molecular studies. The second was the study of the culture bias from an experimental point of view. Finally, we aimed at obtaining new cultures of heterotrophic flagellates. To achieve this we defined more specific objectives:

1. Determine the clonal contribution of 18S rDNA sequences of chrysophytes, choanoflagellates and bicosoecids in marine and freshwater systems, improve the phylogeny of these groups and analyze their sequence novelty.

2. Determine the effect of PCR induced biases by comparing 18S rDNA sequences obtained from the Global Ocean Survey (GOS) metagenomic database (Rusch et al. 2007) and from standard clone libraries (Massana and Pedrós-Alió 2008).

3. Compare the protist diversity inferred from clone libraries both from extracted DNA and extracted RNA from the same sample, in order to delineate the biases

introduced in environmental diversity studies generally based on DNA.

4. Report the effects of different organic matter enrichments to heterotrophic flagellates community structure and put together ideas and concepts related to the culturing bias that had been generally assumed or refused but never specifically addressed.

5. Develop a new culturing approach to isolate previously uncultured heterotrophic flagellates species that might be abundant in the marine plankton.

Discussing culturing biases in the study of heterotrophic protists diversity

A viral crash as the starting point

With the intention of observe the effect of organic matter over the heterotrophic flagellates enriched from an open ocean community we started different incubations from a central Indian Ocean sample (**Paper 1**). Sequencing and FISH analyses from these incubations showed that essentially *Cafeteria roenbergensis* and *Caecitellus paraparvulus* were growing in the enriched conditions. These two species are well-known cultured heterotrophic flagellates (Patterson and Lee 2000). *Caecitellus paraparvulus* grew first and probably was initially more abundant than *Cafeteria roenbergensis*, because it was detected in the clone library done with the *in situ* sample, but *Cafeteria roenbergensis* was the dominant flagellate at the peak. Interestingly, *Cafeteria roenbergensis* numbers rapidly decreased after the population peak and at the 8th day were below detection by FISH. This appeared to be related to the presence of viruses infecting the population. The virus was specific to *Cafeteria roenbergensis*, since only infected cells from this taxa were observed by FISH. In **Paper 1** we have shown that a cultured heterotrophic flagellate and its specific virus were readily selected by enriching an oceanic sample with organic matter. The virus could spread fastly and crash completely the host population. This result is in agreement with the general view of the ecological role of viruses (Bratbak et al. 1993; Garza and Suttle 1995, Tarutani et al. 2000; Brussaard 2004). Our data expands the existing information, since to our knowledge only two viruses infecting heterotrophic microeukaryotes have been isolated and maintained in culture (Garza and Suttle 1995, Takao et al. 2005).

Besides the intrinsic interest of the description of the crashing event of a protozoan population due to a virus, there were two other aspects of these enrichments that captured our attention and drove us to new studies. The first was the negligible presence of *Cecitellus paraparvulus* and *Cafeteria roenbergensis* in the original sample, even being the most successful protists in the enrichments and two of the most reported flagellates in the sea. We decided to look for them and other classically relevant flagellates in environmental surveys, resulting **Paper 2.** The other striking result was the fast growth of *Cafeteria roenbergensis* in the enriched mesocosms, becoming the dominant organism of the community. As this was the second report of this, after Lim et al. 1999 showed a similar case with *Paraphysomonas* sp., we decided to demonstrate that this enrichment bias was a usual phenomenon that pervades the classical culture processes that had been used for years. This concern generated **Paper 4**, which was then continued by **Paper 5**.

Most wanted

As told before in the introduction chrysophytes, choanoflagellates and bicosoecids are considered to account for most of aquatic heterotrophic flagellates, but it is not clear if these cultured strains are ecologically relevant. A literature search on the species most commonly retrieved in culture dependant studies indeed reveals that these groups are widely reported (Table 1). The environmental 18S rDNA libraries have unveiled a large diversity and highlighted new lineages that appeared in most studies in high clonal abundance, however chrysophytes, choanoflagellates and bicosoecids were generally represented by few sequences in marine (Massana and Pedrós-Alió 2008) and freshwater (Lefranc et al. 2005; Richards et al. 2005; Slapeta et al. 2005) individual studies.

Table 1. Most reported heterotrophic flagellates species in culture dependent studies. Choanoflagellates, not included in the table, are reported in all these studies. Data obtained from six studies: Patterson et al. 1993, Vørs et al. 1995, Ivavalko et al. 1997, Tong 1997, Tong et al. 1997 and Tikhonenkov et al. 2006

Organism	Affiliation	Studies
Cafeteria roenbergensis	Stramenopile, Bicoecida	5
Caecitellus parvulus	Stramenopile, Bicoecida	3
Boroka karpovii	Stramenopile, Bicoecida	4
Neobodo designis	Kinetoplastea, Neobodonida	5
Rhynchomonas nasuta	Kinetoplastea, Neobodonida	5
Ancyromonas sygmoides	Incertae sedis, Ancyromonas	5
Pteridomonas danica	Stramenopile, Dictyophyceae	6
Paraphysomonas imperforata	Stramenopile, Chrysophyceae	5

Nevertheless sequences affiliated with minor groups (in terms of clonal abundance) have often been under analyzed, and this hides a potentially relevant source of phylogenetic information. In **Paper 2** we did an effort to analyze the sequences affiliating to chrysophytes, choanoflagellates and bicosoecids from environmental molecular survey existing in public databases. The relative clonal abundance of these three groups suggested that they might be less important than expected in marine systems, since they only accounted for 5% of clones, contrasting with the

large clonal abundance of MAST or MALV. Phylogenetic trees adding environmental complete sequences to the dataset of sequences from cultured strains present the most accurate representation of the diversity of these groups, with the emergence of several new clades formed exclusively by environmental sequences. Exhaustive data mining in sequence databases allowed the identification of new diversity hidden inside chrysophytes, choanoflagellates and bicosoecids. This strategy has been also applied to other protist groups such as Mamiellophyceae (Marin and Melkonian 2010).

We applied a new approach (Massana et al. 2010) to address the novelty of a given dataset based on the similarity against GenBank sequences. The large novelty displayed by the environmental sequences of each group was interpreted in terms of efforts in culturing and environmental sequencing. A low correspondence between environmental sequences and sequences obtained from cultures was the more common situation, and this highlights the culturing bias. On the other hand, sequencing environmental DNA is relatively straightforward and there are little chances to miss quantitatively important major phylogenetic groups. Nevertheless, an insufficient sequencing effort was generally found, suggesting that there is plenty of room to discover additional diversity for these groups using environmental molecular surveys, especially if they take advantage of new high-throughput sequencing technologies (Amaral-Zettler et al. 2009) or use group-specific primers (Bass and Cavalier-Smith 2004). Alternatively, another explanation of this insufficient sequencing effort would be a large endemism of the organisms carrying the "novel" sequences, which might appear only in the studied site. Enhanced culturing and sequencing efforts will be needed to reach a full understanding of protist *in situ* diversity and ecological role. The main contribution to enlarge 18S rDNA databases will be the high-throughput studies that have been already released (Amaral-Zettler et al. 2009, Stoeck et al. 2009, Benhke et al. 2010, Cheung et al. 2010, Edgcomb et al. 2011) or will be in the near future (such as during the BioMarks project: www.biomarks.org). The accurate phylogenetic analysis done in **Paper 2** are fundamental to provide curated trees that will be used as phylogenetic maps to avoid getting lost inside this increasing sea of data.

New insights on protist diversity

The current picture of marine eukaryotic biodiversity may be significantly skewed by PCR amplification biases, occurrence of rDNA genes in multiple copies within a single cell, and the capacity of DNA to persist as extracellular material. Part of this PhD study was driven to investigate how severe were these biases and how the view of protist diversity could be improved. We consider this information relevant and interesting for our objectives because those techniques, as said before, are basic instruments of our research.

In **Paper 3** we performed an in-depth analysis of the metagenomic dataset from the GOS expedition, seeking eukaryotic signatures through the presence of 18S rDNA genes. Metagenomic approaches directly clone and shotgun sequence the DNA from a given sample, without prior PCR. The similarity in diversity patterns between the PCR cloning and metagenomic approaches suggests little impact of the PCR step on the outcome of clone libraries in terms of sequence diversity and relative contribution of specific taxa. Separate analysis of the two size fractions (0.2-0.8 µm and 0.8-3 µ m) from the GOS dataset revealed clear differences in terms of taxonomic composition. As the smallest eukaryotic organism known so far has a cell diameter of 0.8 µm (Courties et al. 1994), some of the 18S rDNA signatures observed in the <0.8 µm fraction might indeed derive from very small protists, but many sequences most likely derive from cell debris or extracellular DNA from larger cells (Jiang and Paul 1995; Dell'Anno and Danovaro 2005; Vlassov et al. 2007). It is likely that a fraction of the extracellular DNA is retained onto 0.2 µm filters, through collection of aggregates or molecular adsorption. Consequently, we believe that it is important to consider the interference of extracellular DNA when assessing the diversity of eukaryotic microbes in ecological studies.

Another alternative approach to investigate microbial diversity is to target directly the 18S rRNA. We compared the phylogenetic patterns from rDNA and reverse transcribed rRNA 18S clone libraries from the same sample harvested in the Mediterranean Sea. It is generally recognised that 18S rDNA diversity surveys are not quantitative with respect to cell abundance (Kirchman 2002). Diversity assessed by rRNA led to a drastically different view of the community as compared to the classical DNA-based approach. The approach avoided the effect of taxon-specific rDNA copy number and the interference of extracellular material, since RNA is much less stable than DNA. We found very little overlap in the sequences retrieved in the DNA and RNA libraries. With respect to heterotrophic protists, the rRNA approach points to MAST organisms as prominent members of the community, which together with their widespread distribution suggest they might actually be the major protistan predators in the oceans (Massana et al. 2006a). Environmental 18S rRNA clone libraries appear to represent a promising means to minimize some important biases and thereby offer new perspectives in the study of the diversity and function of marine protist.

Facing Culturing Bias

Culturing bias paradox appears as a reasonable doubt behind ecological studies, but has never been properly addressed. **Paper 4** was designed to deal with this by

analyzing the effect of organic matter in a confined community of natural heterotrophic flagellates. According to the preliminary information retrieved from **Paper 1** several microcosms were established with an increasing amount of organic matter and different organic matter sources.

The peaks of abundance of bacteria and heterotrophic flagellates increased with organic matter and this was accompanied by a delay in the apparition of both peaks. This delay could be due to the time needed by the community to adapt to the enrichment conditions. In the organic matter enriched incubations, the dominant original populations, adapted to lower food source concentrations, could be inhibited to grow, and some minoritary population, well-adapted to high nutrients concentrations, could fastly develop and adopt a dominant role in the community. Another factor that could delay the peak is the increasing number of cells in the enriched samples, which would then need more time of exponential growth. Most likely, the original dominant heterotrophic flagellates species were not prepared for these enriched conditions. The differences in growth rate are consistent with different species growing in different incubations, with fast growing populations in enriched conditions. Moreover, the very large bacteria in the first peak, together with a large proportion of bacteria in aggregates in enriched treatments, seemed to become a grazing refuge that avoided heterotrophic flagellates exploitation (Jürgens and Güde 1994, Hahn et al. 2000, Simek et al. 2001). Only the heterotrophic flagellates species adapted to eat large free-living bacteria or bacteria in aggregates will be able to proliferate and dominate in these conditions.

Data obtained by DGGE fingerprinting and clone libraries from the heterotrophic flagellates peaks revealed that the unamended treatments were similar to the original sample and highly different from the rest of the enriched treatments. This agreed with previous studies that showed that unamended incubations promoted the growth of heterotrophic flagellates present in the natural assemblage and prevented a great modification of the community structure (Massana et al. 2006b). On the contrary the enrichments were promoting the mergence of other populations not very abundant in the original sample (Lim et al 1999). Our phylogenetic analysis highlighted clearly the bias effect caused by the organic matter. While in the unamended clone library there was a predomination of uncultured protists, the increase of organic matter reversed this trend, and cultured protists became clearly dominant in the most enriched incubations. The selective and homogenizer role of the organic matter was confirmed by a clear decrease of the diversity when increasing organic matter concentration (Shannon index from 2.5 to 1). Organic matter enrichments resulted in communities not only dominated by cultured organisms but also less diverse, here dominated by *Paraphysomonas* spp. and *Oikomonas* spp. These species have a great capacity to eat large bacteria at high abundances (Fenchel 1982, Caron et al. 1985, Eccleston-Parry et al. 1994) and are known to be commonly isolated from marine snow (Davis et al. 1984). They have the potential to outcompete the heterotrophic flagellates that were originally dominant in the oligotrophic initial sample.

One of the main reasons of the culturing bias in heterotrophic flagellates is the use of organic matter in the isolation process, driving a shift on the community to conditions closer to laboratory cultures. Culturing efforts done up to now have been extremely important, but a new culturing impulse using novel strategies is needed to advance on our understanding of protist ecology and evolution.

Looking for alternatives

In order to bring into culture ecologically relevant heterotrophic flagellates, we mimicked oligotrophic marine conditions by amending sterile seawater with a mix of natural bacteria collected from the same sampling site at abundances only slightly higher than *in situ*. Each pre-culture was initiated with a single cell, obtained by serial dilution or by flow cytometry sorting, and incubated in the dark at *in situ* temperature. Strains belonging to lineages only known so far from environmental sequencing were isolated. In this process some pre-cultures were lost and others evolved to different species.

Two clonal and stable cultures were finally obtained. They included an uniflagellated rhizaria related to chlorarachniophytes and a biflagellated stramenopile distantly related to *Developayella* sp. Scanning electron microscopy performed on the two stable cultures revealed extremely small cells with little morphological features. The rhizarian isolate was distant to any described organism, its 18S rDNA sequence being only 90.6% similar to *Chlorarachnion reptans*, and it was highly similar to environmental sequences retrieved from different sites. Its basal position within chlorarachniophytes together with its observed heterotrophic nature suggests it represents an ancient lineage. Pending a formal description, this small heterotrophic flagellate has been named *Minorisa minuta* candidatus.

The functional response of *Minorisa minuta* candidatus yields a half-saturation constant much lower than that of other cultured flagellates (Rodríguez-Martínez, unpublished), suggesting that it is adapted to live at the usual bacterioplankton concentrations in oligotrophic waters. Counts provided by TSA-FISH reveals *Minorisa minuta* candidatus as a significant component of marine heterotrophic flagellates on a global scale, being both widely distributed and abundant. It accounts for 1.8% of heterotrophic flagellates in the Atlantic, Pacific, Indian, and Southern Oceans and the Mediterranean Sea, a value that

increases up to 5% when considering coastal sites only. *Minorisa minuta* candidatus was detected all year round in a coastal oligotrophic station in the NW Mediterranean Sea, ranging from 12 to 120 cells mL^{-1} and accounting for 5% of heterotrophic flagellates on average. Sizing cells in natural marine assemblages using microscopy confirms its picoeukaryotic character with cell size varying from 1 to 3 µm and an averaged size of 1.4 µm. Only *Symbiomonas scintillans* and *Picophagus flagellatus* (Guillou et al. 1999) get close to this size within heterotrophic flagellates, but still *Minorisa minuta* represents the smallest bacterial grazer known to date. Indeed, apart from this size record what differentiate this organism from other small heterotrophic eukaryotes is its wide distribution and its quantitative importance in different sites.

The isolation method presented in this work defines the path to follow in future environmental prospections of looking for new organisms to increase culture collections and the knowledge on heterotrophic protists.

Conclusions

The general conclusions of this thesis are:

1. We have shown that a cultured heterotrophic flagellate and its specific virus were readily selected by enriching an oceanic sample with organic matter. Our study demonstrates that viruses can also control heterotrophic flagellate populations. This strategy can also lead to the isolation of novel marine eukaryotic viruses.

2. Our survey using environmental sequences from public databases highlights a large emergent diversity (sometimes novel) of chrysophytes, choanoflagellates and bicosoecids in aquatic environments. This indicates both a bias in the representation of cultures and an incomplete sequencing effort for these groups. The extent of diversity and novelty is striking even for protist groups considered well characterized. This analysis can be extended to other groups in order to fully benefit from environmental molecular surveys.

3. Our analyses of the 18S rDNA sequences retrieved from the metagenomic GOS dataset do not reveal substantial differences as compared to PCR-based clone libraries. The similar outputs of both approaches suggests little impact of the PCR step in clone libraries in terms of sequence diversity and relative contribution of specific taxa.

4. The differences detected between DNA and RNA libraries from the same sample are consistent with the view that rRNA libraries reduce significantly two of the major biases of rDNA diversity surveys, the rDNA copy number and the occurrence of extracellular DNA. Considering the relative ease of handling ribosomal RNA molecules, extended diversity surveys based on environmental rRNA will undoubtedly provide insights into the ecology of uncultured species. Associated with stronger depth of sequencing this approach will probably help to achieve a nearly exhaustive view of protist diversity.

5. Enrichment incubations with organic matter modify the community dynamics due to a substitution process within both bacterial and heterotrophic flagellate members of the initial community. There is a dramatic decrease on diversity and a gradual increase of cultured organisms when organic matter increases.

6. Classical culturing techniques based on enrichments must be complemented with novel culturing strategies if we really want to catch the organisms responsible for most processes in the sea. Despite the importance of the obtained cultures until today we must expand and grow our collections with new cultures of environmental relevance.

7. We have designed a novel culturing approach of heterotrophic flagellates based on feeding single cells with natural bacteria. Through this approach we have isolated several small protists belonging to previously uncultured taxa and from distant lineages within the eukaryotic tree of life. When applied at different temporal and spatial scales, this strategy will potentially give access to a wealth of heterotrophic protists in culture.

8. The tiny uniflagellated *Minorisa minuta* candidatus stands up as the smallest bacterivore known so far. Moreover, it represents the only heterotrophic representative within the chlorarachniophyte lineage and is of primary interest to study the transition to secondary plastid endosymbiosis. As for its photosynthetic counterparts in the oceans, the genome analysis of *Minorisa minuta* candidatus will certainly reveal unprecedented cellular, biochemical, and evolutionary pathways.

9. *Minorisa minuta* candidatus has a worldwide marine distribution and is a significant member of heterotrophic flagellate assemblages, particularly in coastal waters. The physiological properties of *Minorisa minuta* candidatus can explain its ecological success and set this species as a good model for dominant marine heterotrophic flagellates, whose parameters could be used to improve ecological models. Getting the environmentally relevant bacteria *Pelagibacter ubique* candidatus in culture led to a leap forward towards a better understating of microbes' function in the oceans and opened up several research directions. Taming small marine predators with ecological relevance holds promise for similar future discoveries.

References

Amann RI, Ludwig W, Schleifer KH (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 parrallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS ONE 4: e6372

Andersen RA, van de Peer Y, Potter D, Sexton JP, Kawachi M, LaJeunesse T (1999) Phylogenetic analysis of the SSU rRNA from members of the Chrysophyceae. Protist 150: 71-84

Arndt H, Dietrich D, Auer B, Cleven E, Gräfenhan T, Weitere M, Mylnikov AP (2000) Functional diversity of heterotrophic flagellates in aquatic ecosystems. In Leadbeater, BSC, and Green, JC (eds) The Flagellates: Unity, Diversity and Evolution. London, UK: Taylor & Francis Press, pp 240-268

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad, F (1983) The ecological role of watercolumn microbes in the sea. Mar Ecol Prog Series 10: 257-263

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

Behnke A, Engel M, Christen R, Nebel M, Klein RR, Stoeck T (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. Environ Microbiol 13: 340-349

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

Bratbak G, Egge JK, Heldal M. (1993) Viral mortality of the marine alga Emiliania huxleyi (Haptophyceae) and termination of algal blooms. Mar Ecol Prog Series 93: 39-48

Brussaard CPD (2004) Viral control of phytoplankton populations - a review. J Eukaryot Microbiol 51: 125-138

Bütschli O (1880/9) Protozoa. In Bronn, HG (ed) Klassen und Ordungen des Their-reichs, vol. 1. Leipzig, Germany: CF Winter.

Caron DA, Goldman JC, Andersen OK, Dennett, MR (1985) Nutrient cycling in a microflagellate food chain: II. Population dynamics and carbon cycling. Mar Ecol Prog Ser 24: 243-254

Caron DA, Davis PG, Sieburth JMcN (1989) Factors responsible for the differences in cultural estimates and direct microscopical counts of populations of bacterivorous microflagellates. Microb Ecol 18: 89-104

Caron DA, Countway PD, Brown MV (2004) The growing contributions of molecular biology and immunology to protistan ecology: Molecular signatures as ecological tools. J Eukaryot Microbiol 51: 38-48

Cavalier-Smith, T, Chao EE (2006) Phylogeny and megasystematics of phagotrophic heterokonts (Kingdom Chromista). J Mol Evol 62: 388-420

Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. ISME J 4: 1053-1059

Cohn FJ (1853) Flagellata. Zeitschrift für wissenschaftliche Zoologie 6

Countway PD, Gast RJ, Savai P, Caron DA (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. Journal of Eukaryotic Microbiology 52: 95-106

Courties C, Vaquer A, Trousselier M, Lautier J, Chrétiennot-Dinet MJ, Neveux J, Machado C, Claustre H (1994) Smallest eukaryotic organism. Nature 370: 255

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London, UK: John Murray

Davis PG, Sieburth JM (1984). Estuarine and oceanic microflagellate predation of actively growing bacteria: estimation by frequency of dividing-divided bacteria. Mar Ecol Prog Series 19: 237-246

Dell'Anno A, Danovaro R (2005) Extracellular DNA Plays a Key Role in Deep-Sea Ecosystem Functioning. Science 309: 2179

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

Edgcomb VP, Orsi W, Bunge JA, Jeon S, Christen R, Leslin C, Holder M, Taylor GT, Suarez P, Varela R, Epstein S (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. ISME J doi:10.1038/ismej.2011.6

Epstein S, Lopez-Garcia P (2008) ''Missing'' protists: a molecular prospective. Biodiv Conserv 17: 261-276

Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034-1039

Fenchel T (1982) Ecology of heterotrophic microflagellates. IV. Quantitative occurrence and importance as bacterial consumers. Mar Ecol Prog Ser 9: 35-42

Fenchel T (1986) The ecology of heterotrophic microflagelates. Adv Microb Ecol 9: 57-97

Fenchel T (1988) Marine Plankton Food Chains. Annu Rev Ecol Systemat 19: 19-38

Garza, DR, Suttle CA (1995) Large double-stranded DNA viruses which cause the lysis of a marine heterotrophic nanoflagellate (Bodo sp.) occur in natural marine viral communities. Aquat Microb Ecol 9: 203-210

Giovannoni SJ, Foster RA, Rappé MS, Epstein S (2007) New cultivation strategies bring more microbial plankton species into the laboratory. Oceanography 20: 62-69

Giovannoni SJ, Stingl U (2007) The importance of culturing bacterioplankton in the 'omics' age. Nature Rev Microbiol 5: 820-826

Goldfuss GA (1817) Über die Entwicklungsstufen des Thieres. Nuremberg, Germany: Leonard Schrag

Guillou, L, Chrétiennot-Dinet MJ, Boulben S, Seung Yeo Moon-van Der Staay SY, Vaulot D (1999) *Symbiomonas scintillans* gen. et sp. nov. and *Picophagus flagellatus* gen. et sp. nov. (Heterokonta): two new heterotrophic flagellates of picoplanktonic size. *Protist* 150: 383-98

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R, Scanlan DJ, Worden AZ (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). Environ Microbiol 10: 397-408

Leadbeater BSC, McCready SMM (2000) The flagellates: historical perspectives. In Leadbeater, BSC, and Green, JC (eds) The Flagellates: Unity, Diversity and Evolution. London, UK: Taylor & Francis Press, pp 1-26

Lefranc M, Thénot A, Lepère C, Debroas D (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. Appl Environ Microbiol 71: 5935-5942

Leipe DD, Winright PO, Gunderson JH, Porter D, Patterson DJ, Valois F (1994) 16S-like rRNA sequences from Labyrinthuloides minuta and Cafeteria roenbergensis. Phycologia 33: 369-377

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

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

Hahn, MW, Moore, ERB, Höfle, MG (2000) Role of microcolony formation in the protistan grazing defense of the aquatic bacterium Pseudomonas sp. MWH1. Microb Ecol 39: 175-185

Ikävalko J, Gradinger R (1997) Flagellates and heliozoans in the Greenland Sea ice studied alive using light microscopy. Polar Biol 17: 473-481

Jiang SC, Paul JH (1995) Viral Contribution to Dissolved DNA in the Marine Environment as Determined by Differential Centrifugation and Kingdom Probing. Appl Environ Microbiol 61: 317-325

Jürgens K, Güde H (1994) The potential importance of grazing-resistant bacteria in planktonic systems. Mar Ecol Prog Series 112: 169-188

Jürgens K, Massana R (2008) Protistan Grazing on Marine Bacterioplankton. In Kirchman, DL (ed) Microbial Ecology of the Oceans, Second Edition. John Wiley & Sons, Inc., New York, USA, pp 383-441

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

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437: 543-546

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, 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

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

Massana R, Pernice M, Bunge J, del Campo J (2010) Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. ISME J 5: 184-195

Marin B, Melkonian M (2010). Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist 161: 304-36

Moeseneder MM, Arrieta JM, Herndl GJ (2005) A comparison of DNA-and RNA-based clone libraries from the same marine bacterioplankton community. FEMS Microbiol Ecol 51: 341-352

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 OF (1773) Vermium Terrestrium et Fluviatilium, seu animalium infusorium, helminthicorum et testaceorum, non marinorum, succincta historia. Copenhagen and Leipzig: Heineck and Faber

Müller OG (1786) Animalcula Infusoria Fluviatilia et Marina, qua detexit, systematice descripsit et ad vivum delineari craviti. Copenhagen and Leipzig: Müller

Not F, Valentin K, Romari K, Lovejoy C, Massana R, Töbe K, Vaulot D, Medlin LK (2007) Picobiliphytes: A marine picoplanktonic algal group with unknown affinities to other eukaryotes. Science 315: 252-254

Not F, Latasa M, Scharek R, Viprey M, Karleskind P, Balague V, Ontoria-Oviedo I, Cumino A, Goetze E, Vaulot D, Massana R (2008) Phytoplankton diversity across the Indian Ocean: A focus on the picoplanktonic size fraction. Deep Sea Res Part I 55: 1456-1473

Patterson DJ, Nygaard K, Steinberg G, Turley CM (1993) Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid North Atlantic. J Mar Biol Assoc UK73: 67-95

Patterson DJ, Lee WJ (2000) Geographic distribution and diversity of free-living heterotrophic flagellates. In Leadbeater, BSC, and Green, JC (eds) The Flagellates: Unity, Diversity and Evolution. London, UK: Taylor & Francis Press, pp 269-287

Paul JH, Cazares L, Thurmond J (1990) Amplification of the rbcL Gene from Dissolved and Particulate DNA from Aquatic Environments. Appl Environ Microbiol 56: 1963-1966

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol 14: 257-63

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

Piganeau G, Desdevises Y, Derelle E, Moreau H (2008) Picoeukaryotic sequences in the Sargasso Sea metagenome. Genome Biol 9: R5

Poulsen LK, Ballard G, Stahl DA (1993) Use of rRNA Fluorescence In Situ Hybridization for Measuring the Activity of Single Cells in Young and Established Biofilms. Appl Environ Microbiol 59: 1354-1360

Raes J, Bork P (2008) Molecular eco-systems biology: towards an understanding of community function. Nature Rev Microbiol 6:683-699

Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418, 630-633

Richards TA, Vepritskiy AA, Gouliamova DE, Nierzwicki-Bauer SA (2005) The molecular diversity of freshwater picoeukaryotes from oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. Environ Microbiol 7: 1413-1425

Roger AJ, Simpson AGB (2009) Evolution: revisiting the root of the eukaryote tree. Curr Biol 19:165-1677

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. PloS Biology 5: e77

Scheckenbach F, Wylezich 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. Aquatic Microbial Ecology 38: 239-247.

Schleiden MJ (1838). Beiträge zur Phytogenesis. Archiv für Anatomie, Physiologie und wissenschaftliche Medizin, Leipzig, Germany, pp 137-176

Schwann FT (1839) Mikroskopische Untersuchungen über die Übereinstimmung in der Struktur und dem Wachsthum der Thiere und Planzen. Berlin, Germany: GE Reimer

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

Simek K, Pernthaler J, Weinbauer MG, Hornak K, Dolan JR, Nedoma J, Masin M, Amann R (2001) Changes in bacterial community composition and dynamics and viral mortality rates associated with enhanced flagellate grazing in a mesoeutrophic reservoir. Appl Environ Microbiol 67: 2723-2733

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci 272: 2073-2083

Stoeck T, Zuendorf A, Breiner HW, Behnke A (2007) A molecular approach to identify active microbes in environmental eukaryote clone libraries. Microb Ecol 53: 328-339

Stoeck T, Behnke A, Christen R, Amaral-Zettler L, Rodriguez-Mora MJ, Chistoserdov A, Orsiand W, Edgcomb VP (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biology 7: 72

Suzuki MT, Giovannoni SJ (1996) Bias Caused by Template reannealing in the Amplification of Mixtures of 16S rRNA Genes by PCR. Applied and Environmental Microbiology 62: 625-630.

Takao Y, Nagasaki K, Mise K, Okuno T, Honda, D (2005) Isolation and characterization of a novel singlestranded RNA virus infectious to a marine fungoid protist, Schizochytrium sp. (Thraustochytriaceae, Labyrinthulea). Appl Environ Microbiol 71: 4516-4522

Tarutani K, Nagasaki K, Yamaguchi M (2000) Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton Heterosigma akashiwo. Appl Environ Microbiol 66: 4916-4920

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

Tikhonenkov DV, Mazei YA (2006) Distribution of heterotrophic flagellates at the littoral of estuary of Chernaya River (Kandalaksha Bay, White Sea). Russ J Mar Biol 32: 276-283

Tong SM (1997) Heterotrophic flagellates from the water column in Shark Bay, Western Australia. Mar Biol 128:517-536

Tong SM, Vørs N, Patterson DJ (1997) Heterotrophic flagellates, centrohelid heliozoa and filose amoebae from marine and freshwater sites in the Antarctic. Polar Biol 18:91-106

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

van Leeuwenhoek A (1677) Observations communicated to the publisher by Mr Antony van Leewenhoek in a Dutch letter on the 9th of Octob. 1676 here English'd: Concerning little animals by him observed in rain-, well-, sea- and snow-water, as also in water wherein pepper had lain infused. Phil Trans R Soc 13 (152): 347

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 (≤ 3µm) in marine ecosystems. FEMS Microbiol Ecol 32, 795-820

Vlassov VV, Laktionov PP, Rykova EY (2007) Extracellular nucleic acids. BioEssays 29: 654-667

Vørs N, Buck K, Chavez F, Eikrem W, Hansen L, Ostergaard J, Thomsen HA (1995) Nanoplankton of the equatorial Pacific with emphasis on the heterotrophic protists. Deep Sea Res Part II 42: 585-595

Wintzingerode FV, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev 21: 213-229

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.

Report of the thesis supervisor

Dr. Ramon Massana, as the supervisor of the PhD thesis entitled "**Culturing bias in the study of marine heterotrophic flagellates diversity**" presented by the PhD candidate Javier del Campo García-Ramos, informs about the impact factor of the journals where part of this thesis has been published, as well as the implication of the PhD candidate in each published paper.

Paper 1: "**Crash of a population of the marine heterotrophic flagellate *Cafeteria roenbergensis* by viral infection**", by Ramon Massana, Javier del Campo, Christian Dinter, and Ruben Sommaruga published in *Environmental Microbiology* in year 2007. This journal has an impact factor of 4.909 (2009 JCR Science Edition), appears in the first quartile of its corresponding category (15 of 95 in MICROBIOLOGY), and has the highest impact factor in the field of Microbial Ecology, our specific discipline (journals ranked between 1 and 14 publish review papers or clinical research). Each paper undergoes an exhaustive peer-reviewing process, which warrants the high quality standards of the manuscripts published in this journal.

This study was performed to investigate the effects of increasing organic matter amounts on the development of heterotrophic flagellates. One of the main findings was that at the highest organic matter addition there was a striking dominance of the culturable species *Caecitellus paraparvulus* and *Cafeteria roenbergensis*, both of which were present at very low abundance in the original sample. However, the most interesting issue of this study was a casual observation: the crash of the *Cafeteria roenbergensis* population due a specific and very lethal viral infection. This was one of the few studies that demonstrated that heterotrophic flagellates could also severely suffer from viral infection in certain circumstances. The PhD candidate incorporated at the ICM during the data processing of this study (the experiment was performed during a cruise at the Indian ocean in 2003) and participated in the FISH enumeration of these two species in the enrichments, and actively collaborated in the writing process. This paper was the perfect starting point for the present PhD research, which was further fully developed by the PhD candidate. He is the second author of this paper and the data presented is not included in any other doctoral thesis.

Paper 2: "**Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys**" by Javier del Campo, and Ramon Massana published in *Protist* in year 2011 (doi:10.1016/j.protis.2010.10.003). This journal has an impact factor of 3,853 (2009 JCR Science Edition) and appears in the first quartile of its corresponding category (22 of 95 in MICROBIOLOGY). Each paper undergoes an exhaustive peer-reviewing process, which warrants the high quality standards of the manuscripts published in this journal.

In this study we did an exhaustive screening of public databases in the search for environmental sequences related to three of the best-known taxonomic groups of cultured heterotrophic flagellates. This was used to evaluate their prevalence in marine and freshwater planktonic systems, to detect the existence of novel diversity within supposedly well-known groups, and to obtain the best up-to-date phylogeny representing these three groups, including 18S rDNA sequences from both cultured organisms and environmental surveys. Although these groups were not very abundant in molecular surveys, so probably did not account for the numerically dominant flagellates in natural communities, we identified several new clades composed by environmental sequences only. Striking differences in abundance and novelty were detected between the groups and marine and freshwater environments. Moreover, our phylogenetic trees will surely be used as seed for future phylogenetic analyses. This work was fully under the responsibility of the PhD candidate, who devised the initial scientific plan, did the in silico searches and the phylogeny analysis, and wrote the paper. The PhD candidate is the first and corresponding author of this paper and the data presented is not included in any other doctoral thesis.

Paper 3: " **New Insights into the diversity of marine picoeukaryotes**" by Fabrice Not, Javier del Campo, Vanessa Balagué, Colomban de Vargas, and Ramon Massana, published in *PLoS ONE* in year 2009. This is a relatively new online-only journal (the first paper appeared in December 2006) and has received the first impact factor only recently: 4,351 (2009 JCR Science Edition). This journal appears in the first quartile of its corresponding category (10 of 76 in BIOLOGY). Each paper undergoes an exhaustive peer-reviewing process, which warrants the high quality standards of the manuscripts published in this journal.

In this study we compared the diversity of marine picoeukaryotes obtained by standard clone libraries of 18S rDNA environmental sequences with two independent and complementary approaches, in order to improve our comprehension of the diversity of this ecologically important microbial component of marine ecosystems. In the first approach, the relative abundance of different taxonomic groups obtained from DNA-based clone libraries (reviewed in Massana and Pedrós-Alió 2008. Curr Opin Microbiol 11: 213-218), was compared with the same data obtained from GOS metagenomes. The striking similarity picture derived from both approaches suggested that the PCR biases, which did not apply in the metagenomic approach, were not pervading picoeukaryotic diversity studies. In the second

approach, we compared the diversity obtained from DNA-based and RNA-based libraries in the same sample. In this case substantial differences were found, which were interpreted as varying rDNA copy number among taxonomic groups. The responsibility of the PhD candidate in this study was the analysis of the GOS metagenomes to retrieve 18S rDNA sequences and the data elaboration of this part, as well as a general participation in the writing of the whole manuscript. This paper fits nicely with the scientific topic of the present PhD research. The PhD candidate is the second author of this paper and the data presented is not included in any other doctoral thesis.

Dr. Ramon Massana
Permanent researcher at the ICM - CSIC

Biaix de cultiu en l'estudi de la diversitat dels flagel·lats heterotròfics marins

Introducció als flagel·lats heterotròfics

350 anys d'història natural dels protozous

Els Protists (eucariotes unicel·lulars) són una de les principals forces impulsores dels ecosistemes del planeta i un excel·lent reservori de diversitat biològica (gens, molècules, vies metabòliques i processos cel·lulars) encara per descobrir i representen la major part de la diversitat eucariòtica coneguda. Són també actors principals en els processos macro i microevolutius de la vida a la Terra. Nosaltres hem centrat la nostra investigació en els protozous. Actualment sabem que el concepte original de protozous, que es refereix als organismes unicel·lulars amb un comportament animal (per exemple la motilitat) no té cap significat evolutiu o taxonòmic. En canvi, els protozous poden ser considerats com una classificació funcional i operativa dels protists heteròtrofs. Dins dels protozous, els flagel·lats heterotròfics són els que posseeixen un o més orgànuls anomenats flagels, que utilitzen tant per a la motilitat com per la depredació. El nostre coneixement sobre aquests organismes, inicialment considerats com petits animals, ha anat millorant durant els últims 350 anys, i molts científics rellevants han participat en aquesta aventura col·laborativa.

Els protists van ser descrits per primera vegada pel comerciant de Delft (Holanda) i pare de la microbiologia, Antoine van Leeuwenhoek. Utilitzant els seus primitius microscopis casolans, durant la segona meitat del segle XVII va ser capaç d'observar protists entre molts altres microorganismes. Algunes de les descripcions fetes per Leeuwenhoek corresponen en realitat a flagel·lats, probablement *Monas* spp. o *Bodo* spp.. Però va ser Otto Friedrich Müller, un zoòleg danès especialitzat en invertebrats marins, el primer a descriure formalment diverses espècies de flagel·lats sota el sistema de nomenclatura creat per Linné. Müller també va ser responsable de la taxonomia seminal dels protists heteròtrofs. El 1817 Georg Augustus Goldfuss va introduir el terme Protozous, però sense una definició adequada o una explicació. D'acord amb Goldfuss, els protozous incloïen els pòlips, les meduses, els infusoris i els phytozoa (animals-plantes i plantes-animals).

Al segle XIX sota la influència de la teoria cel·lular per Schelieden (1838) i Schwan (1839) diferents reconeguts naturalistes van observar i descriure el caràcter unicel·lular de diversos microorganismes eucariotes. Entre ells, Meyer i Dujarin van ser els pioners en estudiar protozous sota la influència del paradigma cel·lular i Barry i von Siebold van sistematitzar, explicar i endreçar tot aquest coneixement. Ferdinand Julius Cohn és reconegut com el responsable del terme Flagel·lats, per referir-se a alguns protozous amb flagell. També es va originar una de les grans dicotomies en l'estudi dels protists, quan va confirmar que la coloració verda d'alguns flagel·lats era deguda a la clorofil·la i que aquests organismes tenien la capacitat de fer fotosíntesi. Aquesta troballa va conduir a la inclusió de "plantes" i "animals" dins dels protists, però van ser classificats per separats en els regnes vegetal i animal.

Després de la publicació de "L'Origen de les Espècies" de Charles Darwin (1859) diversos treballs van nàixer sota el prisma de l'evolució establint les bases per a la taxonomia dels protists tal com la coneixem avui en dia. Els grans noms de la biologia, com Haeckel, Jans-Clark i Kent hi van contribuir amb els seus estudis sobre la relació entre les esponges i els coanoflagel·lats. Però hi ha un nom que cal destacar: Otto Bütschli, el gran arquitecte de la sistemàtica dels protozous. Ell va desenvolupar la primera sistemàtica exhaustiva, basada principalment en els trets locomotors, classificant totes les espècies conegudes en cinc classes: Sarcodina (amebes), Mastigophora (flagel·lats), Infusoria (ciliats), Sporozoa (apicomplexes) i Radiolaria (radiolaris). La seva classificació ha influït en la sistemàtica de protists durant tota la primera meitat del segle XX.

Els anys setanta van representar el final de la sistemàtica de Bütschli. Tot i ser molt convenient no reflectia pas relacions evolutives reals entre els organismes. La introducció de la microscopia electrònica en els estudis de protistologia i la facilitat per accedir a dades moleculars van contribuir clarament a aquest final. Un dels més influents protistòlegs dels anys 70 fins a finals del segle XX ha estat l'anglès Thomas Cavalier-Smith, que ha dut a terme un treball meticulós en la sistemàtica i la taxonomia dels protists basant-se en marcadors moleculars, principalment ADNr 18S. L'últim canvi de paradigma ha tingut lloc recentment, durant els primers anys del segle XXI amb la introducció de les biblioteques de clons ambientals de les comunitats de protists, obrint les portes de l'era "òmica" a la protistologia. El 2001 va aparèixer el primer estudi ambiental independent de cultiu en que les seqüències d'ADNr 18S s'utilitzaven per descriure la diversitat de protists marins en mostres de la columna d'aigua, millorant les filogènies conegudes. Aquests estudis seminals van posar en relleu la manca de coneixement que teníem de la diversitat de protists *in situ*.

Aquest és el punt on ens trobem ara, tenim la gran responsabilitat d'integrar aquest nou coneixement amb el valuós patrimoni que ens van deixar els grans científics com Cavalier-Smith, Bütschli o Müller. Ara sabem que els protists són presents en totes les branques de l'arbre de la vida eucariota i representen la forma de

vida més generalitzada dins dels eucariotes. Amb excepció de Archaeplastida, tots els supergrups també contenen protozous.

El paper dels flagel·lats heterotròfics

Dins dels protozous, és ben sabut que els flagel·lats heterotròfics juguen un paper clau en les xarxes tròfiques marines. Aquests organismes diminuts es distribueixen en ambients planctònics en concentracions d'entre 10^2 i 10^5 cèl·lules ml^{-1}, el que representa el 10-30% dels microorganismes eucariotes en aigües marines superficials. Són actors principals en els cicles biogeoquímics globals com a bacterívors i remineralitzadors de nutrients i constitueixen un conjunt d'espècies diverses, tot sovint mal identificades. Els flagel·lats heterotròfics són principalment fagòtrofs depredadors i controlen l'abundància de procariotes i petits eucariotes, però també poden incloure les etapes de dispersió de certs paràsits d'altres organismes marins. Aquest paper central en els ecosistemes marins s'ha traduït en un gran interès en el manteniment d'aquests organismes en condicions controlades de laboratori, per tal d'estudiar les taxes de depredació, les taxes de creixement, la preferència de presa, la ultraestructura, la genòmica o la transcriptòmica. Els cultius han estat essencials per a dur a terme estudis fisiològics i filogenètics, però la rellevància ecològica de les soques cultivades no està gaire clara. Els flagel·lats heterotròfics que tenim en cultiu representen realment els que són dominants en el medi natural?

L'estàndard d'or per l'estudi de la diversitat ambiental

En l'última dècada les biblioteques de clons d'ADNr 18S han esdevingut el mètode més estès per a l'estudi de la diversitat de protists en el medi ambient. Els treballs amb seqüències ambientals reflecteixen els membres dominants a les comunitats naturals i permeten revelar nous llinatges. Aquests estudis, majoritàriament realitzats en la fracció del picoplàncton (0,8-3 μm), han donat a conèixer nous grups d'alt rang taxonòmic, com ara els MALV (alveolats marins) o els MAST (estramenòpils marins) i les picobilífites, molts dels quals dominen la comunitat en termes d'abundància clonal. Aquesta diversitat descoberta recentment ha alterat significativament la descripció de les xarxes tròfiques microbianes marines i els nostres coneixements sobre l'evolució dels eucariotes.

No hi ha cap dubte que els estudis moleculars ofereixen una visió millorada de la diversitat *in situ* de protists petits en comparació amb les estratègies anteriors, basades fonamentalment en observacions microscòpiques o en l'aïllament en cultiu pur. No obstant això, no donen una resposta definitiva. La diversitat microbiana eucariota avaluada per mitjà de seqüències d'ADNr 18S de l'ambient ha generat patrons de composició molt similar en els diferents estudis realitzats fins ara, amb predomini dels grups no fotosintètics, incloent petits paràsits i bacterívors. Per contra, la microscòpia d'epifluorescència típicament revela un predomini en els oceans de cèl·lules fotosintètiques o mixotròfiques sobre les cèl·lules heterotròfiques en els oceans. Això suggereix que les biblioteques ADNr 18S poden presentar biaixos importants pel que fa a la diversitat de protists. Hi ha varies limitacions tècniques inherents a les tècniques independents de cultiu aplicades a l'estudi de la diversitat microbiana que cal destacar.

Són especialment rellevants la selectivitat dels encebadors de PCR, les variacions en el nombre de còpies de l'operó de l'ADNr i l'existència de pseudogens o d'ADN extracel·lular. Una alternativa prometedora que no requereix l'ús de la PCR és la metagenòmica, basada en la clonació directa i seqüenciació per mitjà de la tècnica de "shotgun" de l'ADN ambiental. S'ha de tenir en compte que fins ara els microorganismes eucariotes han rebut molt poca atenció en els estudis de metagenòmica. Un altre enfoc possible consisteix a dirigir-se directament a l'ARNr 18S (els propis ribosomes) com a una aproximació per a l'activitat i la diversitat metabòlica de les cèl·lules. Això evita l'efecte de les diferències en el nombre de còpies d'ADNr i la interferència de l'ADN dissolt en el medi.

El biaix de cultiu

La diversitat de protists coneguda i la informació que tenim de la seva biologia es basa generalment en soques cultivades, en les quals s'han dut a terme estudis ultraestructurals, fisiològics i moleculars. Els flagel·lats heterotròfics cultivats pertanyen a molts grups taxonòmics diferents i tenen representació en tots els supergrups de l'arbre dels eucariotes (sent els Archaeplastida l'única excepció). En mostres aquàtiques, les crisofícies, els coanoflagel·lats i els bicosoècids representen alguns dels grups aïllats amb més freqüència i sempre s'ha considerat que representen una fracció significativa dels flagel·lats heterotròfics. Cal destacar però, que les anàlisis moleculars ambientals han revelat llinatges totalment nous que fins ara no han estat mai cultivats. Per tant, dades fonamentals per a l'estudi de processos evolutius i ecològics ens han pogut passar inadvertides a causa de la nostra incapacitat d'obtenir aquests organismes en cultiu.

En l'actualitat, el biaix de cultiu segueix sent, sense cap mena de dubte, un dels reptes més importants als que s'enfronten els ecòlegs microbians amb l'objectiu d'aconseguir una plena comprensió de la funció ecològica dels microorganismes i suposa un coll d'ampolla en els estudis dels ecosistemes. Les biblioteques de clons ambientals demostren el baix grau en què els esforços fets per a cultivar han estat capaços de capturar la diversitat microbiana *in situ*. S'estima que tan sols entre 0,1 i un 1% dels microorganismes procariotes i eucariotes poden ser fàcilment cultivats. Irònicament, els tàxons més representats en el medi ambient es neguen a ser cultivats mentre que la majoria de les soques que trobem a les col·leccions de cultiu i als laboratoris són molt escasses en el medi ambient.

Els flagel·lats bacterívors s'han cultivat sempre en medis rics compostos per aigua de mar complementada amb grans de cereals o extracte de llevat, medis que promouen el creixement de bacteris en abundàncies molts superiors a les que trobem a la natura. Aquesta estratègia únicament permet l'aïllament, en la major part de les ocasions, d'un mateix tipus d'organismes, que acostumen a ser organismes com *Cafeteria* spp., *Paraphysomonas* spp. o *Bodo* spp., que són considerades, segons els estudis moleculars, espècies poc abundants al medi marí. Mentrestant, tàxons que per mitjà d'estudis moleculars representen organismes abundants al mar, com per exemple els bacterívors MAST, refusen el cultiu en aquestes condicions. Encara que el biaix de cultiu i l'efecte de la matèria orgànica en els enriquiments és un tema conegut (i sempre polèmic) en protistologia, no s'han publicat treballs en els que s'estudiï a fons el seu efecte sobre la comunitat microbiana eucariota. Estudis anteriors en que s'han analitzat les dinàmiques de protists en incubacions enriquides o sense enriquir s'han centrat sempre en l'evolució i les propietats de la comunitat al llarg del temps, però no han estat pas dissenyats per a tractar el tema del biaix de cultiu.

Aquest pot ser superat per mitjà de l'ús d'estratègies de cultiu originals, com s'ha demostrat pels procariotes *Pelagibacter ubique* i els crenarqueus marins, ambdós inicialment detectats a través d'estudis moleculars ambientals i posteriorment identificats com tàxons importants ecològicament. *Pelagibacter ubique* va ser cultivat simulant condicions oligotròfiques i els crenarqueus marins es van cultivar en medis rics en amoníac, després de que les dades moleculars revelessin que eren oxidadors de l'amoníac. Poques vegades s'han aplicat esforços semblants a l'aïllament i cultiu de protists marins. El biaix de cultiu és particularment greu per als protists heteròtrofs, que depenen d'una font de nutrients orgànica per al seu creixement, i es percep com una limitació important per seguir investigant el seu paper funcional i la seva importància ecològica.

Objectius

El nostre principal objectiu en aquest estudi va ser superar el biaix de cultiu en els flagel·lats heterotròfics marins. Aquest objectiu s'estructura en tres aspectes generals. El primer va ser determinar la importància i la representativitat dels flagel·lats cultivats en el medi ambient a partir d'estudis moleculars. El segon va ser l'estudi del biaix de cultiu des del punt de vista experimental. Finalment, el tercer va ser l'obtenció de nous cultius de flagel·lats heterotròfics. Per aconseguir-ho vam definir uns objectius més específics:

1. Determinar la contribució clonal de les seqüències d'ADNr 18S de crisofícies, coanoflagel·lats i bicosoècids en sistemes marins i d'aigua dolça, millorar la filogènia d'aquests grups i analitzar la seva novetat a nivell de seqüència.

2. Determinar l'efecte del biaixos induïts per la PCR mitjançant la comparació de les seqüències d'ADNr 18S de la base de dades de metagenòmica Global Ocean Survey(GOS) i de les biblioteques de clons estàndard.

3. Comparar la diversitat de protists apareguda a les biblioteques de clons a partir d'ADN i d'ARN extrets de la mateixa mostra, per tal de determinar els biaixos en els estudis de la diversitat ambiental basats en l'ADN.

4. Informar sobre els efectes de diferents enriquiments de matèria orgànica a l'estructura de la comunitat de flagel·lats heterotròfics i agrupar idees i conceptes sobre el biaix de cultiu que, acceptats o refusats, mai s'han acabat d'adreçar.

5. Desenvolupar un mètode de cultiu alternatiu per aïllar espècies de flagel·lats heterotròfics prèviament no cultivades que podrien ser abundants en el plàncton marí.

Aprofundint en els biaixos de cultiu en l'estudi de la diversitat dels protists heteròtrofs

Una infecció viral com a punt de partida

Amb la intenció d'observar l'efecte de la matèria orgànica sobre els flagel·lats heterotròfics en un enriquiment d'una comunitat de mar obert vam realitzar diferents incubacions d'una mostra de l'Oceà Índic Central. L'anàlisi de les seqüències i els resultats de FISH d'aquestes incubacions van mostrar que, en principalment *Cafeteria roenbergensis* i *Caecitellus paraparvulus* estaven creixent en les condicions enriquides. Aquestes dues espècies són flagel·lats heterotròfics ben coneguts i cultivats. *Caecitellus paraparvulus* va créixer primer i, probablement, va ser inicialment més abundant que *Cafeteria roenbergensis*, ja que es va detectar a la biblioteca de clons feta a partir de la mostra natural, però *Cafeteria roenbergensis* era el flagel·lat dominant al pic. Curiosament, els números de la població de *Cafeteria roenbergensis* van disminuir ràpidament després del pic i en el 8è dia estaven per sota el límit de detecció per FISH. Això estava relacionat amb la presència d'un virus infectant la població. El virus era específic per *Cafeteria roenbergensis*, ja que només les cèl·lules infectades d'aquest taxó observades per FISH estaven infectades. En el capítol 1, hem demostrat que un flagel·lat heterotròfic cultivat i un virus específics van ser seleccionats amb facilitat mitjançant l'enriquiment d'una mostra oceànica amb matèria orgànica. El virus es va propagar ràpidament, provocant la caiguda de tota la població. Aquest resultat concorda amb l'opinió general sobre la funció ecològica dels virus. Les nostres dades amplien la informació existent, ja que fins ara només dos virus que infecten a microeucariotes heteròtrofs s'han pogut aïllar i s'han mantingut en cultiu.

A més de l'interès intrínsec de la descripció del col·lapse d'una població de protozous a causa d'un virus, hi havia dos aspectes d'aquests enriquiments que van captar la nostra atenció i ens van portar a nous estudis. La primera va ser la presència insignificant de *Cecitellus paraparvulus* i *Cafeteria roenbergensis* en la mostra original, tot i ser els protists més abundants en l'enriquiment i dos dels flagel·lats més reportats en mostres marines. Vam decidir buscar la presencia d'aquests organismes i altres flagel·lats clàssicament rellevants en els estudis ambientals publicats, donant com a resultat el capítol 2. L'altre resultat interessant va ser el ràpid creixement de *Cafeteria roenbergensis* al mesocosmos enriquit, convertint-se en l'organisme dominant de la comunitat. Ja que aquest va ser el segon cop en que es descrivia aquest fet, Lim et al. 1999 van mostrar un cas similar amb *Paraphysomonas* sp., vam decidir demostrar que aquest biaix d'enriquiment és un fenomen habitual que es reprodueix en els processos de cultiu clàssics que s'han utilitzat durant anys. Aquesta preocupació va generar el capítol 4, que va continuar després al capítol 5.

Els més buscats

Com s'ha dit prèviament a la introducció, crisofícies, coanoflagel·lats i bicosoècids es consideren grups majoritaris dintre dels flagel·lats heterotròfics aquàtics, però no està clar si aquestes soques cultivades són ecològicament rellevants. Una cerca en la literatura sobre les espècies més comunament recuperades en estudis dependents de cultiu posa de manifest que aquests grups són trobats molt sovint. L'ADNr 18S de biblioteques ambientals ha donat a conèixer una gran diversitat i ha posat en relleu nous llinatges que van aparèixer en la majoria dels estudis en una alta abundància clonal. Les crisofícies, els coanoflagel·lats i els bicosoècids però, es troben representats amb poques seqüències en ambients marins i d'aigua dolça quan analitzem aquests estudis de forma individual.

No obstant això, les seqüències associades amb els grups de menor importància (en termes d'abundància clonal) sovint han estat poc analitzades, amagant una font potencialment rellevant d'informació filogenètica. En el capítol 2 vam fer un esforç per analitzar les seqüències afiliades a crisofícies, coanoflagel·lats i bicosoècids d'estudis moleculars ambientals existents en bases de dades públiques. L'abundància clonal relativa d'aquests tres grups suggeria que podien ser menys importants del que s'esperava en els sistemes marins, ja que només van representar el 5% dels clons, el que contrasta amb l'abundància clonal de MAST o MALV. Els arbres filogenètics a partir seqüències ambientals completes i seqüències procedents de soques cultivades permeten generar una representació més exacta de la diversitat d'aquests grups, amb l'aparició de diversos subtipus nous formats exclusivament per seqüències ambientals. Un exercici exhaustiu de mineria de dades va permetre la identificació de nova diversitat oculta dins de crisofícies, coanoflagel·lats i bicosoècids. Aquesta estratègia s'ha aplicat també als grups de protists com ara Mamiellophyceae.

Hem aplicat un nou enfocament per analitzar la novetat d'un determinat conjunt de dades basat en la similitud respecte les seqüències presents al GenBank. El grau de novetat mostrat per les seqüències ambientals de cada grup s'interpreta en termes d'esforços de cultiu i de seqüenciació en mostres ambientals. Una baixa correspondència entre les seqüències ambientals i seqüències obtingudes de cultius va ser la situació més comuna, i això posa de manifest el biaix de cultiu. D'altra banda, la seqüenciació de l'ADN del medi ambient és relativament senzill i hi ha poques possibilitats de perdre grans grups quantitativament importants filogenèticament. No obstant això, es va detectar en general un esforç de seqüenciació insuficient,

el que suggereix que resta força diversitat addicional per descobrir en aquests grups en mostres ambientals per mitjà de tècniques moleculars. Aprofitant les noves tecnologies de seqüenciació d'alt rendiment o amb l'ús del de primers específics podrem emplenar aquest buit de coneixement. D'altra banda, una altra explicació d'aquest esforç de seqüenciació insuficient seria un gran endemisme fent que les seqüències siguin molt diferents a la resta degut a que només se troben en certes zones d'estudi. Serà necessari augmentar els esforços de cultiu i la seqüenciació per tal d'arribar a una plena comprensió de la diversitat de protists en situ i la seva funció ecològica. La principal contribució per ampliar les bases de dades d'ADNr 18S seran els estudis d'alt rendiment que s'estan publicant o que es faran públics en un futur proper (per exemple el projecte BioMarks: www.biomarks.org). Les anàlisis filogenètiques com les realitzades al capítol 2 són fonamentals per proporcionar arbres de referència que seran utilitzats com a mapes filogenètics per a no perdre'ns dins d'aquest mar de dades cada dia més gran.

Noves perspectives sobre la diversitat de protists

La imatge actual de la biodiversitat marina eucariota pot ser alterada significativament pels biaixos en l'amplificació per mitjà de PCR, presència de gens d'ADNr multicòpia en una sola cèl·lula, i la capacitat de l'ADN per persistir com a material extracel·lular. Part d'aquest estudi l'hem dedicat a investigar com de greus són aquests biaixos i com l'estudi de la diversitat de protists per mitjà de tècniques moleculars es podria millorar. Considerem que aquesta informació és rellevant i interessant per als nostres objectius, perquè aquestes tècniques són instruments fonamentals per a la nostra investigació.

En el capítol 3 es va realitzar una anàlisi en profunditat del conjunt de dades de metagenòmica de l'expedició GOS, a la recerca de signatures eucariotes a través de la presència de gens d'ADNr 18S. La metagenòmica utilitza el clonatge i la seqüenciació per la tècnica de "shotgun" evitant el pas previ de PCR que si que s'utilitza en les biblioteques de clons. La similitud en els patrons de diversitat entre la clonació per PCR i l'aproximació metagenòmica suggereix un impacte baix de l'etapa de PCR en els resultats de les biblioteques de clons en termes de diversitat de seqüències i contribució relativa dels taxons específics. Una anàlisi per separat de les dues fraccions de mida (0,2 a 0,8 μm i 0,8 a 3 μm) del conjunt de dades del GOS va revelar clares diferències en termes de la composició taxonòmica. Com l'organisme eucariota més petit conegut té un diàmetre de 0,8 μm, algunes de les seqüencies d'ADNr 18S detectades en la fracció <0,8 μm poden derivar-se de protists molt petits, però probablement moltes d'elles deriven de restes cel·lulars o d'ADN extracel·lular de cèl·lules més grans. És probable que una fracció de l'ADN extracel·lular es retingui als filtres de 0,2 μm, en forma d'agregats moleculars o adsorbits. En conseqüència, creiem que és important tenir en compte la interferència d'ADN extracel·lular a l'hora d'avaluar la diversitat dels microbis eucariotes en els estudis ecològics.

Un altre mètode alternatiu per investigar la diversitat microbiana és apuntar directament l'ARNr 18S. Es van comparar els patrons filogenètics de biblioteques d'ADNr 18S i d'ARNr 18S a partir de la mateixa mostra del Mar Mediterrani. En general es considera que la diversitat reflectida en biblioteques de clons d'ADNr 18S no és una bona mesura quantitativa. L'anàlisi de la diversitat en base a l'ARNr va portar a una visió radicalment diferent de la comunitat en comparació amb el clàssic enfocament basat en l'ADN. Aquesta aproximació evita l'efecte de l'ADN multicòpia i la interferència de material extracel·lular, ja que l'ARN és molt menys estable que l'ADN. Hem trobat molt poca superposició en les seqüències recuperades de les biblioteques d'ADN i d'ARN. Pel que fa als protists heteròtrofs, les biblioteques d'ARNr apunten als MAST com els membres prominents de la comunitat, el que unit a la seva àmplia distribució suggereix que podrien ser els principals protists depredadors dels oceans. Les biblioteques de clons d'ARNr 18S ambientals semblen una via prometedora per reduir al mínim alguns biaixos importants, i així oferir noves perspectives en l'estudi de la diversitat i la funció dels protists marins.

Enfrontant-se al biaix cultiu

La paradoxa del biaix de cultiu apareix com un dubte raonable, darrere dels estudis ecològics, però mai ha estat abordat adequadament. L'estudi presentat al capítol 4 va ser dissenyat per tal de fer front a la mateixa analitzant l'efecte de la matèria orgànica en una comunitat natural de flagel·lats heterotròfics confinada. D'acord amb informació preliminar obtinguda del nostre primer treball i de la literatura es van establir diversos microcosmos amb una creixent quantitat de matèria orgànica i diferents fonts de procedència.

Els pics d'abundància de bacteris i flagel·lats heterotròfics van augmentar amb l'increment de matèria orgànica i això va anar acompanyat per un retard en l'aparició dels dos pics. Aquest retard pot ser degut al temps que necessita la comunitat per adaptar-se a les condicions d'enriquiment. En les incubacions enriquides amb matèria orgànica, les poblacions dominants originals, adaptades a concentracions més baixes de nutrients, poden veure el seu creixement inhibit, mentrestant alguns grups minoritaris, ben adaptats a altes concentracions de nutrients, podrien desenvolupar-se ràpidament i adoptar un paper predominant en

la comunitat. Un altre factor que podria endarrerir l'aparició del pic és l'augment del nombre de cèl·lules en les mostres enriquides, que necessitaven més temps per al creixement exponencial. El més probable és que les espècies de flagel·lats heterotròfics dominants originalment no estiguessin preparades per aquestes condicions d'enriquiment. Les diferències en la taxa de creixement són consistents amb les espècies que creixen en les diferents incubacions, amb un ràpid creixement de la població en condicions enriquides. D'altra banda, els bacteris de grans dimensions en el primer pic, juntament amb una gran proporció de bacteris agregats en els tractaments enriquits, són un refugi per evitar l'explotació per part dels flagel·lats heterotròfics. Només les espècies de flagel·lats heterotròfics adaptades a menjar bacteris grans de vida lliure o bacteris en els agregats seran capaces de proliferar i dominar en aquestes condicions.

Les dades obtingudes del pics de flagel·lats heterotròfics per DGGE i biblioteques de clons van revelar que els tractaments sense enriquir eren similars a la mostra original i molt diferent de la resta dels tractaments enriquits. Aquest resultat concorda amb estudis anteriors que van mostrar que les incubacions no enriquides promouen el creixement dels flagel·lats heteròtrofs presents en el conjunt natural i no modifiquen l'estructura de la comunitat. Els enriquiments per contra van promoure el creixement de poblacions no gaire abundants en la mostra original. La nostra anàlisi filogenètica va posar clarament de relleu l'efecte de biaix causat per la matèria orgànica. Mentre que a la biblioteca de clons no enriquida hi va haver un predomini dels protists no cultivats, l'augment de matèria orgànica invertia aquesta tendència, i els protists cultivables es van convertir en dominants en les incubacions més enriquides. El paper selectiu i homogeneïtzador de la matèria orgànica va ser confirmat per una clara disminució de la diversitat en incrementar la concentració de matèria orgànica (l'Índex de Shannon va passar de 2,5 a 1). Un enriquiment amb matèria orgànica dona com a resultat comunitats no només dominada pels organismes cultivats, però també menys diversa, dominada en el nostre cas per *Paraphysomonas* spp. i *Oikomonas* spp. Aquestes espècies tenen una gran capacitat per menjar bacteris grans en altes abundàncies i se sap que són comunament aïllades de la neu marina. Tenen doncs el potencial de competir directament amb els flagel·lats heterotròfics que originalment eren dominants en la mostra oligotròfica inicial.

Podem concloure que una de les principals raons del biaix de cultiu en els flagel·lats heterotròfics és l'ús de matèria orgànica en el procés d'aïllament, impulsant un canvi en la comunitat a condicions més pròximes als cultius de laboratori. Els esforços de cultiu fets fins ara han estat molt importants, però un nous impuls mitjançant noves estratègies de cultiu és necessari per avançar en la nostra comprensió de l'ecologia i l'evolució dels protists.

Buscant alternatives

Per tal d'obtenir en cultiu flagel·lats heterotròfics ecològicament rellevants es van imitar les condicions d'un medi marí oligotròfic mitjançant l'addició a aigua de mar estèril d'una barreja de bacteris naturals recollits en el mateix lloc de mostreig, en una abundància lleugerament superior a l'original. Cada precultiu es va iniciar amb una sola cèl·lula, que s'obtingué per dilució en sèrie o per mitjà de citometria de flux, i es van incubar en la foscor a temperatura *in situ*. Vam aconseguir aïllar soques pertanyents a llinatges que fins a dia d'avui només es coneixien a partir de seqüències ambientals. En aquest procés alguns precultius es van perdre i d'altres van evolucionar a espècies diferents.

Finalment es van obtenir dos cultius clonals i estables. Entre ells un Rhizaria uniflagel·lat relacionats amb els Chlorarachniophyta i un estramenòpil biflagel·lat llunyanament emparentat amb *Developayella* sp. La prospecció duta a terme per microscòpia electrònica de rastreig realitzada en els dos cultius estables va revelar cèl·lules molt petites amb poques característiques morfològiques. El Rhizaria aïllat sembla ser distant a qualsevol organisme descrit, la seva seqüència d'ADNr 18S només és un 90,6% similar a *Chlorarachnion reptans*, i era molt similar a seqüències recuperades de diferents ambients. La seva posició basal dins de *Chlorarachniophyta* juntament amb la seva naturalesa heterotròfica suggereix que representa un llinatge antic. En espera d'una descripció formal, aquest diminut flagel·lat heterotròfic ha estat batejat com a *Minorisa minuta* candidatus.

La resposta funcional de *Minorisa minuta* candidatus produeix una mitjana constant de saturació molt més baixa que la d'altres flagel·lats cultivats, la qual cosa suggereix que s'ha adaptat a viure en les concentracions de bacterioplàncton habituals en aigües oligotròfiques. Els recomptes obtinguts a partir de TSA-FISH, revelen *Minorisa minuta* candidatus com un component important dels flagel·lats heterotròfics marins a nivell mundial, àmpliament distribuïda i abundant. Representen el 1,8% dels flagel·lats heterotròfics a l'Atlàntic, Pacífic, Índic, l'Oceà Austral i el Mar Mediterrani, un valor que augmenta fins a un 5% quan es consideren només els llocs costaners. *Minorisa minuta* candidatus va ser detectat durant tot l'any en una estació d'aigües oligotròfiques de la costa del Mar Mediterrani NO amb una abundància d'entre 12 i 120 cèl·lules mL^{-1}, representant el 5% dels flagel·lats heterotròfics de mitjana. Les observacions al microscopi confirmen el seu caràcter picoeucariòtic amb una mida cèl·lular que varia entre 1 i 3 μm i una mida mitjana de 1,4 μm en mostres ambientals. Només *Symbiomonas scintillans* i *Picophagus flagellatus* s'acosten a aquesta grandària dins dels

flagel·lats heterotròfics, però tot i així *Minorisa minuta* representa el bacterívor més petit conegut. Però a banda de la reduïda mida d'aquest organisme és la seva àmplia distribució i la seva importància quantitativa en diferents ambients el que el fa tant interessant.

El mètode d'aïllament que es presenta en aquest treball defineix el camí a seguir en el futur en les prospeccions ambientals a la recerca de nous organismes per tal d'incrementar les col·leccions de cultius i el coneixement dels protists heteròtrofs.

Conclusions

Les conclusions generals d'aquesta tesi són els següents:

1. Hem mostrat que un flagel·lat heterotròfic cultivat i els seus virus específics van ser seleccionats amb facilitat mitjançant l'enriquiment amb matèria orgànica d'una mostra oceànica. El nostre estudi demostra que els virus també poden controlar les poblacions d'heteròtrofs flagel·lats. Aquesta estratègia també pot conduir a l'aïllament de nous virus d'eucariotes unicel·lulars marins.

2. El nostre estudi utilitzant seqüències ambientals procedents de bases de dades públiques posa de manifest una gran diversitat emergent (de vegades nova) de crisofícies, coanoflagel·lats i bicosoècids en ambients aquàtics. Això indica un biaix en la representació dels cultius i una manca d'esforç de seqüenciació per a aquests grups. La diversitat i la novetat observada és sorprenent fins i tot per als grups considerats protists ben caracteritzats. Aquesta anàlisi es pot estendre a altres grups per tal de beneficiar-se plenament de les biblioteques de clons ambientals.

3. La nostra anàlisi de les seqüències d'ADNr 18S recuperades del conjunt de dades metagenòmiques del GOS no mostren diferències substancials respecte a les biblioteques de clons basades en PCR. Els resultats similars d'ambdós aproximacions suggereixen poc impacte del pas de PCR a les biblioteques clon en termes de diversitat de les seqüències i de la contribució relativa de taxons específics.

4. Les diferències detectades entre les biblioteques d'ADN i ARN a partir de la mateixa mostra són consistents amb l'opinió que les biblioteques d'ARNr permeten reduir significativament dos dels principals biaixos en estudis de diversitat molecular basats en ADNr: el nombre de còpies d'ADNr i la presència d'ADN extracel·lular. Tenint en compte la relativa facilitat per treballar amb ARN, ampliar la realització de biblioteques de clons basades en ARNr, proporcionarà valuosa informació sobre l'ecologia de les espècies no cultivades. Si s'acompanya d'un major esforç de seqüenciació és probable que s'aconsegueixi una visió gairebé exhaustiva de la diversitat dels protists.

5. Els enriquiments amb matèria orgànica modifiquen la dinàmica de la comunitat a causa d'un procés de substitució dels bacteris i els flagel·lats heterotròfics dins la comunitat inicial. Hi ha una important disminució en la diversitat i un augment gradual dels organismes cultivats quan augmenta la matèria orgànica.

6. Les tècniques clàssiques de cultiu basades en l'enriquiment han de ser complementades amb noves estratègies de cultiu si realment es volen obtenir els organismes responsables de la majoria dels processos al mar. Tot i la importància dels cultius obtinguts fins avui, hem d'ampliar i fer créixer les nostres col·leccions amb nous cultius de rellevància ambiental.

7. Hem dissenyat un nou mètode d'aïllament i cultiu de flagel·lats heterotròfics basant-nos en l'ús de bacteris naturals com a font d'aliment. Fent servir aquesta aproximació s'han aïllat diversos protists petits que pertanyen a tàxons prèviament no cultivats i pertanyents a llinatges distants dins l'arbre eucariota de la vida. Quan s'apliqui a diferents escales temporals i espacials, aquesta estratègia potencialment donarà accés a una gran quantitat de protists heteròtrofs en cultiu.

8. El petit uniflagel·lat *Minorisa minuta* candidatus es reivindica com el més petit dels bacterívors conegut fins ara. D'altra banda, és l'únic representant heteròtrof dins del llinatge Chlorarachniophyta i és de primordial interès per estudiar la transició a l'endosimbiosi secundària dels plàstids. Com per als seus homòlegs fotosintètics en els oceans, l'anàlisi del genoma de *Minorisa minuta* candidatus sens dubte revelarà vies cel·lulars, bioquímiques i evolutives sense precedents.

9. *Minorisa minuta* candidatus està amplament distribuït pels oceans i és un membre important de la població de flagel·lats heterotròfics, sobretot en les aigües costaneres. Les propietats fisiològiques de *Minorisa minuta* candidatus pot explicar el seu èxit ecològic i establir aquesta espècie com un bon model per als flagel·lats heteròtrofs dominants en aigües marines, aquests paràmetres poden ser utilitzats per millorar els models ecològics. Obtenir en cultiu un bacteri de rellevància ambiental com *Pelagibacter ubique* candidatus va suposar un salt endavant cap a una millor comprensió de la funció dels microbis en els oceans i va obrir diverses noves vies de recerca. La domesticació de petits depredadors marins amb importància ecològica és una promesa per a futurs descobriments similars.

1

Col·lapse d'una població del flagel·lat heterotròfic marí *Cafeteria reonbergensis* degut a l'acció d'un virus

Els virus són importants agents causants de mortalitat en microorganismes marins. La majoria d'estudis es centren en els virus de bacteris i algues, i existeixen alguns informes sobre virus que infecten protists heteròtrofs marins. En aquest estudi mostrem els resultats de l'anàlisi de la comunitat microbiana en diverses incubacions de mostres procedents de l'Oceà Índic Central enriquides amb diferents quantitats de matèria orgànica. Els flagel·lats heterotròfics van arribar fins a les 30.000 cèl·lules ml^{-1} en la incubació més enriquida. Una biblioteca de clons de ADNr 18S i recomptes de FISH utilitzant sondes dissenyades expressament per aquesta estudi van permetre observar que el pic de flagel·lats estava format per *Cafeteria roenbergensis* i *Caecitellus paraparvulus* (90% i 10% de les cèl·lules, respectivament). Ambdós tàxons estaven per sota del límit de detecció en la mostra original, el que indica un fort biaix al seu favor a l'enriquiment. Durant el pic, al citoplasma de les cèl·lules de *Cafeteria roenbergensis* s'hi van observar partícules similars a virus, i en els quatre últims dies de l'enriquiment no vam poder detectar cèl·lules pertanyents a aquest organisme. Mitjançant microscòpia electrònica de transmissió vam confirmar la naturalesa viral d'aquestes partícules. Eren relativament grans (280 nm), amb ADN de doble cadena. Es tracta d'un virus específic de *Cafeteria roenbergensis* ja que ni *Caecitellus paraparvulus* ni cap dels altres tàxons de fagel·lats que van aparèixer en algun moment a la incubació se'n van veure afectats. Aquest és un dels pocs informes existents sobre un virus que ataca a flagel·lats heterotròfics i en discutim les seves implicacions.

Environmental Microbiology (2007) **9**(11), 2660–2669 doi:10.1111/j.1462-2920.2007.01378.x

Crash of a population of the marine heterotrophic flagellate *Cafeteria roenbergensis* by viral infection

Ramon Massana,[1]* Javier del Campo,[1] Christian Dinter[1] and Ruben Sommaruga[2]
[1]*Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC. Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.*
[2]*Laboratory of Aquatic Photobiology and Plankton Ecology, Institute of Ecology, University of Innsbruck. Technikerstr. 25, 6020 Innsbruck, Austria.*

Summary

Viruses are known as important mortality agents of marine microorganisms. Most studies focus on bacterial and algal viruses, and few reports exist on viruses infecting marine heterotrophic protists. Here we show results from several incubations initiated with a microbial assemblage from the central Indian Ocean and amended with different amounts of organic matter. Heterotrophic flagellates developed up to 30 000 cells ml^{-1} in the most enriched incubation. A 18S rDNA clone library and fluorescent *in situ* hybridization counts with newly designed probes indicated that the peak was formed by *Cafeteria roenbergensis* and *Caecitellus paraparvulus* (90% and 10% of the cells respectively). Both taxa were below detection in the original sample, indicating a strong positive selective bias during the enrichment. During the peak, *C. roenbergensis* cells were observed with virus-like particles in the cytoplasm, and 4 days later this taxa could not be detected. Transmission electron microscopy confirmed the viral nature of these particles, which were large (280 nm), had double-stranded DNA, and were produced with a burst size of ~70. This virus was specific of *C. roenbergensis* as neither *C. paraparvulus* that was never seen infected, nor other flagellate taxa that developed in later stages of the incubation, appeared attacked. This is one of the few reports on a heterotrophic flagellate virus and the implications of this finding in the Indian Ocean are discussed.

Received 27 March, 2007; accepted 24 May, 2007. *For correspondence. E-mail ramonm@icm.csic.es; Tel. (+34) 93 2309500; Fax (+34) 93 2309555.

Introduction

Viruses are abundant and ubiquous members of marine ecosystems, infecting probably all living beings from bacteria to whales, and playing different ecological roles (Suttle, 2005). By lysing marine microorganisms which are at the base of food webs, viruses can compromise the trophic transfer of energy and organic matter, stimulate respiration and nutrient regeneration, and thus influence global biogeochemical cycles (Fuhrman, 1999). Because viruses are generally host-specific, they can affect the dynamics of given populations (Brussaard, 2004) and modulate the diversity of natural assemblages (Wommack and Colwell, 2000). In fact, according to the 'killing the winner' hypothesis (Thingstad and Lignell, 1997), they are responsible for maintaining the high diversity of microbial assemblages by suppressing the most successful population. Viruses have also played an important role in evolution and diversification in prokaryotes, being vectors of extensive horizontal gene transfer among different evolutive lineages (Ochman *et al*., 2000; Weinbauer and Rassoulzadegan, 2004). Moreover, viruses appear to harbour an enormous genetic and biological diversity that remains largely undersampled (Man *et al*., 2003). Therefore, it is widely accepted that viruses are active and important components of marine microbial communities.

Most virus-like particles (VLP) found in aquatic systems are very small, typically below 100 nm in capsid size, and are considered to infect mostly bacteria (Weinbauer, 2004). Large filamentous VLP infecting aquatic bacteria have been observed (Hofer and Sommaruga, 2001), but they are uncommon in marine waters (Middelboe *et al*., 2003). Bacteriophage ecology has typically focus on estimating and explaining their abundance and their impact on bacterial production and diversity (Weinbauer, 2004). This focus provides fundamental insights on the ecological role of phages, but hides the complexity of specific viral–host interactions, because both viral and bacterial assemblages are considered as homogeneous groups (the so-called black-box approach). An exception of this 'black-box' approach on phage ecology are studies on viruses infecting specifically the cyanobacteria *Synechococcus* and *Prochlorococcus* (Suttle and Chan, 1994; Sullivan *et al*., 2003). Marine eukaryotic viruses, on the other hand, are typically larger, less abundant, and their study is generally based on specific viral–host interactions

(Brussaard, 2004). Viral impact on species from most important algal groups has been seen in the field, with viruses preventing or terminating algal blooms. Some examples are reports of viral control on populations of the DMS-producing *Emiliania huxleyi* (Bratbak *et al*., 1993), the harmful bloom-forming *Heterosigma akashiwo* (Tarutani *et al*., 2000), or the abundant coastal picoeukaryote *Micromonas pusilla* (Cottrell and Suttle, 1995). Viruses infecting these and other algal taxa have been isolated and maintained in culture, where their lytic cycle, infectivity range, molecular features and gene content are characterized (Wilson *et al*., 2005; Zingone *et al*., 2006).

Compared with the numerous studies on marine algal viruses very little is known on viruses infecting marine heterotrophic protists. The most abundant of these protists are heterotrophic flagellates of sizes between 2 and 5 mm, which play central roles in microbial food webs as picoplankton grazers (Sherr and Sherr, 2002; Pernthaler, 2005). Even though there are many taxa in culture (Patterson and Larson, 1991), their *in situ* diversity remains largely unknown (Arndt *et al*., 2000), and uncultured groups are relatively abundant in marine samples (Massana *et al*., 2006a). This might explain why there are so few studies of heterotrophic flagellate viruses. One study reported the occurrence of VLP within an unpigmented flagellate associated with a dinoflagellate bloom (Nagasaki *et al*., 1993). Up to 20% of cells appeared infected, and the authors suggested that the virus regulated the dynamics of this population. In another study, VLP were seen within four chrysophyte species and were responsible for their abrupt disappearance in enrichment cultures (Preisig and Hibberd, 1984). To our knowledge, only one virus infecting a heterotrophic flagellate has been isolated and maintained in culture (Garza and Suttle, 1995). This virus was specific for *Bodo* sp. and could diminish cultures of this flagellate. Recently, a virus infecting a marine thraustochytrid has been isolated (Takao *et al*., 2005), which is not strictly a grazer flagellate but a fungoid protist.

Here we report the presence of a virus during an enrichment of heterotrophic flagellates from the central Indian Ocean. We compared the taxa developing in incubations initiated from the same sample but with different levels of organic matter. Flagellates growing in unamended incubations were those that dominated natural assemblages (Massana *et al*., 2006b), and we hypothesized that typical cultured flagellates would grow in enriched incubations. In the most enriched incubation, we observed that the heterotrophic flagellate assemblage suddenly crashed concomitantly with the presence of VLP, first inside the cells and later free in the water. By combining epifluorescence microscopy, flow cytometry, electron microscopy and molecular techniques we assessed whether VLP were in fact viruses, which flagellate taxa was the host of the virus, and the dynamics of both host and viral populations.

Results

Dark incubations of 3 mm filtered surface seawater were initiated in one station at the central Indian Ocean. One control (unamended, OA) and three enriched incubations were prepared with different amounts of rice media: 0.2% v/v (OB), 1% v/v (OC) and 4% v/v (OD). The dynamics of bacteria and protists were followed daily by epifluorescence microscopy (Fig. 1). In the first two incubations (OA and OB) there were minor changes in microbial counts (Fig. 1A), with heterotrophic bacteria fluctuating between 0.7 and 1.0 ¥ 10^6 cells ml^{-1} and heterotrophic flagellates between 0.3 and 1.1 ¥ 10^3 cells ml^{-1}. The effect of the rice media was obvious in the other two cases, with bacterial numbers reaching up to 2.5 ¥ 10^6 cells ml^{-1} and peaks of heterotrophic flagellates (6.5 and 30.4 ¥ 10^3 cells ml^{-1} in OC and OD respectively), which lasted very shortly. Bacterial numbers were suppressed by heterotrophic flagellates only in OD. In all cases, phototrophic flagellates and *Synechococcus* decreased continuously, consistent with the dark incubation (Fig. 1B).

During the epifluorescence microscopic counts, we noticed that some heterotrophic flagellates had the cytoplasm full of regular particles that fluoresced brightly after DAPI staining (Fig. 2A). These VLP were large enough to be retained in the 0.2 mm pore-size filter to count bacteria and were confused initially with very small cocci. They were particularly obvious in OD at day 5, when most heterotrophic flagellates appeared infected (72% of the cells) with a relatively uniform number of VLP (68.5 on average; SE = 2.7; $n = 20$). Fixed flagellates from this sample were concentrated, sectioned and observed by transmission electron microscopy (TEM). Up to eight regular particles were clearly visible inside the cytoplasm of a 3-mm protist, confirming that these particles were indeed viruses (Fig. 2B). They showed an hexagonal or pentagonal profile in section and were therefore probably icosahedral in three-dimensional morphology. Their estimated size was 280 nm.

As reference for flow cytometric analyses, we used a suspension of T_4 (bacteriophage) and EhV-86 (*E. huxleyi* virus), which were detected as distinct populations on flow cytograms (R1 and R2 in Fig. 3A). Virus-like particles appearing in the OD incubation (R3 in Fig. 3B) were identified as a population with side scatter and fluorescence values larger than EhV-86 but lower than heterotrophic bacteria (R4 in Fig. 3B). Flow cytometry counts of VLP with these settings were below detection in OA and OB samples and low in OC: 7.1 ¥ 10^4 VLP ml^{-1} at day 7 (data not shown). In OD, on the other hand, VLP suddenly increased from day 4 to day 6 and remained high

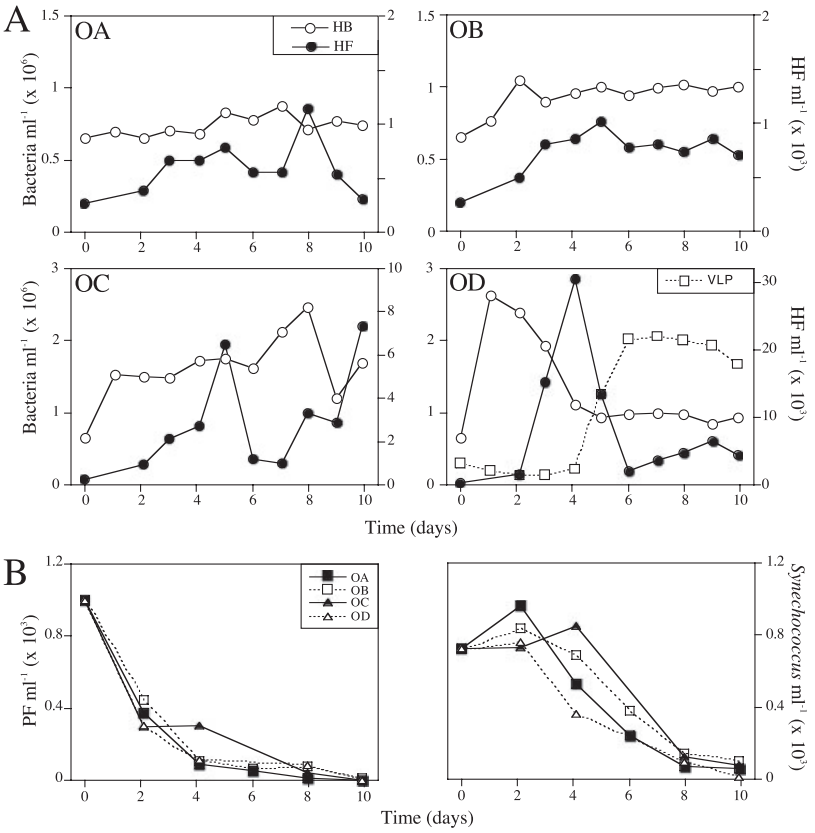


Fig. 1. A. Abundance of heterotrophic bacteria (HB) and heterotrophic flagellates (HF) during the incubations with different amounts of rice media: 0% (OA), 0.2% (OB), 1% (OC) and 4% (OD). Abundance of virus-like particles (VLP, same scale as HB) is shown only in OD.
B. Abundance of phototrophic flagellates (PF, left panel) and *Synechococcus* (right panel) during the four incubations.

(2×10^6 VLP ml^{-1}) for the rest of the incubation (Fig. 1A). The increase in VLP was concomitant to a sudden decrease in the abundance of heterotrophic flagellates.

We then analysed the changes in protist diversity in the four incubations by using the fingerprinting technique DGGE based on 18S rDNAs (Fig. 4A). The high diversity observed in the initial sample was roughly maintained in OA and OB, incubations that presented few changes and shared many bands with the initial sample. Conversely, in OC and OD there was a significant simplification of the protist assemblage, which ended the incubation dominated by a few bands barely detectable at the beginning. This overall trend was exemplified by the grouping of samples in the dendrogram derived from the DGGE fingerprint (Fig. 4B). There was a clear grouping of the three initial samples (IND58, OA0 and OD0) and all samples from OA and OB, indicating minor changes of diversity during the 8 days of the incubation in these treatments. On the other hand, all samples from OC and OD formed a separate cluster, with samples from the intermediate times (days 4 and 6) grouping together in each incubation.

To identify the heterotrophic flagellates forming the largest peak in our study, we constructed a 18S rDNA clone library from sample OD at day 4. A restriction fragment length polymorphism (RFLP) analysis of the first 26 clones from the OD4 library indicated a very low diversity, with a pattern repeated 21 times and another repeated three times. One clone from each RFLP pattern was fully sequenced (the dominant pattern was sequenced twice to confirm their phylogenetic position: clones OD4.1 and OD4.2), and all of them affiliated within the bicosoecids (Fig. 5). The OD4 library was dominated by *Cafeteria roenbergensis* (22 clones out of 26), whereas the remaining four clones affiliated with *Caecitellus paraparvulus*, represented by two slightly different phylotypes. A library from the *in situ* sample (IND58) yielded four clones related

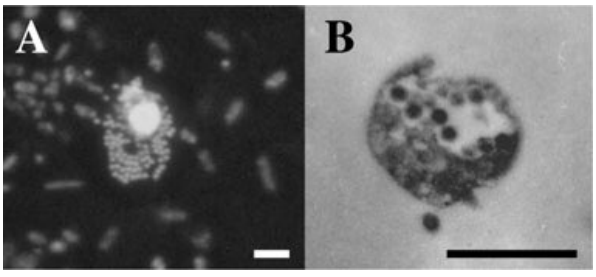


Fig. 2. Micrographs of a *Cafeteria roenbergensis* cell infected by viruses observed by DAPI staining and epifluorescence microscopy (A) and transmission electron microscopy (B). Cells are from the incubation OD at day 5. The scale bar is 2 nm in both pictures.

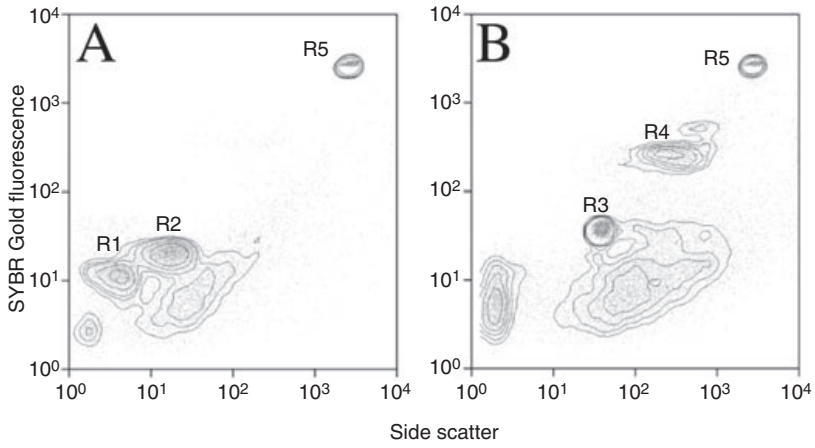


Fig. 3. Flow cytometric analysis of a mix of two isolated viruses (A) and the sample OD at day 10 (B). R1: T4-phage; R2: *Emiliania huxleyi* virus EhV-86; R3: *Cafeteria roenbergensis* virus; R4: Heterotrophic bacteria; R5: Fluorescent microspheres.

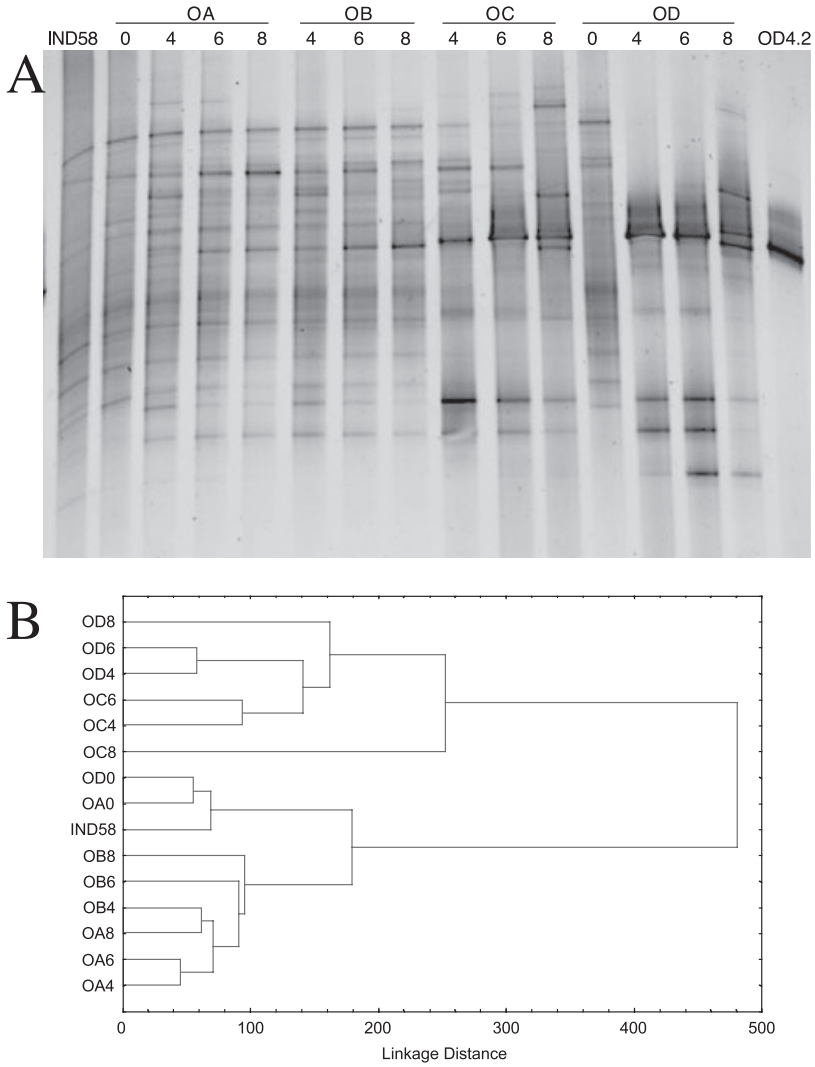


Fig. 4. A. Inverted image of a DGGE gel showing the fingerprint of eukaryotic populations at selected days in the four incubations. The fingerprint from the *in situ* sample (IND58) and from the dominating clone in the OD4 library (OD4.2) is also shown.
B. Cluster analysis relating DGGE fingerprints.

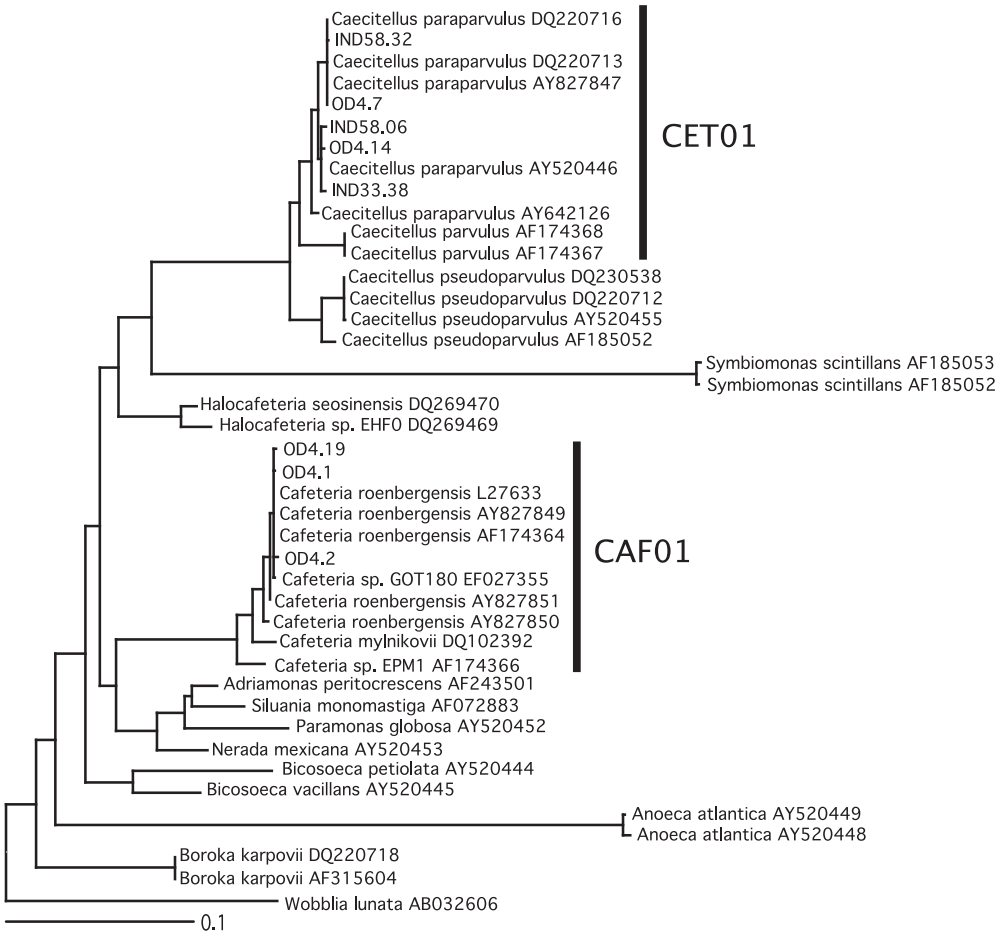


Fig. 5. Phylogenetic tree of 18S rDNA sequences from representative bicosoecids and clones retrieved in the OD4 library and the IND58 library (*in situ* sample). Clone IND33.38 is from another Indian Ocean station. The coverage of the FISH probes designed here is shown.

to *C. paraparvulus* (out of 62 clones), three related to OD4.7 and one to OD4.14. Three clones from the dominant RFLP pattern in OD4 (affiliating with *C. roenbergensis*) were run simultaneously in the DGGE with the samples from the incubations. The three clones showed the same mobility (only OD4.2 is shown in Fig. 4A) and coincided with a bright DGGE band in the OC (days 4–8) and OD (days 4–6) incubations.

Oligonucleotide probes against *C. roenbergensis* and *Caecitellus* spp. (probe coverage in Fig. 5) were designed and applied to quantify these flagellates by fluorescent *in situ* hybridization (FISH) in the incubations OA (unamended) and OD (largest enrichment). We assumed that counts with probe CET01 in this study accounted for *C. paraparvulus* cells, because only sequences from this species were retrieved in libraries OD4 and IND58. Both flagellates were below detection levels in the initial sample and did not develop (*C. roenbergensis*), or only slightly (*C. paraparvulus*, up to 20 cells ml^{-1}), in the OA incubation. However, both taxa reacted very fast in OD (Fig. 6). *Caecitellus paraparvulus* grew first, accounting for 50% of eukaryotic FISH counts at day 2, whereas *C. roenbergensis* dominated at day 4, during the peak of heterotrophic flagellates, accounting for about 90% of eukaryotic cells. The maximum abundance for both taxa was 2070 and 22 900 cells ml^{-1} respectively. The combination of FISH staining (orange cytoplasm under green light) and DAPI staining (blue nucleus and VLP under UV radiation) revealed that the cells infected by the virus belonged exclusively to *C. roenbergensis*.

After the peak of heterotrophic flagellates in OD, *C. roenbergensis* and *C. paraparvulus* cells decreased (both in numbers and in contribution to eukaryotic cells), and at day 8 were undetected by FISH. The dynamics between days 4 and 8 of *C. roenbergensis* (from 2.3×10^4 to 0 cells ml^{-1}) and VLP (from 2.2 to 20.0×10^5 particles ml^{-1}) allowed a rough estimation of the viral burst size, 77 VLP per lytic event, close to the value estimated from counting VLP in infected cells by epifluorescence. A clone library from sample OD at day 8 revealed that the

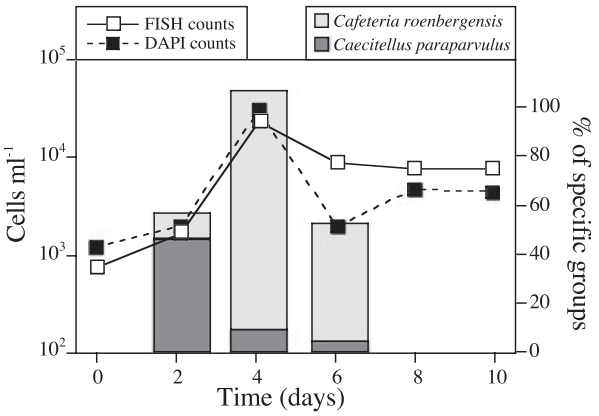


Fig. 6. Abundance of protists during the OD incubation estimated by DAPI (HF plus PF) and FISH counts (eukaryotic probe), and percentage of the later accounted by FISH counts with the specific probes CAF01 and CET01. Bars not seen (days 0, 8 and 10) are near zero.

heterotrophic flagellate assemblage was very different from that at day 4: *Paraphysomonas butcheri* dominated (11 of 14 clones), together with other taxa such as *Amastigomonas debruynei, Ancyromonas sigmoides*, and an unidentified stramenopile. No bicosoecid sequence was found in this library. This revealed a fast replacement of flagellate taxa, apparently unaffected by the *C. roenbergensis* virus, during the OD incubation.

Discussion

Heterotrophic flagellates play a key role in marine food webs as already highlighted in the seminal paper on the microbial loop concept (Azam *et al*., 1983). However, their *in situ* diversity remains largely unknown (Arndt *et al*., 2000; Sherr and Sherr, 2002), and it appears to be significantly composed by groups without cultured representative (Massana *et al*., 2006a). These uncultured taxa can grow in unamended incubations (Massana *et al*., 2006b), and we hypothesize that they would not grow when adding organic matter, as typically done during culturing attempts. In the present study, no flagellate growth was observed in the unamended incubation (OA) and the one with the lowest enrichment (OB), likely because the initial sample was very oligotrophic (Massana *et al*., 2006b). In contrast, the addition of relatively little organic matter (only 1–4% of what is used in culturing; incubations OC and OD) promoted a large development of bacteria and heterotrophic flagellates. Sequencing and FISH analyses from the OD incubation showed that the flagellate peak was formed essentially by *Cafeteria* and *Caecitellus*, well-known cultured genera that appear in the list of the 20 most commonly reported heterotrophic flagellates (Patterson and Lee, 2000). We confirmed the hypothesis that cultured taxa would grow under enriched conditions and uncultured taxa only in unamended conditions (Massana *et al*., 2006b). *Cafeteria* and *Caecitellus* cells were not detected by FISH in the initial sample, so they seem to belong to the rare 'biosphere' of microbial assemblages (Pedrós-Alió, 2006), taxa that are present at very low abundance but can be retrieved by selective culturing.

During the OD incubation, *C. paraparvulus* grew first and probably was initially more abundant than *C. roenbergensis*, because it was detected in the IND58 library (four clones out of 62). However, it only accounted for 10% of cells at the peak at day 4. The dominant flagellate at the peak was *C. roenbergensis*, which accounted for 90% of the cells. This flagellate was not detected initially (neither by FISH nor in the clone library) and grew very fast in the incubation. A very rough estimate using only two points (days 2 and 4) resulted in a growth rate of 2.4 day^{-1} and an initial abundance of 1.5 cells ml^{-1}. This growth rate is lower but comparable to the highest rate measured in a *Cafeteria* sp. culture, 3.5 day^{-1} (Boenigk *et al*., 2007). An intriguing observation is that the three complete 18S rDNA sequences of *C. roenbergensis* were not identical (despite two had the same RFLP pattern), but showed 5–6 differences in the 1716 bp length. These differences are minor but very unlikely from the same cell, meaning that a mixed assemblage of *C. roenbergensis* was present in the initial sample and developed together during the incubation. The ecological significance of this microdiversity is unknown, although it seems to be a property of marine microbial populations (Acinas *et al*., 2004). After the population peak of *C. roenbergensis* at day 4, numbers rapidly decreased and at day 8 were below detection by FISH. This appeared to be related to the presence of VLP infecting the population.

These VLP were initially seen by epifluorescence microscopy as uniform dots that fluoresced brightly after DAPI staining, indicating they contained double-stranded (ds) DNA. By flow cytometry they appeared larger and with more DNA than the two reference viruses used. Ultrastructural analysis by TEM showed icosahedral particles inside flagellate cytoplasm. Their capsid size was around 280 nm, somewhat larger that many dsDNA algal viruses (Brussaard, 2004) but not uncommon (Van Etten and Meints, 1999). The virus was specific to *C. roenbergensis*. Only cells from this taxa were infected after FISH inspection (e.g. *C. paraparvulus* never was), and other flagellate taxa (such as *P. butcheri*) developed moderately after *C. roenbergensis* even though VLP numbers remained high until the end of the incubation. These constant high numbers suggest that the viruses were not significantly grazed by heterotrophic flagellates. Thus, in this case viruses did not seem to be food for flagellates (González and Suttle, 1993). The virus described here is very similar (in capsid size, shape and nucleic acid content) to the only

heterotrophic flagellate virus isolated so far (Garza and Suttle, 1995). In fact, the host of the latter virus has now been properly identified as *Cafeteria* sp. (C. Suttle, pers. comm.), which opens the possibility that it might be the same virus, one isolated in the Gulf of Mexico and the present one in the central Indian Ocean. The existence of viruses with a broad distribution in the marine environment has already been described (Short and Suttle, 2002), and fits with similar observations for prokaryotic (Morris *et al*., 2002) and eukaryotic assemblages (Massana *et al*., 2006a).

In this study, we have shown that a cultured heterotrophic flagellate and its specific virus were readily selected by enriching an oceanic sample with organic matter. The virus could spread fastly and crash completely the host population. The virus was host-specific as it did not affect the dynamics of other flagellate taxa even from the same phylogenetic group. This result is in agreement with the general view of the ecological role of viruses: they exist for virtually all protist taxa, are host-specific and density dependent, and can control population dynamics, as has been seen terminating algal blooms (Bratbak *et al*., 1993; Tarutani *et al*., 2000; Brussaard, 2004). Our study shows that heterotrophic flagellate populations can also be controlled by viruses (Garza and Suttle, 1995). We consider our observation as casual, because there are many reports of *C. roenbergensis* cultures and enrichments where viruses have not been detected. It is also puzzling that a natural assemblage from a remote and oligotrophic place such as the central Indian Ocean contained a virus specific for a rare protist population. A possible explanation is that the virus was lysogenic (i.e. its genome was integrated into that of the flagellate) and switched to a lytic cycle due to the flagellate growth. Cells which are nutrient-limited have usually insufficient energy available for the virus to initiate the lytic cycle. Moreover, lysogenic bacterial viruses seem to dominate in oligotrophic marine regions (Weinbauer, 2004). Alternatively, the virus could be part of a rare 'virosphere' waiting for their opportunity to develop. At any rate, we report the occurrence of a virus terminating a bloom, induced from a natural sample, of the heterotrophic flagellate *C. roenbergensis*.

Experimental procedures

Sampling and seawater incubations

Sampling was performed in the central Indian Ocean (17°10.55′S, 83°40.51′E; maximum depth 5646 m) on 1 June 2003 on board of the R/V *Melville* (Scripps Institution of Oceanography). Surface (5 m) seawater was collected with a Niskin bottle attached to a CTD rosette. Temperature at this depth was 25°C. Seawater was filtered by gravity first through a nylon mesh of 200 μm and later through polycarbonate filters of 3 μm pore-size and dispensed in 2-l Nalgene polycarbonate bottles. Bottles were supplemented with different amounts of rice media (supernatant obtained after autoclaving 1 l of seawater with 40 grains of rice) to promote bacterial growth that serve as food for heterotrophic flagellates. All bottles were incubated in the dark inside a tank on the ship deck continuously fed with surface seawater.

Subsamples for flow cytometry and epifluorescence microscopy were taken daily during 10 days. Aliquotes for flow cytometry were fixed with 1% paraformaldehyde and 0.05% glutaraldehyde (final concentrations), kept in the dark for 10 min and deep-frozen in liquid nitrogen. Aliquotes for epifluorescence microscopy were fixed with ice-cold glutaraldehyde (1% final concentration), stained with DAPI and filtered through 0.2 and 0.6 μm pore-size black polycarbonate filters for counting heterotrophic bacteria (potentially including archaea and *Prochlorococcus*), *Synechococcus* and flagellates (heterotrophic and phototrophic). Counts were done in the laboratory during the first 2 months after sampling. Subsamples for molecular analyses were taken every other day. Aliquotes for microbial biomass (60–150 ml) were filtered on 0.2 μm pore-size Durapore filters, submerged in lysis buffer (40 mM EDTA, 50 mM Tris-HCl and 0.75 M sucrose) and kept frozen (- 50°C on board and - 80°C afterwards). DNA extraction was performed by digesting with lysozyme, proteinase K and SDS, purifying by phenolyzation, and concentrating with a Centricon-100 as described before (Díez *et al*., 2001). Subsamples for FISH counts (50–100 ml) were fixed with 0.2 μm-filtered formaldehyde (3.7% final concentration), kept for 1–24 h at 4°C, filtered on 0.6 μm pore-size polycarbonate filters, and stored frozen.

Viral abundance by flow cytometry

Fixed samples were stained with the nucleic acid stain SYBR Gold (Molecular Probes) as described before (Chen *et al*., 2001). Fluorescent microspheres (1 μm TransFluoSpheres 488/560, Molecular Probes) were added to the sample as counting and internal fluorescence reference. The concentration of microspheres in the stock solution was calculated from the number quantified by flow cytometry in a given volume, estimated as the weight loss of the sample during measurement. Flow cytometric analysis was performed on a MoFlo (DakoCytomation) at the laboratory in Innsbruck. This instrument is equipped with a water-cooled argon ion 4 W Innova 90°C + laser (Coherent) tuned to 488 nm with an output power of 200 mW at TEM00. The orthogonal side scatter (SSC) was measured at 488/10 nm, the green fluorescence of SYBR Gold at 535/50 nm after a 495-nm longpass dichroic beamsplitter, and the yellow signals from the microspheres at 630/40 nm after a 570-nm longpass dichroic beamsplitter. Detectors were R-1477 photomultiplier tubes (Hamamatsu) at 450, 520 and 700 V for SSC, SYBR Gold and yellow signals respectively. Measurements were triggered on logarithmically amplified SYBR Gold signals. Underestimation of particle abundance due to detection of aggregated viruses (i.e. coincidence) was minimized by measuring < 900 events s^{-1}. Suspensions of the bacteriophage T4 (Leiman *et al*., 2003) and of *E. huxleyi* virus EhV-86 (Wilson *et al*., 2005) were used as reference. Their capsid size was 85 and 175 nm respectively. The two viruses were

detected as distinct populations and separated from the electronic noise of the instrument on histograms of SSC versus SYBR Gold fluorescence.

Transmission electron microscopy (TEM)

A paraformaldehyde/glutaraldehyde fixed aliquot of 1.5 ml was concentrated on a Millipore Ultrafree-MC, 0.1-mm filter unit–spin column by centrifugation at 4000 *g* for 2 min. The filter was washed three times with 0.1 M phosphate buffer (10 min incubation) followed each time by centrifugation at 4000 *g* for 2 min. The filter was removed and fixed with osmium tetraoxide during 1 h according to the protocol of Shigenaka and colleagues (1973). Afterwards, the fixative was removed by centrifugation (3000 *g* for 30 s) and subjected to a series of dehydration steps in ethanol (50%, 2 ¥ 70%, 90%, 3 ¥ 100%). The dehydrated sample was embedded in a low viscosity Spurr's resin. Ultrathin sections (90 nm) were cut with a glass knife on an ultramicrotome (Reichert-Jung Ultracut E) and mounted on Formvar-coated grids. Ultrathin sections were stained with a saturated aqueous solution of uranyl acetate mixed with an equal volume of ethanol for 20 min and further stained with alkaline lead citrate for 10 min. Sectioned material was observed with a Zeiss TEM 902 electron microscope at the laboratory in Innsbruck. Viruses were observed inside the flagellates at a magnification of 55 000 ¥ and micrographs were taken at 20 000 ¥.

Denaturing Gradient Gel Electrophoresis (DGGE)

One microlitre of DNA extract was used as template for polymerase chain reaction (PCR) amplification of eukaryotic 18S rDNA using primers Euk1A and Euk516r-GC (Díez *et al*., 2001). DGGE was carried out with a DGGE-2000 system (CBS Scientific Company). Gels of 6% polyacrylamide were prepared with a gradient of denaturant agent from 40% to 65% (100% denaturant agent being 7 M urea and 40% deionized formamide). 800 ng of PCR product were loaded for each sample and the gel was run at 100 V for 16 h at 60°C in 1 ¥ TAE buffer (40 mM Tris [pH 7.4], 20 mM sodium acetate, 1 mM EDTA). The gel was stained with SYBR Gold (Molecular Probes) and visualized with UV radiation in a Fluor-S MultiImager (Bio-Rad). High-resolution images were analysed with the software Quantity One (Bio-Rad) to detect DGGE bands, quantify their intensity and identify the same band position across the different lanes of the gel. A matrix was constructed with the relative intensity of individual bands in each lane. This matrix was used to calculate a distance matrix with City-block distances and a dendrogram with Ward's method using the software Statistica 6.0 (StatSoft).

Genetic libraries and phylogenetic analysis

18S rRNA genes were PCR-amplified with the eukaryotic primers EukA and EukB following the conditions described before (Díez *et al*., 2001). Polymerase chain reaction products from several reactions were cleaned with the QIAGEN PCR purification kit and cloned with the TOPO-TA cloning kit (Invitrogen). The presence of the 18S rDNA insert in the positive colonies was checked by PCR amplification with the same primers. Polymerase chain reaction products of the right insert size were digested with the restriction enzyme HaeIII (Invitrogen) and run in agarose electrophoresis to identify clones with the same RFLP pattern. Complete 18S rDNA sequences were obtained with the Bigdye Terminator Cycle Sequencing kit v.3.0 (PE Biosystems) and an ABI PRISM model 377 (v. 3.3) automated sequencer using five eukaryotic primers. These were aligned with a selection of bicosoecid sequences using ClustalW 1.82 (Thompson *et al*., 1994). Highly variable regions of the alignment were removed using Gblocks (Castresana, 2000) leaving 1553 informative positions. Maximum likelihood analysis was carried out with PAUP 4.0b10 (Swofford, 2002) using the optimal evolutive model and parameters found by ModelTest (Posada and Crandall, 1998). Sequences have been deposited in GenBank under the Accession numbers EF620521–EF620528.

Fluorescent in situ *hybridization (FISH)*

Two oligonucleotide probes against *C. roenbergensis* (including *Cafeteria mylnikovii*) and *Caecitellus parvulus* and *C. paraparvulus* (Hausmann *et al*., 2006) were designed with the ARB package (http://www.arb-home.de/): CAF01 (5′-ACAGTGCTGACACCCTGT-3′) and CET01 (5′-CAGC TCAATACGGACACC-3′) respectively. Closest non-target sequences have at least eight mismatches with the probes, except *Caecitellus pseudoparvulus* sequences which have three central mismatches with CET01. Probes were supplied labelled with a CY3 fluorophore at the 5′ end and were tested against the targeted cultures in a gradient of formamide (0–50%) in the hybridization buffer and constant temperature (46°C) as explained before (Massana *et al*., 2006a). Optimal hybridization signal was obtained with 30% formamide in both cases. Probes gave negative results with all non-target cultures tested: a prasinophyte, a prymnesiophyte, a cryptophyte, an eustigmatophyte and a chrysophyte (listed in Massana *et al*., 2006a). These two new probes, together with probe Euk502 universal for eukaryotes (Lim *et al*., 1999) were applied to samples from the incubations as follows. Filter portions (2 per sample) with protist cells were hybridized for 3 h at 46°C with each probe at 5 ng ml^{-1} in a buffer of 900 nM NaCl, 20 mM Tris-HCl, 0.01% SDS and 30% formamide. Filters were washed for 15 min at 48°C in a buffer of 110 mM NaCl, 20 mM Tris-HCl, 5 mM EDTA and 0.01% SDS. Filters were then dried, counterstained with DAPI, mounted in a slide with a mix of Citifluor and Vecta Shield (4:1), and observed by epifluorescence microscopy under UV excitation (DAPI signal) and green light excitation (CY3 signal) at a magnification of 1250 ¥. Four transects were inspected per sample and mean cell counts and standard errors were calculated. Standard errors were typically 20% of the mean. The detection limit of the counts performed here was around 1 cell ml^{-1}.

Acknowledgements

This study was supported by projects ESTRAMAR (CTM2004-12631/MAR, MEC) and TRANSINDICO

(REN2002-10951-E/MAR, MCyT) to R.M. and by a project from the Austrian Science Fund (19245-BO3) to R.S. The chief-scientist D. Blackman and the crew of R/V *Melville* (Scripps Institution of Oceanography) provided an excellent sampling environment. We thank E. Goetze for her invitation on board, F. Not, I. Forn, F. Unrein (Barcelona) for sampling and laboratory assistance, J. Hofer, W. Salvenmoser and M. Summerer (Innsbruck) for helping with flow cytometric counts and TEM preparations, and W. Wilson for kindly providing the virus EhV-86 to R.S.

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430:** 551–554.

Arndt, H., Dietrich, D., Auer, B., Cleven, E.J., Gräfenhan, T., Weitere, M., and Mylnikov, A.P. (2000) Functional diversity of heterotrophic flagellates in aquatic ecosystems. In *The Flagellates: Unity, Diversity and Evolution*. Leadbeater, B.S.C., and Green, J.C. (eds). London, UK: Taylor & Francis Press, pp. 240–268.

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

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

Bratbak, G., Egge, J.K., and Heldal, M. (1993) Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Series* **93:** 39–48.

Brussaard, C.P.D. (2004) Viral control of phytoplankton populations – a review. *J Eukaryot Microbiol* **51:** 125–138.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540–552.

Chen, F., Lu, J.-R., Binder, B.J., Liu, Y.-C., and Hodson, R.E. (2001) Application of digital image analysis and flow cytometry to enumerate marine viruses stained with SYBR Gold. *Appl Environ Microbiol* **67:** 539–545.

Cottrell, M.T., and Suttle, C.A. (1995) Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate *Micromonas pusilla*. *Limnol Oceanogr* **40:** 730–739.

Díez, B., Pedrós-Alió, C., and Massana, R. (2001) Application of Denaturing Gradient Gel Electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbiol* **67:** 2942–2951.

Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* **399:** 541–548.

Garza, D.R., and Suttle, C.A. (1995) Large double-stranded DNA viruses which cause the lysis of a marine heterotrophic nanoflagellate (*Bodo* sp.) occur in natural marine viral communities. *Aquat Microb Ecol* **9:** 203–210.

González, J.M., and Suttle, C.A. (1993) Grazing by marine nanoflagelaltes on viruses and virus-sized particles: ingestion and digestion. *Mar Ecol Prog Series* **94:** 1–10.

Hausmann, K., Selchow, P., Scheckenbach, F., Weitere, M., and Arndt, H. (2006) Cryptic species in a morphospecies complex of heterotrophic flagellates: the case study of *Caecitellus* spp. *Acta Protozool* **45:** 415–431.

Hofer, J., and Sommaruga, R. (2001) Seasonal dynamics of viruses in an alpine lake: importance of filamentous forms. *Aquat Microb Ecol* **26:** 1–11.

Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60:** 2356–2370.

Lim, E.L., Dennet, M.R., and Caron, D.A. (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.

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003) Bacterial photosynthesis genes in a virus. *Nature* **424:** 741.

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.

Middelboe, M., Glud, R.N., and Finster, K. (2003) Distribution of viruses and bacteria in relation to diagenetic activity in an estuarine sediment. *Limnol Oceanogr* **48:** 1447–1456.

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.

Nagasaki, K., Ando, M., Imai, I., Itakura, S., and Ishida, Y. (1993) Virus-like particles in an apochlorotic flagellate in Hiroshima Bay, Japan. *Mar Ecol Prog Series* **61:** 235–239.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299–304.

Patterson, D.J., and Larson, J. (1991) *The Biology of Free-Living Heterotrophic Flagellates*. Oxford, UK: Clarendon Press. 505 pages.

Patterson, D.J., and Lee, W.J. (2000) Geographic distribution and diversity of free-living heterotrophic flagellates. In *The Flagellates: Unity, Diversity and Evolution*. Leadbeater, B.S.C., and Green, J.C. (eds). London, UK: Taylor & Francis Press, pp. 269–287.

Pedrós-Alió, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiology* **14:** 257–263.

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

Posada, D., and Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Preisig, H.R., and Hibberd, D.J. (1984) Virus-like particles and endophytic bacteria in *Paraphysomonas* and *Chromophysomonas* (Chrysophyceae). *Nord J Bot* **4:** 279–285.

Sherr, E.B., and Sherr, B.F. (2002) Significance of predation by protists in aquatic microbial food webs. *Antonie Van Leeuwenhoek* **81:** 293–308.

Shigenaka, Y., Watanabe, K., and Kaneda, M. (1973) Effects of glutaraldehyde and osmium tetroxide on hypotrichous ciliates, and determination of the most satisfactory fixation methods for electron microscopy. *J Protozool* **20:** 414–420.

Short, S.M., and Suttle, C.A. (2002) Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68:** 1290–1296.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424:** 1047–1051.

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

Suttle, C.A., and Chan, A.M. (1994) Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol* **60:** 3167–3174.

Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sunderland, MA, USA: Sinauer Associates.

Takao, Y., Nagasaki, K., Mise, K., Okuno, T., and Honda, D. (2005) Isolation and characterization of a novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium* sp. (Thraustochytriaceae, Labyrinthulea). *Appl Environ Microbiol* **71:** 4516–4522.

Tarutani, K., Nagasaki, K., and Yamaguchi, M. (2000) Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. *Appl Environ Microbiol* **66:** 4916–4920.

Thingstad, T.F., and Lignell, R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol* **13:** 19–27.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22:** 4673–4680.

Van Etten, J.L., and Meints, R.H. (1999) Giant viruses infecting algae. *Annu Rev Microbiol* **53:** 447–494.

Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28:** 127–181.

Weinbauer, M.G., and Rassoulzadegan, F. (2004) Are viruses driving microbial diversification and diversity? *Environ Microbiol* **6:** 1–11.

Wilson, W.H., Schroeder, D.C., Allen, M.J., Holden, M.T.G., Parkhill, J., Barrell, B.G., *et al.* (2005) Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* **309:** 1090–1092.

Wommack, K.E., and Colwell, R.R. (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64:** 69–114.

Zingone, A., Natale, F., Biffali, E., Borra, M., Forlani, G., and Sarno, D. (2006) Diversity in morphology, infectivity, molecular characteristics and induced host resistance between two viruses infecting *Micromonas pusilla*. *Aquat Microb Ecol* **45:** 1–14.

2

Diversitat emergent dintre de les crisofícies, els coanoflagel·lats i els bicosoècids a partir de dades obtingudes d'estudis moleculars

En els darrers anys, s'han obtingut una quantitat considerable de dades sobre els protists aquàtics a partir d'aproximacions moleculars independents de cultiu, revelant una gran diversitat i l'existència de nous llinatges. No obstant això, les seqüències corresponents a grups de menor importància (en termes d'abundància clonal) sovint no han estat objecte d'una anàlisi en profunditat, aquí s'amaga una font potencialment important d'informació filogenètica. En aquest estudi hem buscat a les bases de dades públiques seqüències d'ADNr 18S de crisofícies, coanoflagel·lats i bicosoècids obtingudes en prospeccions de mostres ambientals amb tècniques moleculars. Aquests tres grups han estat sovint considerats com a constituents de la major part dels flagel·lats heterotròfics marins, i com a tals un important component funcional en les xarxes tròfiques microbianes. Vam trobar que representaven una fracció significativa dels clons en els estudis d'aigua dolça, mentre que la seva abundància clonal relativa va ser baixa en els estudis marins. La novetat mostrada per aquest treball va ser notable. La majoria de les seqüències ambientals van resultar ser distants a seqüències d'organismes cultivats, indicant un biaix significatiu en la representació dels tàxons en cultiu. A més, sovint eren també distants a seqüències d'altres estudis moleculars, la qual cosa suggereix un esforç de seqüenciació insuficient a l'hora de caracteritzar aquests grups a l'ecosistema. Els arbres filogenètics realitzats amb seqüències completes ens han permès generar la representació més acurada d'aquests grups fins al moment, amb l'aparició de diversos subtipus nous constituïts exclusivament per seqüències ambientals. Una feina exhaustiva de mineria de dades va permetre la identificació de la nova diversitat oculta dins les crisofícies, els coanoflagel·lats i els bicosoècids.

Protist, Vol. 162, 435–448, July 2011
http://www.elsevier.de/protis
Published online date 15 January 2011

Protist

ORIGINAL PAPER

Emerging Diversity within Chrysophytes, Choanoflagellates and Bicosoecids Based on Molecular Surveys

Javier del Campo[1], **and Ramon Massana**

Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar, CSIC.
Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Catalonia, Spain

Submitted June 30, 2010; Accepted October 30, 2010
Monitoring Editor: Michael Melkonian

In recent years, a substantial amount of data on aquatic protists has been obtained from culture-independent molecular approaches, unveiling a large diversity and the existence of new lineages. However, sequences affiliated with minor groups (in terms of clonal abundance) have often been under-analyzed, and this hides a potentially relevant source of phylogenetic information. Here we have searched public databases for 18S rDNA sequences of chrysophytes, choanoflagellates and bicosoecids retrieved from molecular surveys of protists. These three groups are often considered to account for most of the heterotrophic flagellates, an important functional component in microbial food webs. They represented a significant fraction of clones in freshwater studies, whereas their relative clonal abundance was low in marine studies. The novelty displayed by this dataset was notable. Most environmental sequences were distant to sequences of cultured organisms, indicating a significant bias in the representation of taxa in culture. Moreover, they were often distant to sequences from other molecular surveys, suggesting an insufficient sequencing effort to characterize the in situ diversity of these groups. Phylogenetic trees with complete sequences present the most accurate representation of the diversity of these groups, with the emergence of several new clades formed exclusively by environmental sequences. Exhaustive data mining in sequence databases allowed the identification of new diversity hidden inside chrysophytes, choanoflagellates and bicosoecids.

Key words: 18S rDNA; bicosoecids; choanoflagellates; chrysophytes; emerging diversity; heterotrophic flagellates; maximum likelihood phylogeny; molecular surveys.

Introduction

Heterotrophic Flagellates (HF) are distributed in planktonic environments at concentrations between 10^2 and 10^5 cells ml^{-1}, representing 10-30% of protist cells in upper marine waters (Jürgens and Massana 2008). HF cells are often phagotrophs that graze and control the abundance of prokaryotes and picoeukaryotes (Pernthaler 2005), but also may include dispersal stages of parasites of other marine organisms (Guillou et al. 2008). Consequently, HF are important actors in microbial food webs and play key roles in global biogeochemical cycles (Chambouvet et al. 2008; Sherr and Sherr 2002;). Traditionally, the diversity of HF assemblages has been studied by microscopy and culturing, yielding the impression that most

[1]Corresponding author; fax 93-2309555
e-mail jdelcampo@icm.cat (J. del Campo).

doi:10.1016/j.protis.2010.10.003

cells belong to chrysophytes, choanoflagellates or bicosoecids (Arndt et al. 2000; Fenchel 1982). However, the in situ diversity and ecological relevance of these taxonomic groups remain poorly investigated.

The chrysophytes is a large group of stramenopiles with about 100 described genera (Lee et al. 2000). They include colorless cells (heterotrophs) and chloroplast-containing cells (phototrophs or mixotrophs) with one or two flagella (Preisig et al. 1991). The majority lives in freshwater but there are also some well-known marine species, such as *Paraphysomonas imperforata*. The phylogeny of chrysophytes using 18S rDNA was presented by Andersen et al. (1999), and currently there are 30 genera represented in GenBank. The choanoflagellates are colorless ovoid cells with about 50 genera described from marine, brackish and freshwater systems (Leadbeater 1991; Lee et al. 2000). They have a collar surrounding a unique flagellum, and some are covered by an intricate lorica. They belong to Opisthokonta and are the closest metazoan relatives, thus attracting the interest of evolutionary biologists (King et al. 2008). Their phylogeny using the 18S rDNA was presented in Carr et al. (2008) and currently there are 16 genera in GenBank's Taxonomy. Bicosoecids are colorless flagellates that belong to the stramenopiles and include 11 genera (Cavalier-Smith and Chao 2006; Lee et al. 2000;), all represented in GenBank's Taxonomy with their 18S rDNA. Cells have typically two flagella. Both marine and freshwater species are known, including the well-known marine species *Cafeteria roenbergensis* (Fenchel and Patterson 1988).

Cultured strains have been essential for delineating the physiology and phylogeny of the three groups (Andersen et al. 1999; Cavalier-Smith and Chao 2006; Leipe et al. 1994), but it is not clear if these cultured strains are ecologically relevant. For instance, a very low abundance of *Paraphysomonas imperforata* (Lim et al. 1999) and *Cafeteria roenbergensis* (Massana et al. 2007) was recorded in samples from which these two species were easily enriched. In situ diversity can be better addressed by culture-independent molecular techniques (Caron et al. 2004). Environmental 18S rDNA libraries targeting microbial eukaryotes highlighted new lineages that appeared in most studies in high clonal abundance, such as MAST (Marine Stramenopiles) (Massana et al. 2006) and MALV (Marine Alveolates) (Guillou et al. 2008), whereas chrysophytes, choanoflagellates or bicosoecids were generally represented by few sequences in marine (Massana and Pedrós-Alió 2008) and freshwater (Lefranc et al. 2005; Richards et al. 2005; Šlapeta et al. 2005) individual studies. These later groups have been under analyzed due to their low clonal abundance, and we hypothesize that new diversity would emerge once we put together sequences from independent studies.

Here, we searched public databases (nucleotide collection nr/nt in GenBank) for chrysophyte, choanoflagellate and bicosoecid 18S rDNA sequences obtained in molecular surveys. We used this sequence dataset to pursue three goals: First, to determine the clonal contribution of these groups in marine and freshwater systems. Second, to analyze the sequence novelty within each group, i.e. the difference between target sequences and those deposited in GenBank (both from cultured strains and from other molecular surveys). This novelty can then be interpreted in terms of sequencing effort and representation of taxa in culture. Third, to present a robust phylogeny of each group combining all available sequences to better describe their diversity and identify new clades formed by environmental sequences only. These phylogenetic trees can serve as a backbone where to map tag sequences that begin to appear by Next Generation Sequencing technologies (Amaral-Zettler et al. 2009; Stoeck et al. 2009). For each of the three taxonomic groups, major differences are found in clonal abundance, novelty pattern and new diversity in marine and freshwater systems.

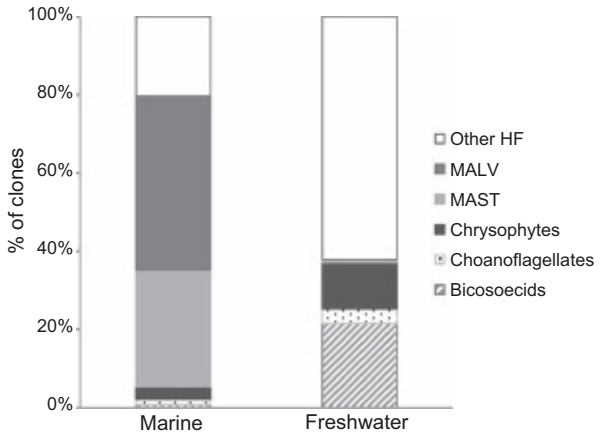


Figure 1. Relative clonal abundance of different taxonomic groups putatively forming the heterotrophic flagellate assemblages in marine and freshwater systems (data from 82 clone libraries of 18S rDNA genes; see Supplementary Table S3).

Table 1. Novelty degree represented by environmental sequences of chrysophytes, choanoflagellates and bicosoecids. In this integrated analysis we show the average similarity (standard error in brackets) with closest environmental match (CEM) and closest cultured match (CCM) for all sequences separated by environments and together. The second column shows to the number of sequences analyzed and the last column the statistical tests (***: p< 0.0001, ns: not significant).

	Environment	n	% CEM (SE)	% CCM (SE)	t-student
Chrysophytes	Marine	144	97.6 (0.2)	94.2 (0.3)	***
	Freshwater	86	95.3 (0.3)	95.8 (0.3)	ns
	All	230	96.8 (0.2)	94.8 (0.2)	***
Choanoflagellates	Marine	69	95.3 (0.3)	94.7 (0.4)	ns
	Freshwater	20	90.8 (0.5)	91.6 (0.7)	ns
	All	89	94.3 (0.3)	94.0 (0.3)	ns
Bicosoecids	Marine	45	98.1 (0.4)	98.3 (0.5)	ns
	Freshwater	31	90.9 (0.4)	90.6 (0.6)	ns
	All	76	95.1 (0.3)	95.0 (0.4)	ns

Results

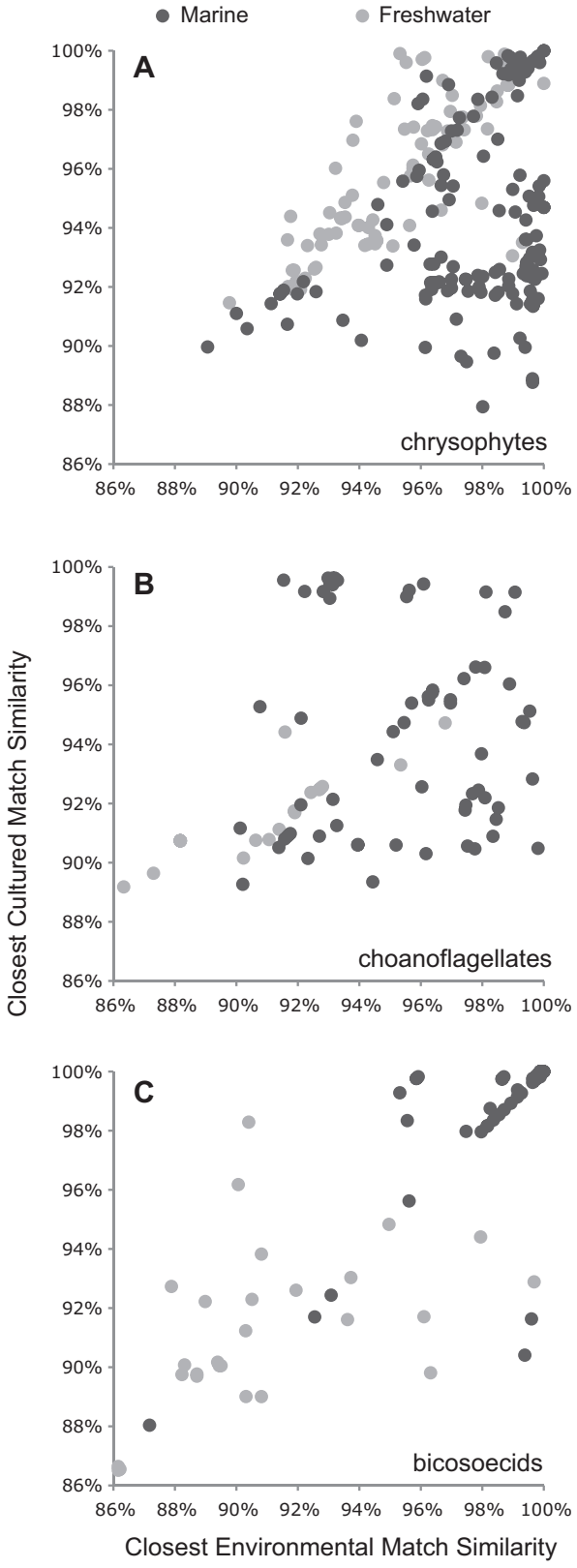
To obtain an exhaustive description of the phylogenetic diversity of chrysophytes, choanoflagellates and bicosoecids, we screened GenBank and our unpublished libraries to retrieve all sequences from these groups obtained in marine and freshwater molecular surveys. The dataset inspected included 292 environmental clone libraries of 18S rDNA genes (representing more than 13000 sequences) that have been published in 58 scientific papers and targeted a large variety of systems, depths in the water column, and physical-chemical settings (Supplementary Table S1). Some studies focused on the smallest eukaryotic microbes (<3-5 μm) and others to the whole water community. Overall, we obtained 230 chrysophyte, 89 choanoflagellate and 76 bicosoecid environmental sequences (listed in the Supplementary Table S2). Sequences were grouped into two categories (marine and freshwater) before further abundance, novelty and diversity analyses.

Relative Clonal Abundance in Environmental Surveys

The representation of chrysophyte, choanoflagellate and bicosoecid sequences in 18S rDNA libraries was addressed considering only the studies that reported the clonal abundance of distinct taxonomic groups (82 libraries published in 14 papers, Supplementary Table S3). In each library, clones were assigned to putative heterotrophic flagellate (HF) groups, to putative phototrophic (PP) protist groups (prasinophytes, dinoflagellates, haptophytes and others) and to other heterotrophic protists (OHP) (ciliates and fungi). Then, the proportion of clones within different HF groups was displayed (Fig. 1). Chrysophyte sequences appeared in most environmental surveys, averaging 3.3% of HF clones in marine and 11.8% in freshwater studies (Fig. 1). The relative clonal abundance of choanoflagellates averaged 1.3% in marine and 3.7% in freshwater systems. Bicosoecids were rarely found in marine surveys (0.6% relative clonal abundance on average) and were rather abundant in freshwater systems (21.6% on average, in some cases up to 50%). The bulk of sequences from putative HF in marine systems affiliated with MALV and MAST. In freshwater systems, other alveolates and cercozoans accounted for a significant number of clones.

Novelty of Environmental Sequences

Figure 2 plots together two values obtained for each environmental sequence after a GenBank search: the similarity against the closest environmental match (CEM) and the similarity against the closest cultured match (CCM). Sequences appeared widely distributed in the graph with each taxonomic group displaying a distinct novelty pattern. Most chrysophyte sequences from marine samples accumulated in two plot regions: those with high CEM-CCM similarity values (above 98%), thus similar to sequences from cultures and molecular surveys, and those with high CEM (above 98%) and low CCM values (below 94%), thus similar only to sequences from molecular surveys (Fig. 2A). Choanoflagellates sequences showed a more uniform dispersion in the graph, with a tendency of freshwater sequences to have lower values in both axis (Fig. 2B). Interestingly, we detected some sequences that were very close to cultured species but had not been retrieved in other molec-



ular surveys (this did not occur in chrysophytes). The novelty pattern for bicosoecids also showed a uniform dispersion of dots in the graph, as the previous example, but here the difference between systems was very marked, with sequences from marine environments being above 98% in both axis (Fig. 2C).

Averaging the similarity values against CEM and CCM for all sequences yielded the novelty degree of a given dataset (Table 1). The difference between CCM similarity and 100% represented the bias in representation of cultures, whereas the difference between CEM similarity and 100% represented the bias in environmental sequencing. Considering all sequences together yielded average similarities of 94-95% in all cases (except chrysophytes against CEM). This general overview obscured clear differences between systems, with choanoflagellates and bicosoecids being significantly more novel in freshwater (91% similarity) than in marine systems (95% and 98%, respectively). The difference between CEM and CCM similarity in each row represented the increase of knowledge gained by environmental sequencing. Surprisingly, in most cases both values were very similar. The only exception was the marine chrysophytes, that showed significant differences between both values (t-student test, $p<0.0001$). Altogether, the novelty degree was larger in freshwater than in marine systems.

Phylogenetic Trees and New Clades

Using complete 18S rDNA sequences, we constructed Maximum Likelihood phylogenetic trees for chrysophytes (Fig. 3), choanoflagellates (Fig. 4) and bicosoecids (Fig. 5). Environmental sequences appeared in the trees in different color depending on their origin (blue: marine; green: freshwater), whereas reference sequences from cultured organisms appeared in black. Trees were divided into separate clades, some of them already defined in published trees and others being new, derived from the present analysis. Clades always contained

Figure 2. Novelty pattern derived from chrysophyte (**A**), choanoflagellate (**B**) and bicosoecid (**C**) environmental sequences. Dots represent the % similarity with the closest environmental match (CEM) and the closest cultured match (CCM) for each sequence within the three taxa (229, 88, and 76 sequences, respectively) and are colored depending the environment where they originate (dark: marine; light: freshwater).

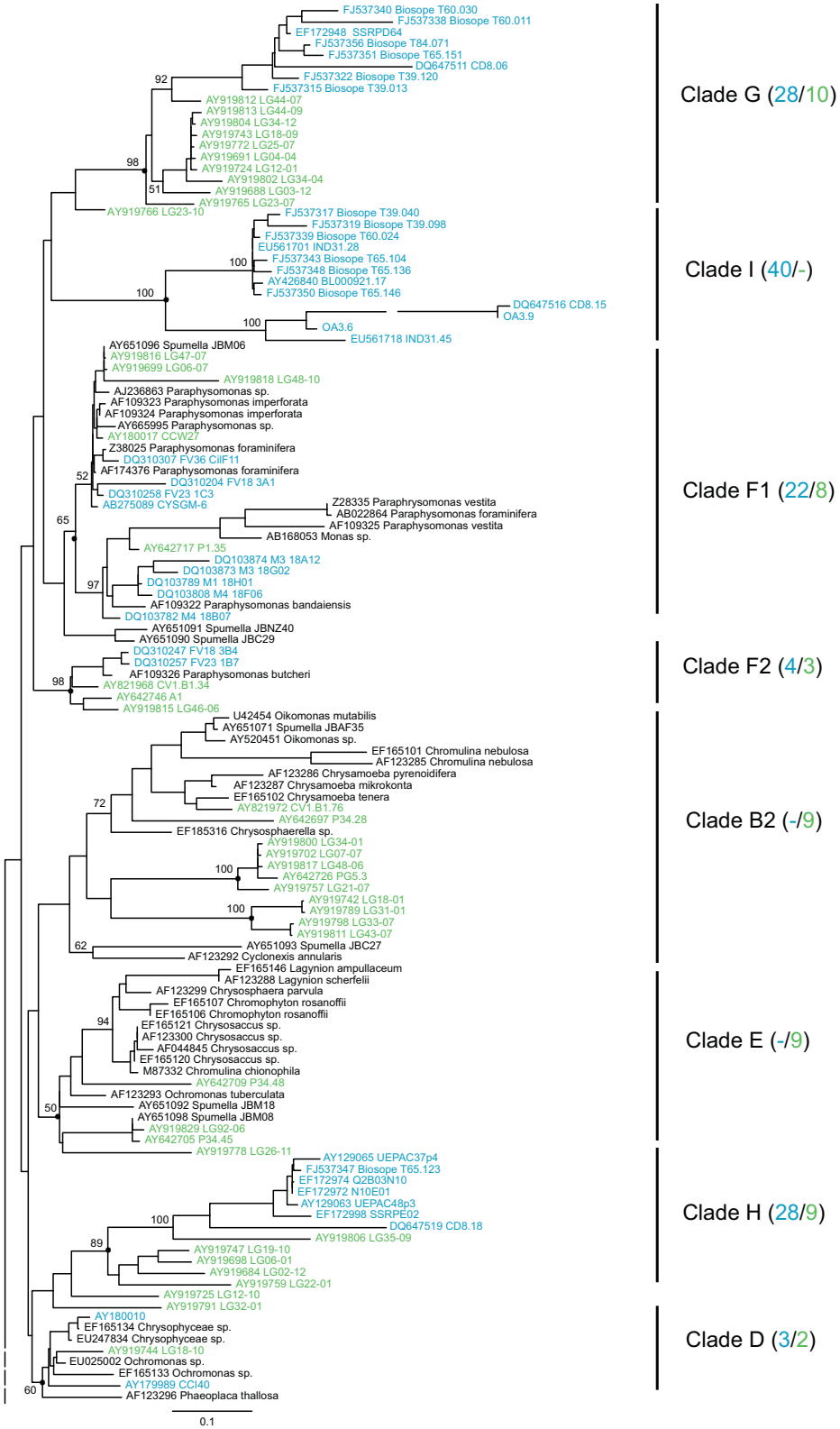

FJ537340 Biospe T60.030
FJ537338 Biospe T60.011
EF172948 SSRPD64
FJ537356 Biospe T84.071
FJ537351 Biospe T65.151
DQ647511 CD8.06
FJ537322 Biospe T39.120
FJ537315 Biospe T39.013
AY919812 LG44-07
AY919813 LG44-09
AY919804 LG34-12
AY919743 LG18-09
AY919772 LG25-07
AY919691 LG04-04
AY919724 LG12-01
AY919802 LG34-04
AY919688 LG03-12
AY919765 LG23-07
AY919766 LG23-10
92
98
51
Clade G (28/10)
FJ537317 Biospe T39.040
FJ537319 Biospe T39.098
FJ537339 Biospe T60.024
EU561701 IND31.28
FJ537343 Biospe T65.104
FJ537348 Biospe T65.136
AY426840 BL000921.17
FJ537350 Biospe T65.146
100
100
Clade I (40/-)
DQ647516 CD8.15
OA3.9
OA3.6
EU561718 IND31.45
100
AY651096 Spumella JBM06
AY919816 LG47-07
AY919699 LG06-07
AY919818 LG48-10
AJ236863 Paraphysomonas sp.
AF109323 Paraphysomonas imperforata
AF109324 Paraphysomonas imperforata
AY665995 Paraphysomonas sp.
AY180017 CCW27
Z38025 Paraphysomonas foraminifera
DQ310307 FV36 CilF11
AF174376 Paraphysomonas foraminifera
DQ310204 FV18 3A1
DQ310258 FV23 1C3
AB275089 CYSGM-6
52
Z28335 Paraphrysomonas vestita
AB022864 Paraphysomonas foraminifera
AF109325 Paraphysomonas vestita
AB168053 Monas sp.
AY642717 P1.35
DQ103874 M3 18A12
DQ103873 M3 18G02
DQ103789 M1 18H01
DQ103808 M4 18F06
AF109322 Paraphysomonas bandaiensis
DQ103782 M4 18B07
AY651091 Spumella JBNZ40
AY651090 Spumella JBC29
65
97
Clade F1 (22/8)
DQ310247 FV18 3B4
DQ310257 FV23 1B7
AF109326 Paraphysomonas butcheri
AY821968 CV1.B1.34
AY642746 A1
AY919815 LG46-06
98
Clade F2 (4/3)
U42454 Oikomonas mutabilis
AY651071 Spumella JBAF35
AY520451 Oikomonas sp.
EF165101 Chromulina nebulosa
AF123285 Chromulina nebulosa
AF123286 Chrysamoeba pyrenoidifera
AF123287 Chrysamoeba mikrokonta
EF165102 Chrysamoeba tenera
AY821972 CV1.B1.76
AY642697 P34.28
EF185316 Chrysosphaerella sp.
72
AY919800 LG34-01
AY919702 LG07-07
AY919817 LG48-06
AY642726 PG5.3
AY919757 LG21-07
AY919742 LG18-01
AY919789 LG31-01
AY919798 LG33-07
AY919811 LG43-07
100
100
AY651093 Spumella JBC27
AF123292 Cyclonexis annularis
62
Clade B2 (-/9)
EF165146 Lagynion ampullaceum
AF123288 Lagynion scherfelii
AF123299 Chrysosphaera parvula
EF165107 Chromophyton rosanoffii
EF165106 Chromophyton rosanoffii
EF165121 Chrysosaccus sp.
AF123300 Chrysosaccus sp.
AF044845 Chrysosaccus sp.
EF165120 Chrysosaccus sp.
M87332 Chromulina chionophila
AY642709 P34.48
AF123293 Ochromonas tuberculata
AY651092 Spumella JBM18
AY651098 Spumella JBM08
AY919829 LG92-06
AY642705 P34.45
AY919778 LG26-11
94
50
Clade E (-/9)
AY129065 UEPAC37p4
FJ537347 Biospe T65.123
EF172974 Q2B03N10
EF172972 N10E01
AY129063 UEPAC48p3
EF172998 SSRPE02
DQ647519 CD8.18
AY919806 LG35-09
100
AY919747 LG19-10
AY919698 LG06-01
AY919684 LG02-12
AY919759 LG22-01
AY919725 LG12-10
AY919791 LG32-01
89
Clade H (28/9)
AY180010
EF165134 Chrysophyceae sp.
EU247834 Chrysophyceae sp.
AY919744 LG18-10
EU025002 Ochromonas sp.
EF165133 Ochromonas sp.
AY179989 CCl40
AF123296 Phaeoplaca thallosa
60
Clade D (3/2)
0.1

Figure 3.

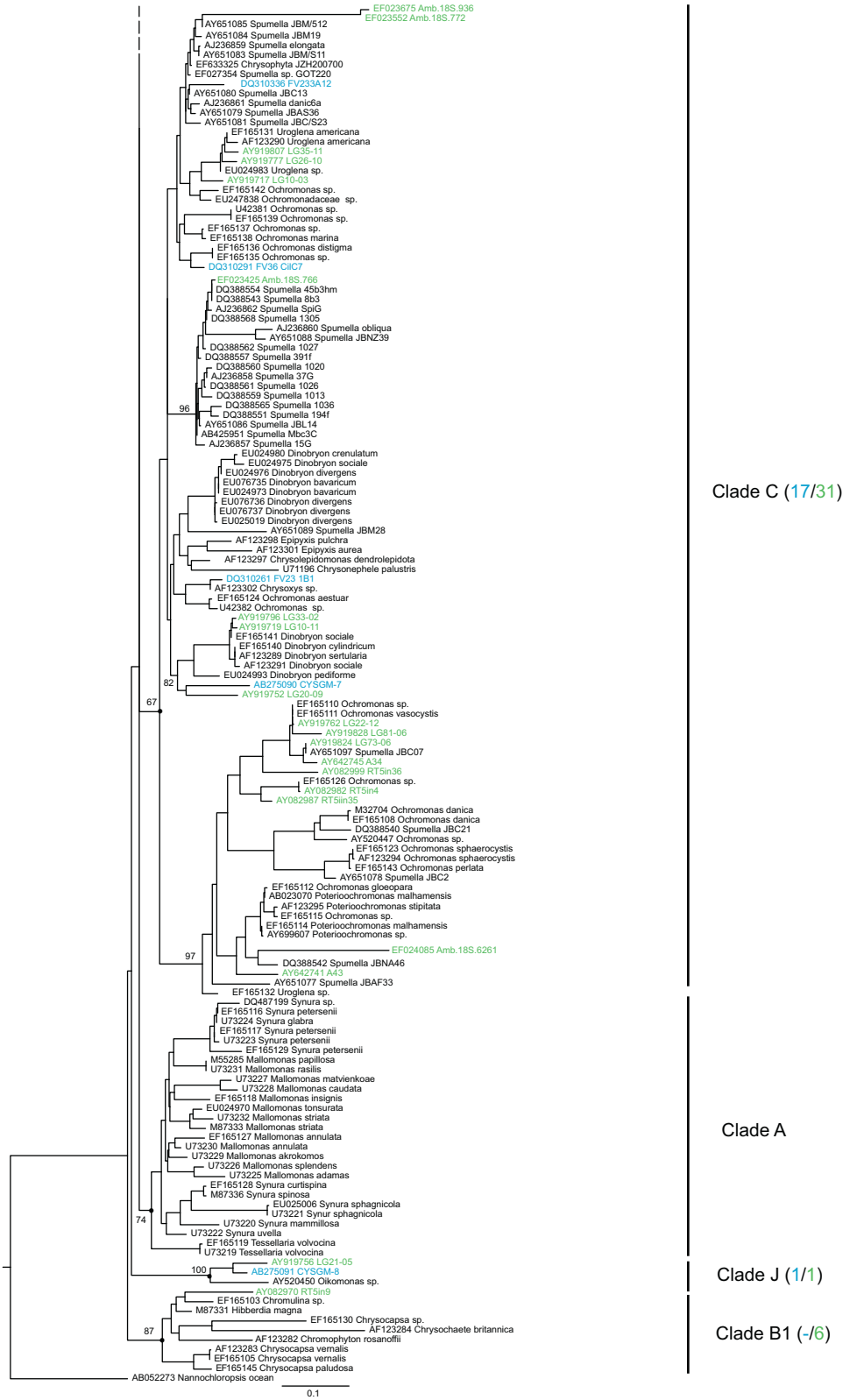


Figure 3. (*Continued*).

sequences from different studies and were generally well supported by high Maximum Likelihood bootstrap values. In addition, Neighbor Joining phylogenetic trees were done to assign partial sequences to the clades delineated by complete sequences (trees not shown). The total number of environmental sequences (complete and partial) within each clade was shown in brackets after the clade name (in blue for marine and green for freshwater sequences). Most clades contained environmental sequences.

The chrysophyte tree obtained here showed good agreement with the topology described in Andersen et al. (1999), displaying the same clades A to F defined there (although clade F was subdivided into two lineages in our tree) plus 4 additional new clades (Fig. 3). In general these clades presented ML bootstrap values above 60%. Except clade A (Synurophyceae), the other eleven clades incorporated environmental sequences. Clades B1, B2 and E contained only freshwater representatives, whereas Clades C, D, F1 and F2 contained sequences from both freshwater and marine systems. New chrysophyte clades described for Lake George (Richards et al. 2005) belonged to clade C (LG-G and LG-H) and clade F1 (LG-I). Many of the environmental sequences affiliated with the four new chrysophyte clades. Clade G contained the Marine A group from Shi et al. (2009), clones from different marine systems and also freshwater sequences from Lake George. Clade I contained only marine sequences, including the ones belonging to Shi's Marine B group. Clade H contained a monophyletic subclade of sequences from marine samples, corresponding to Shi's Marine C group, together with sequences from freshwater origin. Finally, clade J was formed by only few sequences. Since clades G, H and I included sequences from both pigmented cells (Shi et al. 2009) and putative heterotrophic cells growing in unamended dark incubations (Massana et al. 2006), they preferentially included heterotrophic or mixotrophic cells.

The emerging diversity observed in the choanoflagellate tree was also notable, with two new clades (E and F) unveiled by environmental sequences (Fig. 4). All nine defined clades were well supported by high ML bootstrap values (above 85%) and included environmental sequences. Clade C (corresponding to clade 2 of Carr et al. 2008), contained sequences from freshwater origin only, whereas the rest of the clades included only marine representatives. Carr's clade 1 was separated into clades A and B, which are distantly related phylogenetically, and the remaining clades would form Carr's clade 3.

The bicosoecid tree showed a clear separation between a large freshwater clade and several marine clades, all supported by high ML bootstrap values (Fig. 5). Most sequences retrieved from marine systems affiliated with the genera *Caecitellus* and *Cafeteria*. The *Bicosoeca* cluster included sequences previously named as MAST-13 (Zuendorf et al. 2006) that clearly belonged to bicosoecids in our stramenopile tree (not shown) and in recent studies (Park and Simpson 2010). On the other hand, most freshwater sequences appeared in two clades that were already described from Lake George, one of them (LG Heterokonta I) contained exclusively environmental sequences. Several cultured strains formed long branches without a clear position and no environmental representation.

The phylogenetic and novelty analyses could be combined to display the novelty of each clade as its position in the CEM/CCM plot based on the averaged values for all environmental sequences, and the relevance of the clade by sizing the dot proportionally to the number of sequences (Fig. 6). It is interesting to note the distinct placement of each clade within the plotted area. For instance the four new chrysophyte clades (G to J) and the two new choanoflagellate clades (D and E) all appeared below the diagonal revealing higher similarity with CEM than with CCM, confirming the environmental origin of its sequences. Another interesting case was the bicosoecid clades, all distributing along the diagonal, with extreme novelty displayed by the LG Heterokonta I clade.

Discussion

This study is an effort to analyze the data existing in environmental molecular surveys for three protist groups, chrysophytes, choanoflagellates and bicosoecids, which are often observed in aquatic

Figure 3. Maximum Likelihood phylogenetic tree of chrysophytes constructed with 270 complete 18S rDNA sequences (1648 informative positions). Sequences from cultured taxa appear in black and environmental sequences appear in blue (marine) or green (freshwater). ML bootstrap values are shown for the named clades. The number of complete and partial environmental sequences assigned to each clade appear after the clade name. The scale bar indicates 0.1 substitutions per position.

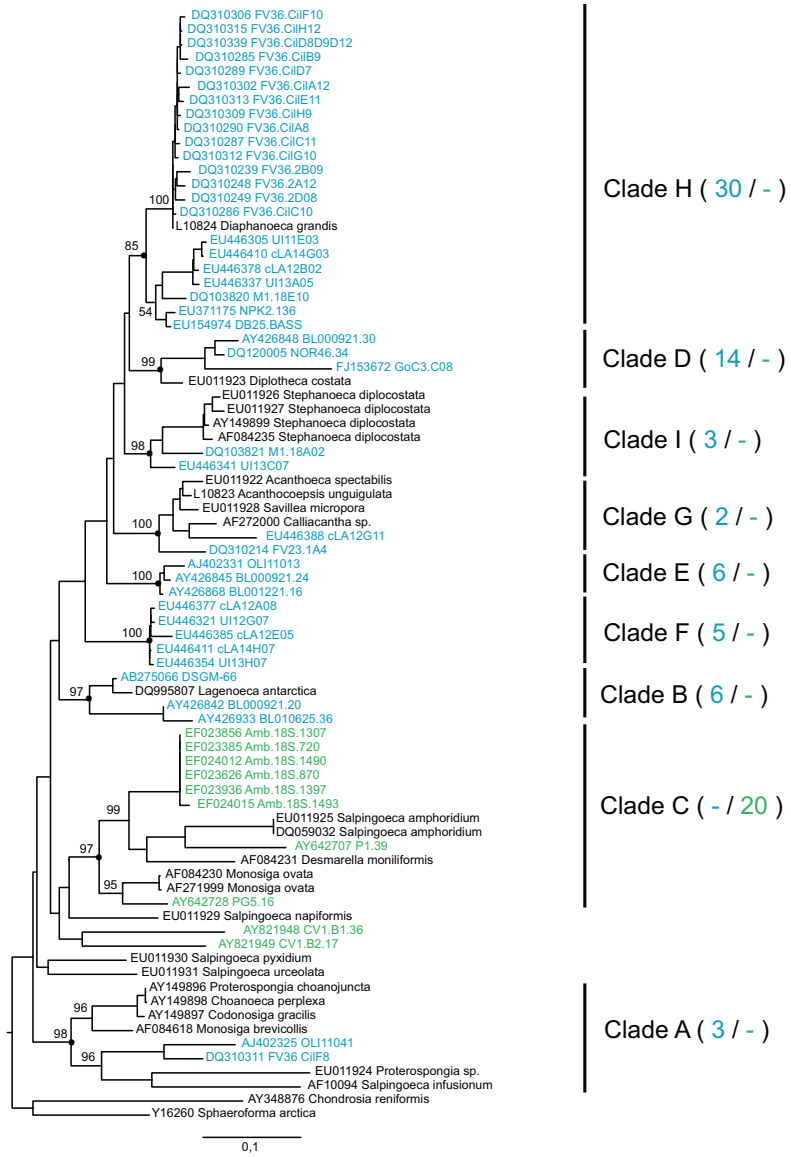


Figure 4. Maximum Likelihood phylogenetic tree of choanoflagellates constructed with 79 complete 18S rDNA sequences (1428 informative positions). Legend as in Figure 3.

samples and thought to account for a significant fraction of heterotrophic flagellates (Arndt et al. 2000; Patterson and Lee 2000). There is little doubt that sequencing of environmental clones offers an enhanced view of in situ diversity for very small protists (Caron et al. 2004; Jürgens and Massana 2008). Environmental sequences highlight the dominant members of natural assemblages and may reveal new and unexpected lineages. We do not assume that the data analyzed here do not face methodological limitations. PCR-based clone libraries suffer a variety of drawbacks that have been discussed in detail (von Wintzingerode et al. 1997). Also, different microbial size fractions were analyzed in each study (see Supplementary Table S1), potentially biasing against protists from certain size classes. In addition, intrinsic differences may occur between marine and freshwater environments, with freshwater systems being generally less homogeneous and undersampled as compared with marine systems. Nevertheless, our analysis clearly identified new

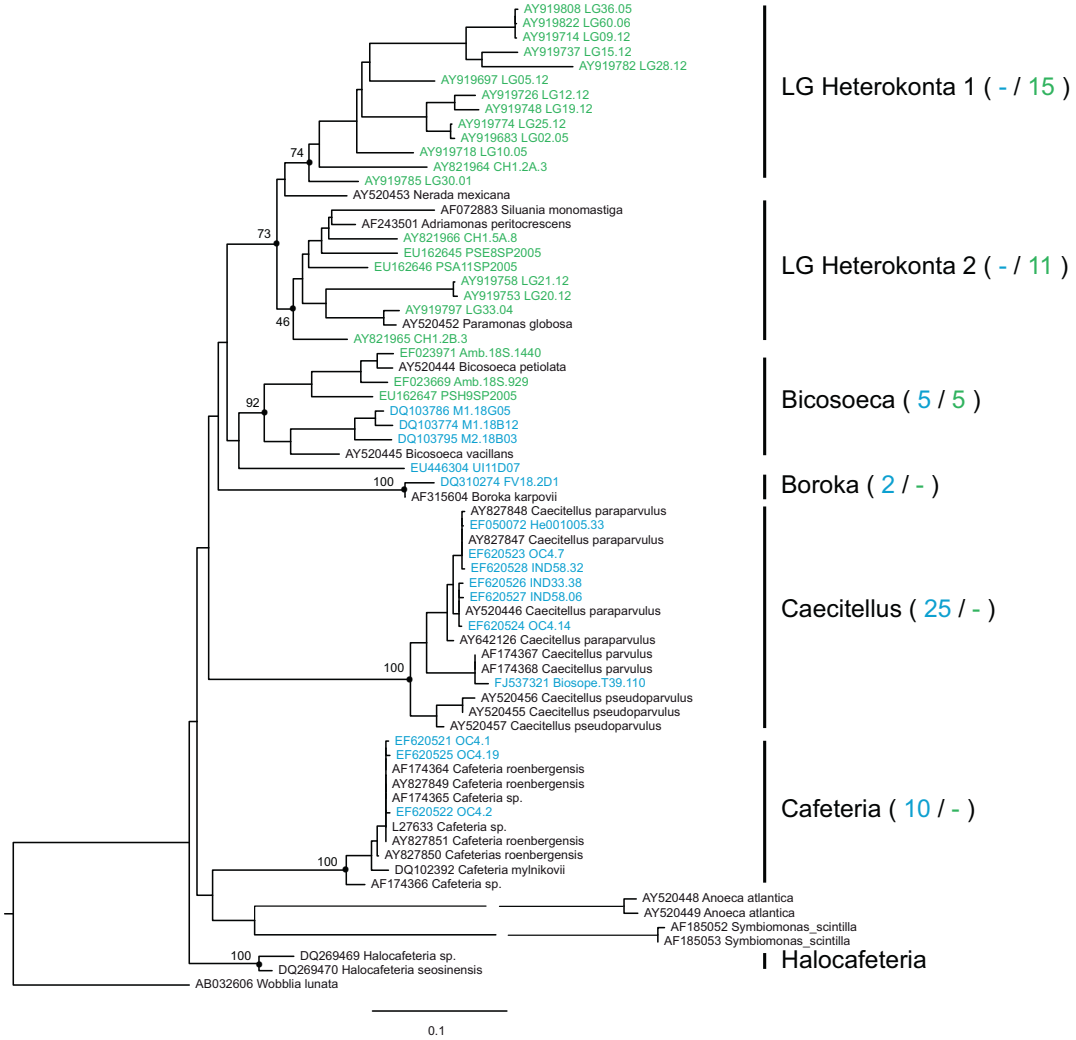


Figure 5. Maximum Likelihood phylogenetic tree of bicosoecids constructed with 66 complete 18S rDNA sequences (1485 informative positions). Legend as in Figure 3.

diversity and reduced the knowledge gaps within these groups. We provide a snapshot of the novelty of the groups that will change in the future depending on the effort of their study.

We first estimated the relative clonal abundance of chrysophytes, choanoflagellates and bicosoecids with respect to other groups of putative heterotrophic flagellates. This exercise should not be translated into absolute abundances, but instead used for a relative comparison among groups. In marine systems, only 5% of clones belonged to chrysophytes, choanoflagellates and bicosoecids, a low number given that these groups were proposed to account for most of the marine heterotrophic flagellates (Arndt et al. 2000; Brandt and Sleigh 2000; Patterson and Lee 2000), and in contrast with the large clonal abundance of the marine uncultured MAST or MALV (Massana and Pedrós-Alió 2008). This contribution could still be lower, since a fraction of environmental chrysophyte sequences could derive from chlorophyll-containing cells (Fuller et al. 2006). Also, half of the studies analyze small protists (Supplementary Table S1) and in these samples the contribution of choanoflagellates could have been underestimated, since these cells are usually larger than 3-5 μm and some are covered by a mineral lorica. However, choanoflagellates are thought to be less abundant than stramenopile flagellates (Arndt et al. 2000; Brandt and Sleigh 2000),

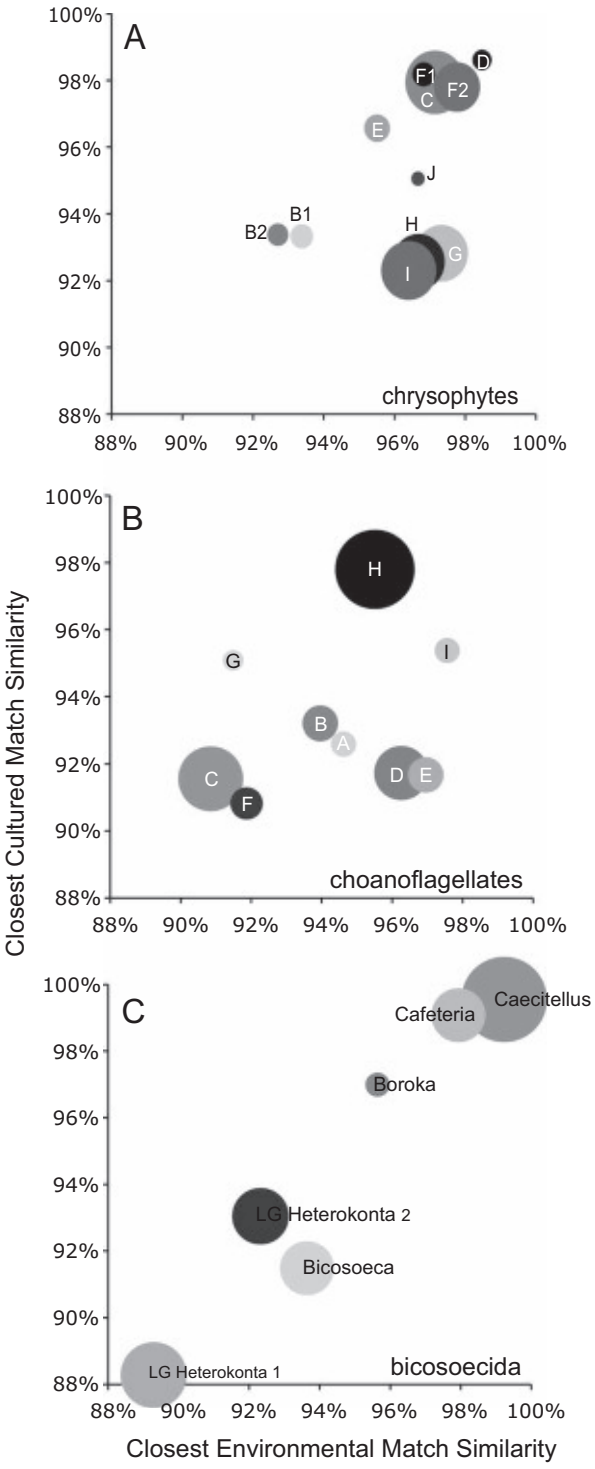


Figure 6. Novelty pattern derived from each described clade within chrysophytes (**A**), choanoflagellates (**B**) and bicosoecids (**C**). Dots representing the novelty of the clades (average similarity against CEM and CCM for all environmental sequences within the clade) have a size proportional to the number of sequences. Different grey tones are used for convenience.

although they may reach up to 20% of the heterotrophic flagellates in polar systems (Leakey et al. 2002). A very different situation occurs in freshwater systems, where bicosoecids represent 22% and chrysophytes 12% of clonal abundance, matching the importance given to these organisms in freshwater systems (Arndt et al. 2000; Carrias et al. 1998).

The estimates of relative clonal abundance suggested that chrysophytes, choanoflagellates and bicosoecids might be less important than expected in marine systems. The presence of these three groups was independently assessed by the analysis of GOS metagenomes (Rusch et al. 2007), which were built by sequencing the environmental DNA directly, and so were free of PCR biases. From the 115 sequences of eukaryotic 18S rDNA retrieved from all samples (Not et al. 2009), only one affiliated with choanoflagellates and two to chrysophytes. As comparison, other groups such as MAST or MALV were much more represented in the GOS metagenomes (15 and 36 sequences, respectively). This PCR-independent approach does not give a definitive answer, either, since it could be strongly affected by the variable copy number of the rDNA operon in different taxa (Zhu et al. 2005). To validate the cell abundance of chrysophytes, choanoflagellates and bicosoecids in the marine plankton, quantitative methods such as FISH (or quantitative-PCR with the proper controls) are needed.

We propose a new approach (Massana et al. 2010) to address the novelty of a given dataset based on the similarity against GenBank sequences. Overall, the novelty displayed by the environmental sequences of each group was rather large, and this was interpreted in terms of efforts in culturing and environmental sequencing. In our context the correspondence of environmental sequences with sequences derived from cultures means that ecologically relevant protists have been cultured. It combines the culturing effort with the ability of a given taxa to grow in the laboratory. In our dataset, such correspondence was apparent only in a few cases, like in marine bicosoecids. A low correspondence between environmental sequences and sequences obtained from cultures was the more common situation, being extreme for freshwater bicosoecids and choanoflagellates

whose environmental sequences only shared 91% similarity with CCM. Enhanced efforts and novel culturing strategies will be needed to bring more ecologically relevant (i.e. abundant) protists into culture, in a similar manner that has been so successful with dominant marine prokaryotes (Könneke et al. 2005; Rappé et al. 2002).

On the other hand, sequencing environmental DNA is relatively straightforward and there are little chances to miss quantitatively important major phylogenetic groups. An insufficient sequencing effort was generally found in our study, with low averaged similarity values of our target sequences against those from other molecular surveys. In addition, similarities against CCM and CEM for different sequence sets were rather similar (Table 1), with the exception of marine chrysophytes for which sequencing was decreasing the novelty. This suggests that there is plenty of room to discover additional diversity for these groups using environmental molecular surveys, which should also take advantage of new high-throughput sequencing technologies (Amaral-Zettler et al. 2009; Stoeck et al. 2009) or use group-specific primers (Bass and Cavalier-Smith 2004). Alternatively, another explanation for low similarity with CEM would be a large endemism of the studied sequences, which might appear only in the studied site. At any rate, our novelty analysis showed that the three protists groups studied here (except marine bicosoecids) need further sequencing effort to reach a full understanding of the in situ diversity.

Our use of environmental sequences from public databases improved the chrysophyte, choanoflagellate and bicosoecid phylogeny and identified emergent new diversity. Thus, four novel clades appeared within chrysophytes, two within choanoflagellates and two within bicosoecids. The tree topologies and clade divisions promise to be very useful as a backbone reference for future studies. An interesting observation from the bicosoecid and choanoflagellate trees was the appearance of a single monophyletic freshwater clade nested within several marine clades. This could be a sign of a single and perhaps ancient transition event from marine to freshwater systems in both protist groups (Logares et al. 2007). In marine systems, chrysophytes harbored an important new diversity, suggesting that uncultured chrysophytes, unlike the easily cultured *Spumella* or *Paraphysomonas*, may be ecologically more relevant (Lim et al. 1999). The same applied for marine choanoflagellates, which showed a great discrepancy between their representation in culture and their abundance in clone libraries. In contrast, marine bicosoecids were highly similar to cultured organisms. Finally, the three groups contained a significant hidden diversity in freshwater systems, specially bicosoecids and choanoflagellates.

In summary, our culture-independent analysis highlighted a large diversity of chrysophytes, choanoflagellates and bicosoecids in aquatic environments that was accompanied with a high novelty degree. This indicated a bias in the representation of cultures and an incomplete sequencing effort for these groups. This analysis should be extended to other protist groups in order to fully benefit from environmental molecular surveys (e.g. Marin and Melkonian 2010). Increasing the effort of environmental sequencing of aquatic protists is already on the research agenda of several laboratories worldwide (Amaral-Zettler et al. 2009; Stoeck et al. 2009). On the other hand, it is equally important to increase the culturing efforts, to match the diversity of protist cultures with the in situ diversity of ecologically relevant protists. Besides culturing efforts, other techniques such as FISH should be applied to assess the abundance and ecological role of new taxa (Chambouvet et al. 2008; Massana et al. 2006). The extent of environmental diversity and novelty is striking even for protist groups that were considered well characterized.

Methods

Sequence dataset retrieval: Environmental 18S rDNA sequences of chrysophytes, choanoflagellates and bicosoecids were obtained from GenBank in a two-step screening. First, sequences found by the NCBI Taxonomy Application were retrieved and checked by BLAST (Altschul et al. 1997) to confirm their placement. Second, we used these and other published sequences from cultures or environmental surveys that belong to the target groups (but are not labeled as such in GenBank) to retrieve additional sequences by BLAST. Putative chimeric sequences were checked by KeyDNATools (www.keydnatools.com) as described before (Guillou et al. 2008). Neighbor Joining phylogenetic trees (see later) were constructed with a wide taxon coverage to find out whether or not ambiguous divergent sequences belong to a given group. Related sequences from cultured organisms were also retrieved from GenBank and pruned to keep only a few representatives for phylogeny.

Two 18S rDNA clone libraries were constructed from dark unamended incubations done in March 2006 and October 2007 with Blanes Bay (Mediterranean Sea) seawater prefiltered by a 3 μm filter. These incubations are known to promote the growth of uncultured HF (Massana et al. 2006). Picoplanktonic biomass was collected on filters, and community DNA was extracted. Complete 18S rDNA genes were PCR-amplified with eukaryote-specific primers, and the PCR products were cloned. Details of the filtering setup, DNA extraction protocol, and PCR and cloning conditions are described elsewhere (Massana et al. 2004, 2006). Twenty-five and 44 clones were partially sequenced with the primer 528f by the MACROGEN

Genomics Sequencing Services. Sequences were identified and inspected for chimeras by BLAST and KeyDNATools, yielding 18 target sequences (accession numbers HQ437173 – HQ437184 and HQ437193 – HQ437196). Ten clones from these libraries and from published libraries (BL in Massana et al. 2004; IND in Not et al. 2008) were completely sequenced with five internal primers by the same service. The final sequence dataset consisted in 395 complete or partial environmental sequences from the three target groups.

Novelty analysis: To infer the novelty of an environmental sequence dataset, we noted for each sequence its similarity in a BLAST search with the closest environmental match (CEM) and the closest cultured match (CCM). The CEM is the first sequence in the output that derives from a molecular survey (excluding those from the same library), and the CCM is the first sequence in the output that belongs to a known organism (often cultured). Both similarity values for all sequences are plotted in a 2D dispersion graph, giving the "novelty pattern" of the dataset. Dots with high CCM similarity (i.e. above 98%) represent environmental sequences close to cultured organisms, whereas dots with low CCM similarity (i.e. below 94%) highlight environmental sequences with no cultured counterpart. Conversely, sequences with high CEM similarity indicate an optimal sequencing effort (they have been found in other environmental surveys), and those with low CEM similarity highlight an insufficient sequencing effort. Finally, the "novelty degree" of the dataset is obtained by averaging the similarity values for all sequences.

Phylogenetic analyses: 18S rDNA sequences were aligned using MAFFT (Katoh et al. 2002) using a close relative as outgroup. Alignments were checked with Seaview 3.2 (Galtier et al. 1996) and highly variable regions of the alignment were removed using Gblocks (Castresana 2000). Neighbor Joining trees were first done with PAUP 4.0b10 (Swofford 2002) with all partial sequences in order to define all possible diversity, and to assure that each clade has at least one clone with the complete sequence. Then, Maximum likelihood (ML) phylogenetic trees with complete sequences were constructed with the fast ML method RAxML (Stamatakis 2006) using the evolutionary model GTRMIXI. Phylogenetic analyses were done in the freely available University of Oslo Bioportal (www.bioportal.uio.no). Repeated runs on distinct starting trees were carried out to select the tree with the best topology (the one having the best Likelihood of 1000 alternative trees). Bootstrap ML analysis was done with 1000 pseudo-replicates and the consensus tree was computed with MrBayes (Huelsenbeck and Ronquist 2001). Trees were edited with FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Acknowledgements

This study was supported by projects GEMMA (CTM2007-63753-C02-01/MAR, MEC) and MICROVIS (CTM2007-62140/MAR, MEC) and the European Funding Agencies from the ERA-net program BiodivERsA under the BioMarKs project. Javier del Campo was funded by I3P program (I3PPRE-06-00676, CSIC). We thank Raquel Rodríguez-Martínez for her help in the unamended incubations, Vanessa Balagué for her laboratory assistance and Joseph Jr. Campo for his help in the English.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.protis.2010.10.003.

References

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

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parrallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS ONE **4**:e6372

Andersen RA, van de Peer Y, Potter D, Sexton JP, Kawachi M, LaJeunesse T (1999) Phylogenetic analysis of the SSU rRNA from members of the Chrysophyceae. Protist **150**:71–84

Arndt H, Dietrich D, Auer B, Cleven EJ, Gräfenhan T, Weitere M, Mylnikov AP (2000) Functional Diversity of Heterotrophic Flagellates in Aquatic Ecosystems. In Leadbeater BSC, Green JC (eds) The Flagellates: Unity, Diversity and Evolution. Taylor & Francis Press, London, pp 240–268

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

Brandt SM, Sleigh MA (2000) The quantitative occurrence of different taxa of heterotrophic flagellates in Southampton water, U.K. Estuar Coast Shelf Sci **51**:91–102

Caron DA, Countway PD, Brown MV (2004) The growing contributions of molecular biology and immunology to protistan ecology: Molecular signatures as ecological tools. J Eukaryot Microbiol **51**:38–48

Carr M, Leadbeater BSC, Hassan R, Nelson M, Baldauf SL (2008) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. Proc Natl Acad Sci USA **105**:16641–16646

Carrias JF, Amblard C, Quiblier-Lloberas C, Bourdier G (1998) Seasonal dynamics of free and attached heterotrophic nanoflagellates in an oligomesotrophic lake. Freshwat Biol **39**:91–101

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17**:540–552

Cavalier-Smith T, Chao EE (2006) Phylogeny and megasystematics of phagotrophic heterokonts (Kingdom Chromista). J Mol Evol **62**:388–420

Chambouvet A, Morin P, Marie D, Guillou L (2008) Control of toxic marine dinoflagellate blooms by serial parasitic killers. Science **322**:1254–1257

Fenchel T (1982) Ecology of heterotrophic microflagellates. IV. Quantitative occurrence and importance as bacterial consumers. Mar Ecol Prog Ser **9**:35–42

Fenchel T, Patterson DJ (1988) *Cafeteria roenbergensis* nov. gen., nov. sp., a heterotrophic microflagellate from marine plankton. Mar Microb Food Webs **3**:9–19

Fuller NJ, Campbell C, Allen DJ, Pitt FD, Zwirglmaier K, LeGall F, Vaulot D, Scanlan DJ (2006) Analysis of photosynthetic picoeukaryotic diversity at open ocean sites in the Arabian Sea using a PCR biased towards marine algal plastids. Aquat Microb Ecol **43**:79–93

Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_Win, two graphic tools for sequence alignment and molecular phylogeny. Comput Applic Biosci **12**:543–548

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R, Scanlan DJ, Worden AZ (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). Environ Microbiol **10**:397–408

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinfomatics **17**:754–755

Jürgens K, Massana R (2008) Protistan Grazing on Marine Bacterioplankton. In Kirchman, DL (ed) Microbial Ecology of the Oceans, Second Edition. John Wiley & Sons, Inc., New York, USA, pp 383–441

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

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, M Marr M, et al. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature **451**: 783–788

Könneke M, Bernhard AE, de la Torre JR, et al. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature **437**:543–546

Leadbeater BSC (1991) Choanoflagellate Organization with Special Reference to the Loricate Taxa. In Patterson DJ, Larson J (eds) The Biology of the Free-living Heterotrophic Flagellates. Clarendon Press, Oxford, pp 241–258

Leakey RJG, Leadbeater BSC, Mitchell E, Mccready SMM, Murray AWA (2002) The abundance and biomass of choanoflagellates and other nanoflagellates in waters of contrasting temperature to the north-west of South Georgia in the Southern Ocean. Eur J Protistol **38**:333–350

Lee JJ, Leedale GF, Bradbury P (2000) The Illustrated Guide to the Protozoa, 2nd ed. Lawrence: Society of Protozoologists

Lefranc M, Thénot A, Lepère C, Debroas D (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. Appl Environ Microbiol **71**:5935–5942

Leipe DD, Winright PO, Gunderson JH, Porter D, Patterson DJ, Valois F (1994) 16S-like rRNA sequences from *Labyrinthuloides minuta* and *Cafeteria roenbergensis*. Phycologia **33**:369–377

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

Logares R, Shalchian-Tabrizi K, Boltovskoy A, Rengefors K (2007) Extensive dinoflagellate phylogenies indicate infrequent marine–freshwater transitions. Mol Phylogen Evol **45**: 887–903

Marin B, Melkonian M (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist **161**:304–336

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

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

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, Pernice M, Bunge J, del Campo J (2010) Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. ISME J doi:10.1038/ismej.2010.104

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

Not F, Latasa M, Scharek R, Viprey M, Karleskind P, Balagué V, Ontoria-Oviedo I, Cumino A, Goetze E, Vaulot D, Massana R (2008) Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. Deep Sea Res Part I **55**:1456–1473

Not F, del Campo J, Balagué V, de Vargas C, Massana R (2009) New insights into the diversity of marine picoeukaryotes. PLoS ONE **4**:e7143

Patterson DJ, Lee WJ (2000) Geographic Distribution and Diversity of Free-Living Heterotrophic Flagellates. In Leadbeater BSC, Green JC (eds) The Flagellates: Unity, Diversity and Evolution. Taylor & Francis Press, London, pp 269–287

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

Preisig HR, Vørs N, Hällfors G (1991) Diversity of Heterotrophic Heterokont Flagellates. In Patterson DJ, Larson J (eds) The Biology of the Free-living Heterotrophic Flagellates. Clarendon Press, Oxford, pp 361–399

Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature **418**:630–633

Richards TA, Vepritskiy AA, Gouliamova DE, Nierzwicki-Bauer SA (2005) The molecular diversity of freshwater picoeukaryotes from oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. Environ Microbiol **7**:1413–1425

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biology **5**:398–431

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

Shi XL, Marie D, Jardillier L, Scanlan DJ, Vaulot D (2009) Groups without cultured representatives dominate eukaryotic

picophytoplankton in the oligotrophic South East Pacific Ocean. PLoS ONE **4**:e7657

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci **272**:2073–2083

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

Stoeck T, Behnke A, Christen R, Amaral-Zettler L, Rodriguez-Mora MJ, Chistoserdov A, Orsiand W, Edgcomb VP (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biology **7**:72

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Mass

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

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

Zuendorf A, Bunge J, Behnke A, Barger KJA, Stoeck T (2006) Diversity estimates of microeukaryotes below the chemocline of the anoxic Mariager Fjord, Denmark. FEMS Microbiol Ecol **58**:476–491

Suplementary table 1. List of all studies from where we have retrieved the environmental sequences for this work. For each study sampling information, the number of libraries (lib) analyzed, the total number or sequences in each study, and the number of sequences within chrysophytes (Chrysos), choanoflagellates (Choanos) and bicosoecids (Bicos) are indicated.

Year	Author	Habitat	System	Size fraction (µm)	lib	Total	Chrysos	Choanos	Bicos	Reference
2001	Díez	Marine	Offshore	0.2 - 2	5	75	1	0	0	Appl Environ Microbiol **68**:4554-4558
2001	López-García	Marine	Offshore	0.2 - 5	5	24	0	0	0	Nature **409**:603-607
2001	Moon-van der Staay	Marine	Offshore	0.2 - 3	1	35	0	2	0	Nature **409**:607-610
2002	Amaral-Zettler	Freshwater	Sediments	Whole	6	33	4	0	0	Nature **417**:137
2002	Dawson	Freshwater	Sediments	Whole	3	125	0	0	0	Proc Natl Acad Sci USA **99**:8324-8329
2002	Edgcomb	Marine	Sediments	Whole	7	275	8	0	1	Proc Natl Acad Sci USA **99**:7658-7662
2003	Stoeck	Marine	Sediments	0.2 - 5	3	80	3	0	0	Appl Environ Microbiol **69**:2657-2663
2003	Stoeck	Marine	Offshore	Whole	3	134	1	0	0	Appl Environ Microbiol **69**:5656-5663
2004	Berney	Freshwater	Sediments	Whole	1	48	0	0	1	BMC Biology **2**:13
2004	Corredor	Marine	Offshore	Whole	2	9	0	1	0	Appl Environ Microbiol **70**:5459-5468
2004	Massana	Marine	Coastal	0.2 -3	4	117	2	5	0	Appl Environ Microbiol **70**:3528-3534
2004	Romari	Marine	Coastal	0.2 - 3	8	409	0	4	0	Limnol Oceanogr **49**: 784-798
2004	Savin	Marine Oxic	Coastal	5 - 100	3	64	0	1	0	Microb Ecol **48**: 51-65
2004	Yuan	Marine	Coastal	Whole	2	64	0	1	0	FEMS Microbiol Let **240**: 163-170
2005	Lefranc	Freshwater	Lake	0.2 - 5	3	56	10	2	1	Appl Environ Microbiol **71**:5935-5942
2005	Luo	Freshwater	Sediments	Whole	3	106	2	0	0	Appl Environ Microbiol **71**:6175-6184
2005	O'Brien	Soil	Soil	Whole	1	181	0	1	0	Appl Environ Microbiol **71**:5544-5550
2005	Richards	Freshwater	Lake	0.2 - 5	12	153	44	0	15	Environ Microbiol **7**: 1413-1425
2005	Slapeta	Freshwater	Sediments / Lake	Whole / 0.2 - 5	7	83	2	2	3	Proc Roy Soc Lond B **272**: 2073-2081
2005	Takishita	Marine	Sediments	Whole	1	29	0	0	0	Extremophyles **9**: 185-196
2005	Unrein	Freshwater	Lake	3 - 20	-	5	5	0	0	Aquat Microb Ecol **40**:269-282
2006	Behnke	Marine	Coastal	Whole	3	183	9	0	1	Appl Environ Microbiol **72**:3626-3636
2006	Brown	Freshwater	Sediments	Whole	4	224	0	0	0	J Eukaryot Microbiol **53**: 420-331
2006	Lovejoy	Marine	Offshore	0.2 - 3	8	245	1	2	0	Appl Environ Microbiol **72**:3085-3095
2006	Massana	Marine	Coastal	0.2 - 3	5	49	6	1	0	Aquat Microb Ecol **45**:171-180
2006	Medlin	Marine	Coastal	0.2 - 3	7	293	6	1	11	Microb Ecol **52**: 53-71
2006	Moon-van der Staay	Soil	Soil	Whole	1	41	0	0	0	FEMS Microbiol Ecol **57**: 420-428
2006	Stoeck	Marine	Offshore	Whole	3	99	0	19	0	Protist **157**: 31-43
2006	Worden	Marine	Coastal	0.45 - 2	5	80	2	0	0	Aquat Microb Ecol **43**:165-175
2006	Zuendorf	Marine	Coastal	Whole	4	113	5	2	3	FEMS Microbiol Ecol **58**: 476-491
2007	Bass	Marine	Offshore / Sediments	Whole	10	24	0	1	0	Proc Roy Soc Lond B **274**: 3069-3077
2007	Countway	Marine	Offshore	0.2 - 200	6	1095	1	1	0	Environ Microbiol **9**: 1219-1232
2007	Lefèvre	Freshwater	Lake	0.6 - 5	4	53	0	0	0	Environ Microbiol **9**: 61-71
2007	López-García	Marine	Offshore	Whole	4	50	0	0	0	Environ Microbiol **9**: 546-554
2007	Massana	Marine	Offshore	0.2 - 3	3	8	0	0	8	Environ Microbiol **9**: 2260-2269

2007	Not	Marine	Offshore	0.2 - 2	6	229	4	0	0	Environ Microbiol **9**: 1233-1252
2007	Stoeck	Marine	Sediments	Whole	5	221	0	0	0	PLoS ONE **2**:e728
2007	Stoeck	Marine	Coastal	Whole	2	202	1	2	4	Microb Ecol **53**: 328-339
2007	Takishita	Freshwater	Saline Lake	Whole	3	37	0	0	0	Protist **158**: 51-64
2007	Takishita	Marine	Sediments	Whole	3	112	3	1	0	Extremophyles **11**: 565-576
2008	Amaral-Zettler	Marine	Coastal	Whole	28	192	5	0	0	Environ Sci Technol **42**: 9072-9080
2008	Chen	Freshwater	Lake	0.8 - 20	6	165	7	1	0	Microb Ecol **56**: 572-583
2008	Guillou	Marine	Coastal	various	61	2276	0	2	0	Environ Microbiol **10**:3349-3365
2008	Lefèvre	Freshwater	Lake	0.6 - 5	6	32	0	0	3	PLoS ONE **3**:e2424
2008	Lepere	Freshwater	Lake	0.2 - 5	2	123	0	8	5	Appl Environ Microbiol **74**:2940-2949
2008	Lesaulnier	Soil	Soil	Whole	1	994	4	6	2	Environ Microbiol **10**: 926-941
2008	Not	Marine	Offshore	0.2 - 3	8	514	17	0	7	Deep Sea Res Part I **55**: 1456-1473
2009	Alexander	Marine	Offshore	Whole	2	119	0	11	1	Environ Microbiol **11**: 360-381
2009	Amacher	Marine	Offshore	Whole	5	291	4	0	0	Deep Sea Res Part I **56**: 2206-2215
2009	Caron	Marine	Offshore	0.2 - 200	2	2332	0	5	0	Appl Environ Microbiol **75**:5797-5808
2009	Costello	Soil	Soil	Whole	4	277	1	0	1	Appl Environ Microbiol **75**:735-747
2009	Luo	Marine	Offshore	0.2 - 50 / Whole	6	281	0	1	0	Hydrobiologia **636**: 233-248
2009	Not	Marine	Coastal	0.6 - 3	2	176	7	0	0	PLoS ONE **4**:e7143
2009	Park	Marine	Sediments	Whole	2	261	0	1	0	J Microbiol **46**: 244-249
2009	Potvin	Marine	Offshore	0.2 - 3	9	135	0	2	0	J Eukaryot Microbiol **56**: 174-181
2009	Shi	Marine	Offshore	0.2 - 3	15	407	37	0	4	PLoS ONE **4**:e7657
2009	Terrado	Marine	Offshore	0.2 - 3 / 3 - whole	8	231	7	2	5	Aquat Microb Ecol **56**:25-39
2009	Zhang	Freshwater	Glacier	Whole	5	41	2	0	0	FEMS Microbiol Ecol 6**7**: 21-29
2010	del Campo	Marine	Coastal	0.2 - 3	2	69	18	0	0	Current study
					292	13015	229	88	76	

Sumpplentary table 2. List of all analyzed sequences (partial and complete). For each sequence we looked by BLAST for its closest match, closes environmental match and closest cultured match and the correspondant similarity. For the sequences analyzed and for the results retrieved by BLAST the Accession Number and Name of the sequence is given. Each sequence is assigned to a phylogenetic group according to our phylogenetic analysis and to an environment according to the information obtained from the original paper.

Chrysophytes

Name	Accession Number	read	Closest match	Accession Number	similarity	%	Closest environmental match	Accession Number	similarity	%	Phylogenetic group	Closest cultured match	Accession Number	similarity	%	Environment
RT1n9	AY082970	1781	Hibberdia magna	M87331	1713/1784	96,0%	SA2_1F7	EF527128	1651/1771	93,2%	Clade B1	Hibberdia magna	M87331	1713/1784	96,0%	Freshwater
PG5.3	AY642726	1271	LG07-07	AY919702	1264/1277	99,0%	LG07-07	AY919702	1264/1277	99,0%	Clade B1	Chrysamoeba mikrokonta	AF123287	1179/1267	93,1%	Freshwater
LG07-07	AY919702	1708	Chrysamoeba mikrokonta	AF123287	1594/1722	92,6%	AND30	AY965871	1516/1650	91,9%	Clade B1	Chrysamoeba mikrokonta	AF123287	1594/1722	92,6%	Freshwater
LG21-07	AY919757	1714	Ochromonas sp. CCMP1393	EF165142	1593/1726	92,3%	10182008-Euk18S-Clone9	AB520724	1533/1662	92,2%	Clade B1	Ochromonas sp. CCMP1393	EF165142	1593/1726	92,3%	Freshwater
LG32-01	AY919791	1716	Mallomonas annulata	U73230	1625/1722	94,4%	Amb_18S_766	EF023425	1621/1733	93,5%	Clade B1	Mallomonas annulata	U73230	1625/1722	94,4%	Freshwater
LG34-01	AY919800	1714	Chrysamoeba mikrokonta	AF123287	1593/1728	92,2%	10182008-Euk18S-Clone9	AB520724	1530/1664	91,9%	Clade B1	Chrysamoeba mikrokonta	AF123287	1593/1728	92,2%	Freshwater
LG48-06	AY919817	1708	Chrysamoeba mikrokonta	AF123287	1593/1721	92,6%	AND30	AY965871	1515/1650	91,8%	Clade B1	Chrysamoeba mikrokonta	AF123287	1593/1721	92,6%	Freshwater
P34.28	AY642697	1776	Chrysamoeba mikrokonta	AF123287	1666/1780	93,6%	SA2_1F7	EF527128	1627/1775	91,7%	Clade B2	Chrysamoeba mikrokonta	AF123287	1666/1780	93,6%	Freshwater
CV1_B1_76	AY821972	1582	Chrysamoeba mikrokonta	AF123287	1540/1582	97,3%	10182008-Euk18S-Clone16	AB520731	1517/1589	95,5%	Clade B2	Chrysamoeba mikrokonta	AF123287	1540/1582	97,3%	Freshwater
LG18-01	AY919742	1724	Spumella-like JBM/S11	EF043285	1604/1731	92,7%	P34.45	AY642705	1612/1741	92,6%	Clade B2	Spumella-like JBM/S11	EF043285	1604/1731	92,7%	Freshwater
LG31-01	AY919789	1724	Spumella-like JBM/S11	EF043285	1603/1731	92,6%	P34.45	AY642705	1611/1741	92,5%	Clade B2	Spumella-like JBM/S11	EF043285	1603/1731	92,6%	Freshwater
LG33-07	AY919798	1719	Chrysamoeba mikrokonta	AF123287	1592/1732	91,9%	P34.45	AY642705	1606/1744	92,1%	Clade B2	Chrysamoeba mikrokonta	AF123287	1592/1732	91,9%	Freshwater
LG43-07	AY919811	1719	Chrysamoeba mikrokonta	AF123287	1592/1732	91,9%	P34.45	AY642705	1590/1731	91,9%	Clade B2	Chrysamoeba mikrokonta	AF123287	1592/1732	91,9%	Freshwater
RT5in4	AY082982	1790	Ochromonas sp. CCMP2761	EF165126	1769/1776	99,6%	HAVOmat-euk48	EF032799	1664/1742	95,5%	Clade C	Ochromonas sp. CCMP2761	EF165126	1769/1776	99,6%	Freshwater
RT5iin35	AY082987	1778	Ochromonas sp. CCMP2761	EF165126	1729/1774	97,5%	HAVOmat-euk48	EF032799	1676/1739	96,4%	Clade C	Ochromonas sp. CCMP2761	EF165126	1729/1774	97,5%	Freshwater
RT5in36	AY082999	1779	Ochromonas sp. CCMP2767	EF165110	1685/1755	96,0%	HAVOmat-euk48	EF032799	1665/1739	95,7%	Clade C	Ochromonas sp. CCMP2767	EF165110	1685/1755	96,0%	Freshwater
A43	AY642741	1273	Ochromonas sp. ACOI-1258	EF165115	1243/1276	97,4%	RT5iin35	AY082987	1221/1275	95,8%	Clade C	Ochromonas sp. ACOI-1258	EF165115	1243/1276	97,4%	Freshwater
A34	AY642745	1305	LG73-06	AY919824	1294/1309	98,9%	LG73-06	AY919824	1294/1309	98,9%	Clade C	Spumella-like JBC07	AY651097	1280/1295	98,8%	Freshwater
Zeuk73	AY916579	1127	Spumella-like JBM/S11	AY651083	1049/1098	95,5%	10182008-Euk18S-Clone18	AB520733	1037/1094	94,8%	Clade C	Spumella-like JBM/S11	AY651083	1049/1098	95,5%	Freshwater
Zeuk72	AY916585	848	Spumella-like JBM/S11	EF043285	814/840	96,9%	10182008-Euk18S-Clone9	AB520724	816/840	97,1%	Clade C	Spumella-like JBM/S11	EF043285	814/840	96,9%	Freshwater
LG10-03	AY919717	1721	Spumella-like JBM/S11	EF043285	1684/1722	97,8%	10182008-Euk18S-Clone9	AB520724	1607/1643	97,8%	Clade C	Spumella-like JBM/S11	EF043285	1684/1722	97,8%	Freshwater
LG10-11	AY919719	1722	Dinobryon sociale	EF165141	1716/1721	99,7%	10182008-Euk18S-Clone9	AB520724	1578/1643	96,0%	Clade C	Dinobryon sociale	EF165141	1716/1721	99,7%	Freshwater
LG18-10	AY919744	1670	Chrysophyceae sp. CCCM41	EF165134	1621/1655	97,9%	CCl40	AY179989	1595/1645	97,0%	Clade C	Chrysophyceae sp. CCCM41	EF165134	1621/1655	97,9%	Freshwater
LG20-09	AY919752	1722	Ochromonas sp. CCMP1393	EF165142	1650/1723	95,8%	AND29	AY994316	1587/1659	95,7%	Clade C	Ochromonas sp. CCMP1393	EF165142	1650/1723	95,8%	Freshwater
LG22-12	AY919762	1717	Ochromonas sp. CCMP2767	EF165110	1715/1717	99,9%	Uncultured chrysophyte	AB455149	1620/1641	98,7%	Clade C	Ochromonas sp. CCMP2767	EF165110	1715/1717	99,9%	Freshwater
LG26-10	AY919777	1720	Uroglena americana	AF123290	1696/1722	98,5%	10182008-Euk18S-Clone9	AB520724	1596/1645	97,0%	Clade C	Uroglena americana	AF123290	1696/1722	98,5%	Freshwater
LG33-02	AY919796	1722	Dinobryon sociale	EF165141	1717/1721	99,8%	10182008-Euk18S-Clone9	AB520724	1588/1652	96,1%	Clade C	Dinobryon sociale	EF165141	1717/1721	99,8%	Freshwater
LG35-11	AY919807	1701	Uroglena americana	EF165131	1676/1693	99,0%	10182008-Euk18S-Clone9	AB520724	1588/1642	96,7%	Clade C	Uroglena americana	EF165131	1676/1693	99,0%	Freshwater
LG73-06	AY919824	1718	Ochromonas sp. CCMP2767	EF165110	1699/1718	98,9%	10182008-Euk18S-Clone21	AB520736	1639/1639	100,0%	Clade C	Ochromonas sp. CCMP2767	EF165110	1699/1718	98,9%	Freshwater
LG81-06	AY919828	1721	Ochromonas sp. CCMP2767	EF165110	1698/1726	98,4%	RT5iin35	AY082987	1642/1726	95,1%	Clade C	Ochromonas sp. CCMP2767	EF165110	1698/1726	98,4%	Freshwater
P80E-2	DQ104072	485	Spumella-like JBM/S11	EF043285	476/485	98,1%	N4aC71	EU333110	475/485	97,9%	Clade C	Spumella-like JBM/S11	EF043285	476/485	98,1%	Freshwater
P200E-19	DQ104087	661	Spumella sp. GOT220	EF027354	651/660	98,6%	10182008-Euk18S-Clone9	AB520724	650/660	98,5%	Clade C	Spumella sp. GOT220	EF027354	651/660	98,6%	Freshwater
Amb_18S_766	EF023425	1776	Spumella sp. Mbc_3C	AB425951	1731/1746	99,1%	AND30	AY965871	1648/1664	99,0%	Clade C	Spumella sp. Mbc_3C	AB425951	1731/1746	99,1%	Freshwater
Amb_18S_772	EF023552	1789	Spumella-like JBM/S11	EF043285	1663/1773	93,8%	LG10-03	AY919717	1604/1730	92,7%	Clade C	Spumella-like JBM/S11	EF043285	1663/1773	93,8%	Freshwater
Amb_18S_936	EF023675	1834	Spumella-like JBM/S11	EF043285	1656/1773	93,4%	LG10-03	AY919717	1597/1730	92,3%	Clade C	Spumella-like JBM/S11	EF043285	1656/1773	93,4%	Freshwater
Amb_18S_6261	EF024085	1782	Poteriochromonas malhamensis	AB023070	1683/1783	94,4%	RT5iin35	AY082987	1638/1785	91,8%	Clade C	Poteriochromonas malhamensis	AB023070	1683/1783	94,4%	Freshwater
LC109.60	EU143916	800	Spumella sp. GOT220	EF027354	769/800	96,1%	N4aC71	EU333110	765/799	95,7%	Clade C	Spumella sp. GOT220	EF027354	769/800	96,1%	Freshwater

WB44.101	EU143929	800	Spumella sp. GOT220	EF027354	797/800	99,6%	N4aC71	EU333110	791/799	99,0%	Clade C	Spumella sp. GOT220	EF027354	797/800	99,6%	Freshwater
L73_ML_034	FJ353894	1048	Uroglena sp. CCMP2768	EF165132	1020/1048	97,3%	HAVOmat-euk48	EF032799	1012/1050	96,4%	Clade C	Uroglena sp. CCMP2768	EF165132	1020/1048	97,3%	Freshwater
Lc2z_ML_106	FJ355238	1049	Uroglena sp. CCMP2768	EF165132	1022/1049	97,4%	HAVOmat-euk48	EF032799	1014/1051	96,5%	Clade C	Uroglena sp. CCMP2768	EF165132	1022/1049	97,4%	Freshwater
051102_S1_W_T_SDP18_098	FJ355409	1048	Spumella-like JBNA46	DQ388542	1043/1048	99,5%	Amb_18S_6261	EF024085	1033/1048	98,6%	Clade C	Spumella-like JBNA46	DQ388542	1043/1048	99,5%	Freshwater
051102_S1_W_T_SDP18_164	FJ355424	1045	Spumella-like JBAF33	AY651077	1044/1045	99,9%	RT5iin35	AY082987	998/1047	95,3%	Clade C	Spumella-like JBAF33	AY651077	1044/1045	99,9%	Freshwater
E12_SE2A	FJ592281	1042	CD8.15	DQ647516	526/528	99,6%	10182008-Euk18S-Clone9	AB520724	1026/1042	98,5%	Clade C	Spumella-like JBM/S11	EF043285	1024/1042	98,3%	Freshwater
ANTNAN_1		454	Ochromonas sp. CCMP1899	EF165133	497/498	99,8%	DL-2-5	EU078267	429/432	99,3%	Clade C	Spumella-like JBM08	AY651098	432/462	93,5%	Freshwater
ANTNAN_14		498	Hydrurus foetidus	FM955256	495/498	99,4%	LG18-10	AY919744	489/498	98,2%	Clade D	Ochromonas sp. CCMP1899	EF165133	497/498	99,8%	Freshwater
ANTNAN_21		497	EBM47.87	FJ410713	482/491	98,2%	LG18-10	AY919744	492/498	98,8%	Clade D	Hydrurus foetidus	FM955256	495/498	99,4%	Freshwater
P34.45	AY642705	1774	Spumella-like JBM08	AY651098	1685/1695	99,4%	LG92-06	AY919829	1697/1717	98,8%	Clade E	Spumella-like JBM08	AY651098	1685/1695	99,4%	Freshwater
LG26-11	AY919778	1713	Chrysosaccus sp. CCMP368	EF165121	1610/1716	93,8%	SA2_1F7	EF527128	1615/1732	93,2%	Clade E	Chrysosaccus sp. CCMP368	EF165121	1610/1716	93,8%	Freshwater
LG92-06	AY919829	1714	P34.45	AY642705	1697/1717	98,8%	P34.45	AY642705	1697/1717	98,8%	Clade E	Spumella-like JBM08	AY651098	1640/1658	98,8%	Freshwater
EB55.130	EU143899	800	Chrysosaccus sp. CCMP295	AF123300	776/795	97,6%	SA2_1F7	EF527128	754/803	93,9%	Clade E	Chrysosaccus sp. CCMP295	AF123300	776/795	97,6%	Freshwater
LC19.29	EU143909	800	Ochromonas tuberculata	AF123293	758/797	95,1%	SA24G10	EF526928	753/803	93,8%	Clade E	Ochromonas tuberculata	AF123293	758/797	95,1%	Freshwater
LC38.36	EU143911	640	Ochromonas tuberculata	AF123293	611/628	97,3%	SA2_1F7	EF527128	610/634	96,2%	Clade E	Ochromonas tuberculata	AF123293	611/628	97,3%	Freshwater
LC121.54	EU143917	800	Ochromonas tuberculata	AF123293	757/798	94,9%	SA24G10	EF526928	752/804	93,5%	Clade E	Ochromonas tuberculata	AF123293	757/798	94,9%	Freshwater
MLB7.146	EU143918	800	Ochromonas tuberculata	AF123293	758/802	94,5%	SA2_1F7	EF527128	748/804	93,0%	Clade E	Ochromonas tuberculata	AF123293	758/802	94,5%	Freshwater
ANTNAN_3		491	Synura curtispina	EF165128	443/523	84,7%	EBM47.87	FJ410713	482/491	98,2%	Clade E	Ochromonas tuberculata	AF123293	478/491	97,4%	Freshwater
P1.35	AY642717	1718	Paraphysomonas bandaiensis	AF109322	1659/1713	96,8%	CCW27	AY180017	1640/1708	96,0%	Clade F1	Paraphysomonas bandaiensis	AF109322	1659/1713	96,8%	Freshwater
PG5.22	AY642735	1193	Paraphysomonas foraminifera	Z38025	1159/1201	96,5%	SA1_4E10	EF526905	1156/1201	96,3%	Clade F1	Paraphysomonas foraminifera	Z38025	1159/1201	96,5%	Freshwater
A42	AY642748	1156	LG47-07	AY919816	1128/1158	97,4%	LG47-07	AY919816	1128/1158	97,4%	Clade F1	Paraphysomonas foraminifera	AF174376	1127/1158	97,3%	Freshwater
LG06-07	AY919699	1722	Paraphysomonas imperforata	AF109323	1705/1725	98,8%	CCW27	AY180017	1647/1667	98,8%	Clade F1	Paraphysomonas imperforata	AF109323	1705/1725	98,8%	Freshwater
LG23-10	AY919766	1723	Paraphysomonas imperforata	AF109323	1671/1726	96,8%	CCW27	AY180017	1616/1671	96,7%	Clade F1	Paraphysomonas imperforata	AF109323	1671/1726	96,8%	Freshwater
LG47-07	AY919816	1724	Paraphysomonas imperforata	AF109323	1708/1725	99,0%	CCW27	AY180017	1650/1667	99,0%	Clade F1	Paraphysomonas imperforata	AF109323	1708/1725	99,0%	Freshwater
LG48-10	AY919818	1711	Paraphysomonas imperforata	AF109323	1629/1728	94,3%	CCW27	AY180017	1579/1672	94,4%	Clade F1	Paraphysomonas imperforata	AF109323	1629/1728	94,3%	Freshwater
MLB92.168	EU143924	800	SA1_4E10	EF526905	772/802	96,3%	SA1_4E10	EF526905	772/802	96,3%	Clade F1	Synura uvella	U73222	764/799	95,6%	Freshwater
A1	AY642746	1211	Paraphysomonas butcheri	AF109326	1181/1214	97,3%	LG46-06	AY919815	1177/1215	96,9%	Clade F2	Paraphysomonas butcheri	AF109326	1181/1214	97,3%	Freshwater
CV1_B1_34	AY821968	1619	Paraphysomonas butcheri	AF109326	1567/1603	97,8%	LG46-06	AY919815	1549/1591	97,4%	Clade F2	Paraphysomonas butcheri	AF109326	1567/1603	97,8%	Freshwater
LG46-06	AY919815	1720	Paraphysomonas butcheri	AF109326	1632/1683	97,0%	P34.45	AY642705	1616/1723	93,8%	Clade F2	Paraphysomonas butcheri	AF109326	1632/1683	97,0%	Freshwater
LG03-12	AY919688	1712	Spumella sp. Mbc_3C	AB425951	1610/1724	93,4%	Biosope_T39.013	FJ537315	1646/1731	95,1%	Clade G	Spumella sp. Mbc_3C	AB425951	1610/1724	93,4%	Freshwater
LG04-04	AY919691	1714	Spumella-like JBM/S11	EF043285	1615/1726	93,6%	Biosope_T39.013	FJ537315	1636/1730	94,6%	Clade G	Spumella-like JBM/S11	EF043285	1615/1726	93,6%	Freshwater
LG12-01	AY919724	1713	Spumella-like JBM/S11	EF043285	1613/1726	93,5%	Biosope_T39.013	FJ537315	1617/1715	94,3%	Clade G	Spumella-like JBM/S11	EF043285	1613/1726	93,5%	Freshwater
LG18-09	AY919743	1714	Spumella-like JBM/S11	EF043285	1614/1725	93,6%	Biosope_T39.013	FJ537315	1636/1730	94,6%	Clade G	Spumella-like JBM/S11	EF043285	1614/1725	93,6%	Freshwater
LG23-07	AY919765	1712	Mallomonas akrokomos	U73229	1617/1720	94,0%	Biosope_T39.013	FJ537315	1636/1735	94,3%	Clade G	Mallomonas akrokomos	U73229	1617/1720	94,0%	Freshwater
LG25-07	AY919772	1716	Spumella-like JBM/S11	EF043285	1615/1728	93,5%	Biosope_T39.013	FJ537315	1636/1731	94,5%	Clade G	Spumella-like JBM/S11	EF043285	1615/1728	93,5%	Freshwater
LG34-04	AY919802	1716	Spumella-like JBM/S11	EF043285	1622/1724	94,1%	Biosope_T39.013	FJ537315	1628/1732	94,0%	Clade G	Spumella-like JBM/S11	EF043285	1622/1724	94,1%	Freshwater
LG34-12	AY919804	1714	Synura uvella	U73222	1616/1724	93,7%	Biosope_T39.013	FJ537315	1636/1731	94,5%	Clade G	Synura uvella	U73222	1616/1724	93,7%	Freshwater
LG44-07	AY919812	1715	Synura uvella	U73222	1622/1724	94,1%	Biosope_T39.013	FJ537315	1643/1718	95,6%	Clade G	Synura uvella	U73222	1622/1724	94,1%	Freshwater
LG44-09	AY919813	1718	Spumella-like JBM/S11	EF043285	1615/1729	93,4%	Biosope_T39.013	FJ537315	1618/1718	94,2%	Clade G	Spumella-like JBM/S11	EF043285	1615/1729	93,4%	Freshwater
P34.48	AY642709	1239	SA2_1F7	EF527128	1214/1239	98,0%	SA2_1F7	EF527128	1214/1239	98,0%	Clade H	Ochromonas tuberculata	AF123293	1175/1239	94,8%	Freshwater
LG02-12	AY919684	1715	Chrysophyceae sp. CCCM41	EF165134	1606/1719	93,4%	Amb_18S_766	EF023425	1615/1741	92,8%	Clade H	Chrysophyceae sp. CCCM41	EF165134	1606/1719	93,4%	Freshwater
LG06-01	AY919698	1715	Chrysophyceae sp. CCCM41	EF165134	1624/1722	94,3%	CCl40	AY179989	1618/1732	93,4%	Clade H	Chrysophyceae sp. CCCM41	EF165134	1624/1722	94,3%	Freshwater
LG12-10	AY919725	1725	Chrysophyceae sp. CCCM41	EF165134	1633/1731	94,3%	CCl40	AY179989	1625/1740	93,4%	Clade H	Chrysophyceae sp. CCCM41	EF165134	1633/1731	94,3%	Freshwater
LG19-10	AY919747	1717	Chrysophyceae sp. CCCM41	EF165134	1614/1721	93,8%	Amb_18S_766	EF023425	1622/1744	93,0%	Clade H	Chrysophyceae sp. CCCM41	EF165134	1614/1721	93,8%	Freshwater
LG22-01	AY919759	1718	Chrysophyceae sp. CCCM41	EF165134	1623/1725	94,1%	Uncultured chrysophyte	AY180010	1572/1673	94,0%	Clade H	Chrysophyceae sp. CCCM41	EF165134	1623/1725	94,1%	Freshwater
LG26-08	AY919776	1722	Mallomonas akrokomos	U73229	1590/1728	92,0%	P34.45	AY642705	1598/1743	91,7%	Clade H	Mallomonas akrokomos	U73229	1590/1728	92,0%	Freshwater
LG35-09	AY919806	1715	Hydrurus foetidus	FM955256	1575/1722	91,5%	CD8.18	DQ647519	1545/1721	89,8%	Clade H	Hydrurus foetidus	FM955256	1575/1722	91,5%	Freshwater
ANTNAN_9		519	Mallomonas annulata	U73230	1650/1790	92,2%	WD4-73	GQ844505	441/521	84,6%	Clade H	Synura curtispina	EF165128	443/523	84,7%	Freshwater
LG21-05	AY919756	1718	CYSGM-8	AB275091	1592/1647	96,7%	CYSGM-8	AB275091	1592/1647	96,7%	Clade J	Oikomonas sp. SA-2.1	AY520450	1630/1723	94,6%	Freshwater
CYSGM-7	AB275090	1653	Spumella-like JBC13	AY651080	1588/1655	96,0%	10182008-Euk18S-Clone9	AB520724	1585/1652	95,9%	Clade C	Spumella-like JBC13	AY651080	1588/1655	96,0%	Marine
He000427_193	AJ965013	537	Spumella-like JBM/S11	EF043285	537/537	100,0%	AND29	AY994316	537/537	100,0%	Clade C	Spumella-like JBM/S11	EF043285	537/537	100,0%	Marine
He000427_18	AJ965015	545	Spumella-like JBM/S11	EF043285	544/545	99,8%	AND29	AY994316	544/545	99,8%	Clade C	Spumella-like JBM/S11	EF043285	544/545	99,8%	Marine
He001206_O28	AJ965058	546	Spumella-like JBM/S11	EF043285	546/546	100,0%	AND29	AY994316	546/546	100,0%	Clade C	Spumella-like JBM/S11	EF043285	546/546	100,0%	Marine
C2_E001	AY046792	1195	Spumella-like JBM19	AY651084	1152/1196	96,3%	10182008-Euk18S-Clone9	AB520724	1153/1196	96,4%	Clade C	Spumella-like JBM19	AY651084	1152/1196	96,3%	Marine
C2_E015	AY046805	1195	Spumella-like JBM19	AY651084	1152/1195	96,4%	10182008-Euk18S-Clone9	AB520724	1153/1195	96,5%	Clade C	Spumella-like JBM19	AY651084	1152/1195	96,4%	Marine
C2_E045	AY046830	1195	Spumella-like JBM19	AY651084	1153/1197	96,3%	10182008-Euk18S-Clone9	AB520724	1154/1197	96,4%	Clade C	Spumella-like JBM19	AY651084	1153/1197	96,3%	Marine

FV36_3A7	DQ310261	1394	Chrysoxys sp. CCMP591	AF123302	1383/1395	99,1%	10182008-Euk18S-Clone18	AB520733	1335/1388	96,2%	Clade C	Chrysoxys sp. CCMP591	AF123302	1383/1395	99,1%	Marine
FV36_CilC7	DQ310291	1326	Ochromonas distigma	EF165136	1294/1309	98,9%	Amb_18S_772	EF023552	1282/1323	96,9%	Clade C	Ochromonas distigma	EF165136	1294/1309	98,9%	Marine
FV23_3A12	DQ310336	1438	Spumella-like JBC13	AY651080	1410/1442	97,8%	10182008-Euk18S-Clone18	AB520733	1409/1442	97,7%	Clade C	Spumella-like JBC13	AY651080	1410/1442	97,8%	Marine
CS123L13	FJ169696	916	Chrysophyta sp. JZH-2007-002	EF633325	914/916	99,8%	10182008-Euk18S-Clone18	AB520733	909/916	99,2%	Clade C	Chrysophyta sp. JZH-2007-002	EF633325	914/916	99,8%	Marine
CS123L36	FJ169701	889	Spumella-like JBC13	AY651080	876/890	98,4%	GoC4_D05	FJ153700	875/890	98,3%	Clade C	Spumella-like JBC13	AY651080	876/890	98,4%	Marine
CS050S12	FJ169705	901	Spumella-like JBM/S11	EF043285	900/902	99,8%	AND29	AY994316	900/902	99,8%	Clade C	Spumella-like JBM/S11	EF043285	900/902	99,8%	Marine
CS050S28	FJ169708	904	Spumella-like JBM/S11	EF043285	902/905	99,7%	AND29	AY994316	902/905	99,7%	Clade C	Spumella-like JBM/S11	EF043285	902/905	99,7%	Marine
CS050L09	FJ169709	873	Spumella-like JBM/S11	EF043285	870/873	99,7%	AND29	AY994316	870/873	99,7%	Clade C	Spumella-like JBM/S11	EF043285	870/873	99,7%	Marine
CS123S01	FJ775653	901	Chrysophyta sp. JZH-2007-002	EF633325	894/901	99,2%	10182008-Euk18S-Clone18	AB520733	889/901	98,7%	Clade C	Chrysophyta sp. JZH-2007-002	EF633325	894/901	99,2%	Marine
CursaT4.41		823	Hydrurus foetidus	FM955256	1633/1770	92,3%	GoC4_D05	FJ153700	789/823	95,9%	Clade C	Spumella-like JBM/S11	EF043285	788/823	95,7%	Marine
CCl40	AY179989	1775	Hydrurus foetidus	FM955256	1710/1785	95,8%	LG18-10	AY919744	1602/1656	96,7%	Clade D	Hydrurus foetidus	FM955256	1710/1785	95,8%	Marine
Uncultured chrysophyte	AY180010	1706	Chrysophyceae sp. CCCM41	EF165134	1676/1686	99,4%	SA2_1H12	EF527131	1552/1560	99,5%	Clade D	Chrysophyceae sp. CCCM41	EF165134	1676/1686	99,4%	Marine
Lc2z_ML_285	FJ355234	1050	SA2_3C8	EF526938	1041/1050	99,1%	SA2_3C8	EF526938	1041/1050	99,1%	Clade D	Chrysophyceae sp. CCCM41	EF165134	1034/1050	98,5%	Marine
CYSGM-6	AB275089	1656	Paraphysomonas foraminifera	AF174376	1649/1656	99,6%	CCW27	AY180017	1646/1657	99,3%	Clade F1	Paraphysomonas foraminifera	AF174376	1649/1656	99,6%	Marine
He000327_98	AJ965002	500	CYSGM-6	AB275089	496/500	99,2%	CYSGM-6	AB275089	496/500	99,2%	Clade F1	Paraphysomonas imperforata	AF109323	495/500	99,0%	Marine
He010218_12	AJ965068	600	Paraphysomonas imperforata	AF109323	600/601	99,8%	LG47-07	AY919816	594/601	98,8%	Clade F1	Paraphysomonas imperforata	AF109323	600/601	99,8%	Marine
CCW27	AY180017	1713	Paraphysomonas foraminifera	AF174376	1697/1707	99,4%	CYSGM-6	AB275089	1646/1657	99,3%	Clade F1	Paraphysomonas foraminifera	AF174376	1697/1707	99,4%	Marine
M4_18B07	DQ103782	1399	Paraphysomonas bandaiensis	AF109322	1375/1398	98,4%	P1.35	AY642717	1368/1398	97,9%	Clade F1	Paraphysomonas bandaiensis	AF109322	1375/1398	98,4%	Marine
M1_18H01	DQ103789	1562	Paraphysomonas bandaiensis	AF109322	1520/1562	97,3%	P1.35	AY642717	1518/1562	97,2%	Clade F1	Paraphysomonas bandaiensis	AF109322	1520/1562	97,3%	Marine
M4_18F06	DQ103808	1400	Paraphysomonas bandaiensis	AF109322	1361/1399	97,3%	P1.35	AY642717	1357/1399	97,0%	Clade F1	Paraphysomonas bandaiensis	AF109322	1361/1399	97,3%	Marine
M3_18G02	DQ103873	1810	Paraphysomonas bandaiensis	AF109322	1688/1766	95,6%	P1.35	AY642717	1643/1722	95,4%	Clade F1	Paraphysomonas bandaiensis	AF109322	1688/1766	95,6%	Marine
M3_18A12	DQ103874	1771	Paraphysomonas bandaiensis	AF109322	1674/1766	94,8%	P1.35	AY642717	1629/1722	94,6%	Clade F1	Paraphysomonas bandaiensis	AF109322	1674/1766	94,8%	Marine
400F5	DQ248164	374	Paraphysomonas sp.	AJ236863	358/372	96,2%	CCW27	AY180017	361/374	96,5%	Clade F1	Paraphysomonas sp.	AJ236863	358/372	96,2%	Marine
FV18_3A1	DQ310204	1397	Paraphysomonas foraminifera	AF174376	1360/1403	96,9%	D2P04E01	EF100249	1359/1404	96,8%	Clade F1	Paraphysomonas foraminifera	AF174376	1360/1403	96,9%	Marine
FV23_1C3	DQ310258	1396	Paraphysomonas foraminifera	AF174376	1387/1396	99,4%	CCW27	AY180017	1383/1397	99,0%	Clade F1	Paraphysomonas foraminifera	AF174376	1387/1396	99,4%	Marine
FV36_CilF11	DQ310307	1339	LG47-07	AY919816	1296/1310	98,9%	LG47-07	AY919816	1296/1310	98,9%	Clade F1	Paraphysomonas foraminifera	Z38025	1286/1294	99,4%	Marine
FV36_2D09	DQ310331	801	M4_18F06	DQ103808	789/801	98,5%	M4_18F06	DQ103808	789/801	98,5%	Clade F1	Paraphysomonas bandaiensis	AF109322	777/801	97,0%	Marine
IND60.63	EU561983	973	Paraphysomonas bandaiensis	AF109322	969/973	99,6%	M4_18B07	DQ103782	958/973	98,5%	Clade F1	Paraphysomonas bandaiensis	AF109322	969/973	99,6%	Marine
IND2.10	EU562152	837	Paraphysomonas foraminifera	AF174376	831/837	99,3%	111.2.32	GQ913148	828/833	99,4%	Clade F1	Paraphysomonas foraminifera	AF174376	831/837	99,3%	Marine
Biosope_T65.123	FJ537347	1751	UEPAC48p3	AY129063	1732/1752	98,9%	UEPAC48p3	AY129063	1732/1752	98,9%	Clade F1	Hydrurus foetidus	FM955256	1617/1757	92,0%	Marine
CursaT4.3		828	GoC4_D05	FJ153700	789/823	95,9%	LG47-07	AY919816	818/826	99,0%	Clade F1	Paraphysomonas imperforata	AF109323	823/826	99,6%	Marine
OA3.18		835	Paraphysomonas imperforata	AF109323	819/838	97,7%	CYSGM-6	AB275089	815/838	97,3%	Clade F1	Paraphysomonas imperforata	AF109323	819/838	97,7%	Marine
OA3.23		839	Paraphysomonas imperforata	AF109323	836/838	99,8%	LG47-07	AY919816	829/838	98,9%	Clade F1	Paraphysomonas imperforata	AF109323	836/838	99,8%	Marine
CursaT4.27		749	Paraphysomonas imperforata	AF109323	823/826	99,6%	CYSGM-6	AB275089	748/749	99,9%	Clade F1	Paraphysomonas foraminifera	AF174376	746/749	99,6%	Marine
NA11-11	AF363223	612	Paraphysomonas butcheri	AF109326	607/612	99,2%	SIF_3F10	EF527051	605/612	98,9%	Clade F2	Paraphysomonas butcheri	AF109326	607/612	99,2%	Marine
He000427_47	AJ965005	547	Paraphysomonas butcheri	AF109326	543/547	99,3%	SIF_3F10	EF527051	541/547	98,9%	Clade F2	Paraphysomonas butcheri	AF109326	543/547	99,3%	Marine
FV18_3B4	DQ310247	1395	Paraphysomonas butcheri	AF109326	1319/1341	98,4%	LG46-06	AY919815	1317/1371	96,1%	Clade F2	Paraphysomonas butcheri	AF109326	1319/1341	98,4%	Marine
FV23_1B7	DQ310257	1392	Paraphysomonas butcheri	AF109326	1316/1340	98,2%	LG46-06	AY919815	1313/1369	95,9%	Clade F2	Paraphysomonas butcheri	AF109326	1316/1340	98,2%	Marine
CD8.06	DQ647511	1763	Biosope_T39.120	FJ537322	1664/1769	94,1%	Biosope_T39.120	FJ537322	1664/1769	94,1%	Clade G	Mallomonas annulata	U73230	1600/1774	90,2%	Marine
AD6S.11	DQ647541	553	Biosope_T84.071	FJ537356	543/551	98,5%	Biosope_T84.071	FJ537356	543/551	98,5%	Clade G	Mallomonas annulata	U73230	524/554	94,6%	Marine
SSRPD64	EF172948	1661	Biosope_T84.071	FJ537356	1628/1661	98,0%	Biosope_T84.071	FJ537356	1628/1661	98,0%	Clade G	Mallomonas annulata	U73230	1546/1674	92,4%	Marine
IND31.17	EU561691	918	Biosope_T39.120	FJ537322	914/918	99,6%	Biosope_T39.120	FJ537322	914/918	99,6%	Clade G	Synura mammillosa	U73220	842/921	91,4%	Marine
IND31.30	EU561703	881	Biosope_T39.120	FJ537322	878/881	99,7%	Biosope_T39.120	FJ537322	878/881	99,7%	Clade G	Synura mammillosa	U73220	800/875	91,4%	Marine
IND31.48	EU561720	868	Biosope_T39.120	FJ537322	865/868	99,7%	Biosope_T39.120	FJ537322	865/868	99,7%	Clade G	Synura mammillosa	U73220	795/870	91,4%	Marine
IND70.35	EU562031	689	Biosope_T39.120	FJ537322	682/689	99,0%	Biosope_T39.120	FJ537322	682/689	99,0%	Clade G	Synura uvella	U73222	636/693	91,8%	Marine
IND2.12	EU562153	840	Biosope_T35.018	FJ537509	807/832	97,0%	Biosope_T35.018	FJ537509	807/832	97,0%	Clade G	Paraphysomonas imperforata	AF109323	778/846	92,0%	Marine
Biosope_T39.013	FJ537315	1779	SSRPD64	EF172948	1613/1662	97,1%	SSRPD64	EF172948	1613/1662	97,1%	Clade G	Mallomonas annulata	U73230	1660/1791	92,7%	Marine
Biosope_T39.120	FJ537322	1780	SSRPD64	EF172948	1626/1661	97,9%	SSRPD64	EF172948	1626/1661	97,9%	Clade G	Mallomonas annulata	U73230	1645/1788	92,0%	Marine
Biosope_T60.011	FJ537338	1750	SSRPD64	EF172948	1602/1666	96,2%	SSRPD64	EF172948	1602/1666	96,2%	Clade G	Hydrurus foetidus	FM955256	1613/1761	91,6%	Marine
Biosope_T60.030	FJ537340	1749	SSRPD64	EF172948	1628/1662	98,0%	SSRPD64	EF172948	1628/1662	98,0%	Clade G	Mallomonas annulata	U73230	1616/1760	91,8%	Marine
Biosope_T65.151	FJ537351	1749	SSRPD64	EF172948	1619/1661	97,5%	SSRPD64	EF172948	1619/1661	97,5%	Clade G	Ochromonas sp. CCMP1393	EF165142	1619/1755	92,3%	Marine
Biosope_T84.071	FJ537356	1779	SSRPD64	EF172948	1628/1661	98,0%	SSRPD64	EF172948	1628/1661	98,0%	Clade G	Haramonas viridis	AB365026	1568/1783	87,9%	Marine
Biosope_T35.018	FJ537509	859	IND2.12	EU562153	807/832	97,0%	IND2.12	EU562153	807/832	97,0%	Clade G	Chrysophyceae sp. CCMP2296	EU247834	798/865	92,3%	Marine
Biosope_T35.061	FJ537518	858	IND31.30	EU561703	854/859	99,4%	IND31.30	EU561703	854/859	99,4%	Clade G	Ochromonas CCMP1278	U42382	740/785	94,3%	Marine
Biosope_T35.062	FJ537519	630	IND31.41	EU561714	627/630	99,5%	IND31.41	EU561714	627/630	99,5%	Clade G	Ochromonadaceae sp. CCMP2298	EU247838	598/629	95,1%	Marine
Biosope_T39.013	FJ537535	862	IND31.30	EU561703	835/862	96,9%	IND31.30	EU561703	835/862	96,9%	Clade G	Synura mammillosa	U73220	791/861	91,9%	Marine

Biospe_T39.120	FJ537559	853	IND31.48	EU561720	850/853	99,6%	IND31.48	EU561720	850/853	99,6%	Clade G	Synura mammillosa	U73220	781/855	91,3%	Marine
Biospe_T60.006	FJ537614	519	cRFM1.28	GQ344712	516/519	99,4%	cRFM1.28	GQ344712	516/519	99,4%	Clade G	Chrysosaccus sp. CCMP1156	EF165120	457/493	92,7%	Marine
Biospe_T60.030	FJ537622	658	cRFM1.28	GQ344712	657/658	99,8%	cRFM1.28	GQ344712	657/658	99,8%	Clade G	Chrysolepidomonas dendrolepidota	AF123297	583/611	95,4%	Marine
Biospe_T60.038	FJ537626	712	IND31.48	EU561720	691/712	97,1%	IND31.48	EU561720	691/712	97,1%	Clade G	Spumella-like JBM/S12	AY651085	604/633	95,4%	Marine
Biospe_T60.045	FJ537630	593	IND70.35	EU562031	587/593	99,0%	IND70.35	EU562031	587/593	99,0%	Clade G	Ochromonas sp. CCMP1393	EF165142	548/575	95,3%	Marine
Biospe_T65.151	FJ537658	796	CD8.06	DQ647511	776/796	97,5%	CD8.06	DQ647511	776/796	97,5%	Clade G	Paraphysomonas imperforata	AF109324	722/807	89,5%	Marine
cRFM1.21	GQ344706	839	SSRPD64	EF172948	839/839	100,0%	SSRPD64	EF172948	839/839	100,0%	Clade G	Ochromonas CCMP1278	U42382	714/754	94,7%	Marine
cRFM1.28	GQ344712	839	C3_E031	AY046860	914/915	99,9%	SSRPD64	EF172948	839/839	100,0%	Clade G	Ochromonas CCMP1278	U42382	714/754	94,7%	Marine
cRFM1.75	GQ344758	905	SSRPD64	EF172948	934/949	98,4%	IND31.17	EU561691	897/905	99,1%	Clade G	Mallomonas annulata	U73230	832/910	91,4%	Marine
cDNA.91	GQ344774	949	IND72.46	EU562123	773/804	96,1%	SSRPD64	EF172948	934/949	98,4%	Clade G	Mallomonas annulata	U73230	875/954	91,7%	Marine
UEPAC48p3	AY129063	1782	Biospe_T65.123	FJ537347	1732/1752	98,9%	Biospe_T65.123	FJ537347	1732/1752	98,9%	Clade H	Hydrurus foetidus	FM955256	1648/1786	92,3%	Marine
UEPAC37p4	AY129065	1781	Biospe_T65.123	FJ537347	1725/1751	98,5%	Biospe_T65.123	FJ537347	1725/1751	98,5%	Clade H	Hydrurus foetidus	FM955256	1639/1785	91,8%	Marine
E222	AY256286	1221	LG35-09	AY919806	1101/1178	93,5%	LG35-09	AY919806	1101/1178	93,5%	Clade H	Ochromonadaceae sp. CCMP2298	EU247838	1095/1205	90,9%	Marine
ENI42482.00342	AY938194	644	CD8.18	DQ647519	639/644	99,2%	CD8.18	DQ647519	639/644	99,2%	Clade H	Ochromonas aestuarii	EF165124	584/647	90,3%	Marine
CD8.18	DQ647519	1771	UEPAC48p3	AY129063	1626/1774	91,7%	UEPAC48p3	AY129063	1626/1774	91,7%	Clade H	Ochromonas aestuarii	EF165124	1616/1781	90,7%	Marine
AD6S.18	DQ647542	552	SIF_4A6	EF527064	535/552	96,9%	SIF_4A6	EF527064	535/552	96,9%	Clade H	Ochromonas aestuarii	EF165124	527/555	95,0%	Marine
N10E01	EF172972	1663	Biospe_T65.123	FJ537347	1662/1663	99,9%	Biospe_T65.123	FJ537347	1662/1663	99,9%	Clade H	Chrysophyceae sp. CCMP2296	EU247834	1544/1670	92,5%	Marine
Q2B03N10	EF172974	1663	Biospe_T65.123	FJ537347	1656/1663	99,6%	Biospe_T65.123	FJ537347	1656/1663	99,6%	Clade H	Chrysophyceae sp. CCMP2296	EU247834	1545/1670	92,5%	Marine
SSRPE02	EF172998	1663	UEPAC48p3	AY129063	1623/1664	97,5%	UEPAC48p3	AY129063	1623/1664	97,5%	Clade H	Chrysophyceae sp. CCMP2296	EU247834	1534/1670	91,9%	Marine
IND31.27	EU561700	864	SSRPE02	EF172998	860/864	99,5%	SSRPE02	EF172998	860/864	99,5%	Clade H	Mallomonas annulata	U73230	807/868	93,0%	Marine
IND31.41	EU561714	888	SSRPE02	EF172998	883/888	99,4%	SSRPE02	EF172998	883/888	99,4%	Clade H	Mallomonas annulata	U73230	828/892	92,8%	Marine
IND70.23	EU562021	832	UEPAC48p3	AY129063	829/832	99,6%	UEPAC48p3	AY129063	829/832	99,6%	Clade H	Paraphysomonas foraminifera	Z38025	773/836	92,5%	Marine
IND70.41	EU562035	896	SSRPE02	EF172998	891/897	99,3%	SSRPE02	EF172998	891/897	99,3%	Clade H	Mallomonas annulata	U73230	833/901	92,5%	Marine
IND72.46	EU562123	872	111.2.76	GQ913170	864/867	99,7%	111.2.76	GQ913170	864/867	99,7%	Clade H	Chromophyton rosanoffi	EF165107	810/875	92,6%	Marine
Biospe_T17.037	FJ537451	753	CD8.18	DQ647519	747/754	99,1%	CD8.18	DQ647519	747/754	99,1%	Clade H	Ochromonas CCMP1278	U42382	623/659	94,5%	Marine
Biospe_T65.123	FJ537647	548	Q2B03N10	EF172974	546/548	99,6%	Q2B03N10	EF172974	546/548	99,6%	Clade H	Chrysophyceae sp. CCMP2296	EU247834	490/552	88,8%	Marine
Biospe_T84.004	FJ537665	809	IND70.23	EU562021	806/809	99,6%	IND70.23	EU562021	806/809	99,6%	Clade H	Paraphysomonas foraminifera	Z38025	737/791	93,2%	Marine
CS123S02	FJ775654	916	E222	AY256286	899/917	98,0%	E222	AY256286	899/917	98,0%	Clade H	Ochromonadaceae sp. CCMP2298	EU247838	783/812	96,4%	Marine
111.1.28	GQ913101	1056	E222	AY256286	993/1028	96,6%	E222	AY256286	993/1028	96,6%	Clade H	Ochromonas aestuarii	EF165124	954/1035	92,2%	Marine
111.1.40	GQ913110	1060	E222	AY256286	1004/1042	96,4%	E222	AY256286	1004/1042	96,4%	Clade H	Ochromonas aestuarii	EF165124	967/1052	91,9%	Marine
111.2.76	GQ913170	1070	E222	AY256286	1041/1077	96,7%	E222	AY256286	1041/1077	96,7%	Clade H	Ochromonadaceae sp. CCMP2298	EU247838	895/924	96,9%	Marine
CD8.S28		544	SSRPD64	EF172948	839/839	100,0%	CD8S.27	DQ647526.1	541/544	99,4%	Clade H	Ochromonas tuberculata	AF123293	512/547	93,6%	Marine
CD8.S17		551	CD8S.27	DQ647526.1	541/544	99,4%	CD8.18	DQ647519	536/539	99,4%	Clade H	Phaeoplaca thallosa	AF123296	509/551	92,4%	Marine
OA3.17		838	IND72.46	EU562123	820/838	97,9%	IND72.46	EU562123	820/838	97,9%	Clade H	Chromophyton rosanoffii	EF165107	777/841	92,4%	Marine
OA3.19		823	IND72.46	EU562123	803/824	97,5%	IND72.46	EU562123	803/824	97,5%	Clade H	Chrysosaccus sp. CCMP1156	EF165120	763/827	92,3%	Marine
OA3.22		824	UEPAC48p3	AY129063	811/824	98,4%	UEPAC48p3	AY129063	811/824	98,4%	Clade H	Chrysophyceae sp. CCCM41	EF165134	764/826	92,5%	Marine
OA3.5		833	E222	AY256286	821/833	98,6%	E222	AY256286	821/833	98,6%	Clade H	Mallomonas annulata	U73230	775/837	92,6%	Marine
CursaT4.21		803	CYSGM-6	AB275089	748/749	99,9%	IND72.46	EU562123	773/804	96,1%	Clade H	Paraphysomonas foraminifera	Z38025	725/806	90,0%	Marine
C3_E006	AY046836	1193	Mallomonas annulata	U73230	1103/1197	92,1%	Biospe_T60.024	FJ537339	1151/1195	96,3%	Clade I	Mallomonas annulata	U73230	1103/1197	92,1%	Marine
C3_E015	AY046844	1194	Hydrurus foetidus	FM955256	1097/1196	91,7%	Biospe_T60.024	FJ537339	1149/1195	96,2%	Clade I	Hydrurus foetidus	FM955256	1097/1196	91,7%	Marine
C3_E023	AY046852	1193	Mallomonas annulata	U73230	1103/1197	92,1%	Biospe_T60.024	FJ537339	1152/1195	96,4%	Clade I	Mallomonas annulata	U73230	1103/1197	92,1%	Marine
C3_E031	AY046860	1193	Mallomonas annulata	U73230	1103/1197	92,1%	Biospe_T60.024	FJ537339	1154/1195	96,6%	Clade I	Mallomonas annulata	U73230	1103/1197	92,1%	Marine
C3_E036	AY046864	1193	Mallomonas annulata	U73230	1103/1197	92,1%	Biospe_T60.024	FJ537339	1152/1195	96,4%	Clade I	Mallomonas annulata	U73230	1103/1197	92,1%	Marine
BL000921.17	AY426840	1777	CD8.18	DQ647519	536/539	99,4%	Biospe_T65.146	FJ537350	1650/1790	92,2%	Clade I	Mallomonas annulata	U73230	1650/1790	92,2%	Marine
NOR50.37	DQ062500	739	CD8.15	DQ647516	729/741	98,4%	CD8.15	DQ647516	729/741	98,4%	Clade I	Spumella-like JBC29	AY651090	666/742	89,8%	Marine
CD8.15	DQ647516	1627	Chrysophyceae sp. CCCM41	EF165134	1482/1636	90,6%	LG18-10	AY919744	1479/1637	90,3%	Clade I	Chrysophyceae sp. CCCM41	EF165134	1482/1636	90,6%	Marine
CD8S.27	DQ647526	544	Biospe_T39.040	FJ537317	522/545	95,8%	Biospe_T39.040	FJ537317	522/545	95,8%	Clade I	Ochromonas tuberculata	AF123293	511/547	93,4%	Marine
IND31.28	EU561701	1758	Ochromonas tuberculata	AF123293	1627/1773	91,8%	Biospe_T60.024	FJ537339	1734/1739	99,7%	Clade I	Hydrurus foetidus	FM955256	1633/1770	92,3%	Marine
IND31.45	EU561718	1754	Paraphysomonas bandaiensis	AF109322	969/973	99,6%	Biospe_T60.024	FJ537339	1606/1746	92,0%	Clade I	Ochromonas tuberculata	AF123293	1627/1773	91,8%	Marine
IND31.49	EU561721	889	104.2.05	GQ913009	836/881	94,9%	104.2.05	GQ913009	836/881	94,9%	Clade I	Ochromonas tuberculata	AF123293	830/895	92,7%	Marine
IND31.90	EU561761	908	Biospe_T65.146	FJ537350	879/909	99,6%	Biospe_T65.146	FJ537350	879/909	99,6%	Clade I	Synura uvella	U73222	835/909	91,9%	Marine
IND31.110	EU561777	828	Biospe_T60.024	FJ537339	799/829	99,6%	Biospe_T60.024	FJ537339	799/829	99,6%	Clade I	Paraphysomonas foraminifera	Z38025	760/830	91,6%	Marine
Biospe_T39.040	FJ537317	1777	Hydrurus foetidus	FM955256	1628/1787	91,1%	CCI40	AY179989	1611/1790	90,0%	Clade I	Hydrurus foetidus	FM955256	1628/1787	91,1%	Marine
Biospe_T39.098	FJ537319	1778	Mallomonas annulata	U73230	1614/1794	90,0%	CCI40	AY179989	1596/1792	89,1%	Clade I	Mallomonas annulata	U73230	1614/1794	90,0%	Marine
Biospe_T60.024	FJ537339	1747	Mallomonas annulata	U73230	1620/1763	91,9%	LG23-10	AY919766	1579/1725	91,5%	Clade I	Mallomonas annulata	U73230	1620/1763	91,9%	Marine
Biospe_T65.104	FJ537343	1747	Hydrurus foetidus	FM955256	1614/1759	91,8%	LG23-10	AY919766	1577/1725	91,4%	Clade I	Hydrurus foetidus	FM955256	1614/1759	91,8%	Marine

Biosope_T65.136	FJ537348	1747	Mallomonas annulata	U73230	1612/1763	91,4%	LG23-10	AY919766	1572/1725	91,1%	Clade I	Mallomonas annulata	U73230	1612/1763	91,4%	Marine
Biosope_T65.146	FJ537350	1747	Mallomonas annulata	U73230	1618/1763	91,8%	LG23-10	AY919766	1577/1725	91,4%	Clade I	Mallomonas annulata	U73230	1618/1763	91,8%	Marine
Biosope_T39.040	FJ537543	817	IND31.28	EU561701	795/817	97,3%	IND31.28	EU561701	795/817	97,3%	Clade I	Paraphysomonas foraminifera	Z38025	719/802	89,7%	Marine
Biosope_T39.098	FJ537552	776	IND31.28	EU561701	751/773	97,2%	IND31.28	EU561701	751/773	97,2%	Clade I	Spumella-like JBC29	AY651090	650/715	90,9%	Marine
Biosope_T60.009	FJ537615	717	IND31.28	EU561701	691/717	96,4%	IND31.28	EU561701	691/717	96,4%	Clade I	Spumella-like JBC29	AY651090	661/699	94,6%	Marine
Biosope_T60.014	FJ537617	730	IND31.28	EU561701	729/730	99,9%	IND31.28	EU561701	729/730	99,9%	Clade I	Paraphysomonas foraminifera	Z38025	663/711	93,2%	Marine
Biosope_T60.019	FJ537618	644	cRFM1.74	GQ344757	639/644	99,2%	cRFM1.74	GQ344757	639/644	99,2%	Clade I	Spumella-like JBC29	AY651090	568/593	95,8%	Marine
Biosope_T60.023	FJ537619	823	IND31.90	EU561761	821/823	99,8%	IND31.90	EU561761	821/823	99,8%	Clade I	Spumella-like JBC29	AY651090	748/798	93,7%	Marine
Biosope_T60.024	FJ537620	546	IND31.110	EU561777	544/545	99,8%	IND31.110	EU561777	544/545	99,8%	Clade I	Paraphysomonas foraminifera	Z38025	491/536	91,6%	Marine
Biosope_T60.027	FJ537621	607	IND31.28	EU561701	606/607	99,8%	IND31.28	EU561701	606/607	99,8%	Clade I	Spumella-like JBC29	AY651090	558/587	95,1%	Marine
Biosope_T60.031	FJ537623	603	IND31.90	EU561761	601/603	99,7%	IND31.90	EU561761	601/603	99,7%	Clade I	Spumella-like JBC29	AY651090	543/573	94,8%	Marine
Biosope_T60.044	FJ537629	663	IND31.28	EU561701	663/663	100,0%	IND31.28	EU561701	663/663	100,0%	Clade I	Spumella-like JBC29	AY651090	608/636	95,6%	Marine
Biosope_T65.064	FJ537638	658	E1aC26	DQ647516	654/658	99,4%	E1aC26	DQ647516	654/658	99,4%	Clade I	Mallomonas annulata	U73230	600/667	90,0%	Marine
RFM1.11	GQ344631	839	IND31.28	EU561701	836/839	99,6%	IND31.28	EU561701	836/839	99,6%	Clade I	Paraphysomonas foraminifera	Z38025	768/840	91,4%	Marine
RFM1.58	GQ344673	840	IND31.28	EU561701	835/840	99,4%	IND31.28	EU561701	835/840	99,4%	Clade I	Spumella-like JBC29	AY651090	733/783	93,6%	Marine
cRFM1.74	GQ344757	915	IND31.17	EU561691	897/905	99,1%	C3_E031	AY046860	914/915	99,9%	Clade I	Mallomonas annulata	U73230	854/919	92,9%	Marine
104.2.05	GQ913009	966	IND31.49	EU561721	836/881	94,9%	IND31.49	EU561721	836/881	94,9%	Clade I	Ochromonas sp. LO128-162	EU076745	847/900	94,1%	Marine
AD6.S29		543	DL-2-5	EU078267	429/432	99,3%	CD8.15	DQ647516	526/528	99,6%	Clade I	Ochromonas tuberculata	AF123293	505/547	92,3%	Marine
OA3.14		837	IND31.49	EU561721	807/838	96,3%	IND31.49	EU561721	807/838	96,3%	Clade I	Spumella-like JBAS36	AY651079	783/844	92,8%	Marine
OA3.10		838	IND31.49	EU561721	808/838	96,4%	IND31.49	EU561721	808/838	96,4%	Clade I	Spumella-like JBAS36	AY651079	783/844	92,8%	Marine
OA3.13		838	IND31.49	EU561721	810/838	96,7%	IND31.49	EU561721	810/838	96,7%	Clade I	Spumella-like JBAS36	AY651079	785/844	93,0%	Marine
OA3.6		1763	Ochromonas aestuarii	EF165124	1632/1777	91,8%	Biosope_T60.024	FJ537339	1612/1741	92,6%	Clade I	Ochromonas aestuarii	EF165124	1632/1777	91,8%	Marine
OA3.9		1764	CD8.15	DQ647516	1621/1627	99,6%	CD8.15	DQ647516	1621/1627	99,6%	Clade I	Hydrurus foetidus	FM955256	1584/1782	88,9%	Marine
CYSGM-8	AB275091	1665	LG21-05	AY919756	1592/1647	96,7%	LG21-05	AY919756	1592/1647	96,7%	Clade J	Oikomonas sp. SA-2.1	AY520450	1591/1667	95,4%	Marine

Choanoflagellates

Name	Accession Number	read	Closest match	Accession Number	similarity	%	Closest environmental match	Accession Number	similarity	%	Phylogenetic group	Closest cultured match	Accession Number	similarity	%	Environment
P1.39	AY642707	1376	Salpingoeca amphoridium	DQ059032	1264/1402	90,2%	Amb_18S_1490	EF024012	1256/1392	90,2%	Clade C	Salpingoeca amphoridium	DQ059032	1264/1402	90,2%	Freshwatwe
PG56	AY642728	1288	Monosiga ovata	AF084230	1218/1290	94,4%	DSGM-66	AB275066	1164/1271	91,6%	Clade C	Monosiga ovata	AF084230	1218/1290	94,4%	Freshwatwe
CV1_B1_36	AY821948	1707	Salpingoeca pyxidium	EU011930	1426/1599	89,2%	Amb_18S_1490	EF024012	1453/1683	86,3%	Clade C	Salpingoeca pyxidium	EU011930	1426/1599	89,2%	Freshwatwe
CV1_B2_17	AY821949	1552	Salpingoeca pyxidium	EU011930	1424/1569	90,8%	DSGM-66	AB275066	1412/1558	90,6%	Clade C	Salpingoeca pyxidium	EU011930	1424/1569	90,8%	Freshwatwe
dfmo4345.018	AY969234	730	B5	EF196791	676/709	95,3%	B5	EF196791	676/709	95,3%	Clade C	Lagenoeca antarctica	DQ995807	683/732	93,3%	Freshwatwe
Amb_18S_720	EF023385	1835	Monosiga ovata	AF271999	1666/1836	90,7%	NPK2_136	EU371175	1618/1835	88,2%	Clade C	Monosiga ovata	AF271999	1666/1836	90,7%	Freshwatwe
Amb_18S_870	EF023626	2115	Monosiga ovata	AF271999	1666/1836	90,7%	NPK2_136	EU371175	1618/1835	88,2%	Clade C	Monosiga ovata	AF271999	1666/1836	90,7%	Freshwatwe
Amb_18S_1307	EF023856	1853	Monosiga ovata	AF271999	1666/1836	90,7%	NPK2_136	EU371175	1618/1835	88,2%	Clade C	Monosiga ovata	AF271999	1666/1836	90,7%	Freshwatwe
Amb_18S_1397	EF023936	1837	Monosiga ovata	AF271999	1666/1836	90,7%	NPK2_136	EU371175	1618/1835	88,2%	Clade C	Monosiga ovata	AF271999	1666/1836	90,7%	Freshwatwe
Amb_18S_1490	EF024012	1835	Monosiga ovata	AF271999	1666/1836	90,7%	NPK2_136	EU371175	1618/1835	88,2%	Clade C	Monosiga ovata	AF271999	1666/1836	90,7%	Freshwatwe
Amb_18S_1493	EF024015	1843	Monosiga ovata	AF271999	1653/1844	89,6%	NPK2_136	EU371175	1609/1843	87,3%	Clade C	Monosiga ovata	AF271999	1653/1844	89,6%	Freshwatwe
BA34	EF196686	680	Amb_18S_1490	EF024012	573/592	96,8%	Amb_18S_1490	EF024012	573/592	96,8%	Clade C	Desmarella moniliformis	AF084231	557/588	94,7%	Freshwatwe
B35	EF196724	901	DSGM-66	AB275066	837/903	92,7%	DSGM-66	AB275066	837/903	92,7%	Clade C	Lagenoeca antarctica	DQ995807	835/903	92,5%	Freshwatwe
B62	EF196742	807	DSGM-66	AB275066	747/813	91,9%	DSGM-66	AB275066	747/813	91,9%	Clade C	Lagenoeca antarctica	DQ995807	745/812	91,7%	Freshwatwe
B64	EF196744	749	DSGM-66	AB275066	690/755	91,4%	DSGM-66	AB275066	690/755	91,4%	Clade C	Lagenoeca antarctica	DQ995807	688/755	91,1%	Freshwatwe
BI42	EF196771	884	DSGM-66	AB275066	825/889	92,8%	DSGM-66	AB275066	825/889	92,8%	Clade C	Lagenoeca antarctica	DQ995807	823/889	92,6%	Freshwatwe
BI53	EF196774	711	DSGM-66	AB275066	652/716	91,1%	DSGM-66	AB275066	652/716	91,1%	Clade C	Lagenoeca antarctica	DQ995807	650/716	90,8%	Freshwatwe
B5	EF196791	957	DSGM-66	AB275066	885/963	91,9%	DSGM-66	AB275066	885/963	91,9%	Clade C	Lagenoeca antarctica	DQ995807	883/963	91,7%	Freshwatwe
B360	EF196793	893	DSGM-66	AB275066	829/894	92,7%	DSGM-66	AB275066	829/894	92,7%	Clade C	Lagenoeca antarctica	DQ995807	827/894	92,5%	Freshwatwe
WB4502	EU144008	800	Monosiga ovata	AF084230	739/800	92,4%	NPK2_136	EU371175	732/792	92,4%	Clade C	Monosiga ovata	AF084230	739/800	92,4%	Freshwatwe
OLI11041	AJ402325	1758	Monosiga brevicollis	AF100940	1597/1789	89,3%	RM2-SGM70	AB505578	1520/1685	90,2%	Clade A	Monosiga brevicollis	AF100940	1597/1789	89,3%	Marine

500F5	DQ248177	470	Salpingoeca infusionum	AF100941	466/470	99,1%	IAFDv90	AY835693	423/427	99,1%	Clade A	Salpingoeca infusionum	AF100941	466/470	99,1%	Marine
FV36_CilF8	DQ310311	1313	OLI11041	AJ402325	1239/1312	94,4%	OLI11041	AJ402325	1239/1312	94,4%	Clade A	Acanthoeca spectabilis	AF084233	1183/1324	89,4%	Marine
DSGM-66	AB275066	1662	Lagenoeca antarctica	DQ995807	1346/1352	99,6%	NPK2_136	EU371175	1525/1666	91,5%	Clade B	Lagenoeca antarctica	DQ995807	1346/1352	99,6%	Marine
He000803_1	AJ965245	543	BL010625.36	AY426933.1	542/544	99,6%	BL010625.36	AY426933.1	542/544	99,6%	Clade B	Lagenoeca antarctica	DQ995807	505/544	92,8%	Marine
c9d2t4	AY378159	521	NS371L73	AJ829808	489/517	94,6%	NS371L73	AJ829808	489/517	94,6%	Clade B	Diaphanoeca grandis	AF084234	488/522	93,5%	Marine
BL000921.20	AY426842	1779	DSGM-66	AB275066	1551/1663	93,3%	DSGM-66	AB275066	1551/1663	93,3%	Clade B	Monosiga brevicollis	AF100940	1628/1784	91,3%	Marine
BL010625.36	AY426933	1786	DSGM-66	AB275066	1540/1668	92,3%	DSGM-66	AB275066	1540/1668	92,3%	Clade B	Diaphanoeca grandis	AF084234	1628/1806	90,1%	Marine
9_132	EU545768	972	DSGM-66	AB275066	909/987	92,1%	DSGM-66	AB275066	909/987	92,1%	Clade B	Lagenoeca antarctica	DQ995807	892/970	92,0%	Marine
RA000412.126	AY295370	545	05M101r1.08	EU682617	511/524	97,5%	05M101r1.08	EU682617	511/524	97,5%	Clade D	Diaphanoeca grandis	AF084234	499/551	90,6%	Marine
RA010613.125	AY295679	550	05M101r1.08	EU682617	523/535	97,8%	05M101r1.08	EU682617	523/535	97,8%	Clade D	Diaphanoeca grandis	AF084234	503/556	90,5%	Marine
RA010613.3	AY295700	550	05M101n.21	EU682616	530/531	99,8%	05M101n.21	EU682616	530/531	99,8%	Clade D	Diaphanoeca grandis	AF084234	504/557	90,5%	Marine
RA010613.56	AY295712	550	05M101r1.08	EU682617	521/533	97,7%	SIF_2B10	EF527008	527/548	96,2%	Clade D	Diaphanoeca grandis	AF084234	503/557	90,3%	Marine
BL000921.30	AY426848	1787	NOR46.34	DQ120005	1713/1758	97,4%	NOR46.34	DQ120005	1713/1758	97,4%	Clade D	Diaphanoeca grandis	AF084234	1652/1800	91,8%	Marine
NOR46.34	DQ120005	1764	SCM28C135	AY665096	1629/1749	93,1%	SCM28C135	AY665096	1629/1749	93,1%	Clade D	Diaphanoeca grandis	AF084234	1630/1769	92,1%	Marine
ENVP10203.00258	DQ918052	437	SCM28C135	AY665096	416/437	95,2%	SCM28C135	AY665096	416/437	95,2%	Clade D	Lagenoeca antarctica	DQ995807	395/436	90,6%	Marine
05M101n.21	EU682616	947	SIF_2B10	EF527008	922/946	97,5%	SIF_2B10	EF527008	922/946	97,5%	Clade D	Diaphanoeca grandis	AF084234	879/956	91,9%	Marine
05M101r1.08	EU682617	903	SIF_2B10	EF527008	884/905	97,7%	SIF_2B10	EF527008	884/905	97,7%	Clade D	Diaphanoeca grandis	AF084234	843/913	92,3%	Marine
BAFRACTnano 3	EU785325	545	SCM28C135	AY665096	512/545	93,9%	SCM28C135	AY665096	512/545	93,9%	Clade D	Diaphanoeca grandis	AF084234	501/553	90,6%	Marine
BAFRACTnano 5	EU785327	546	SCM28C135	AY665096	513/546	94,0%	SCM28C135	AY665096	513/546	94,0%	Clade D	Diaphanoeca grandis	AF084234	502/554	90,6%	Marine
GoC3_C08	FJ153672	1239	NOR46.34	DQ120005	1065/1109	96,0%	NOR46.34	DQ120005	1065/1109	96,0%	Clade D	Diaphanoeca grandis	AF084234	1021/1103	92,6%	Marine
CS050S19	FJ169746	964	NOR46.34	DQ120005	854/898	95,1%	NOR46.34	DQ120005	854/898	95,1%	Clade D	Diaphanoeca grandis	AF084234	831/880	94,4%	Marine
MO010_1.00381	GQ382459	653	CS050S19	FJ169746	623/651	95,7%	CS050S19	FJ169746	623/651	95,7%	Clade D	Diaphanoeca grandis	AF084234	539/565	95,4%	Marine
OLI11013	AJ402331	1780	Diaphanoeca grandis	AF084234	1641/1800	91,2%	MA1_CilE8	EF526984	1606/1782	90,1%	Clade E	Diaphanoeca grandis	AF084234	1641/1800	91,2%	Marine
BL000921.24	AY426845	1789	OLI11013	AJ402331	1750/1788	97,9%	OLI11013	AJ402331	1750/1788	97,9%	Clade E	Diaphanoeca grandis	AF084234	1666/1802	92,5%	Marine
BL001221.16	AY426868	1778	OLI11013	AJ402331	1742/1776	98,1%	OLI11013	AJ402331	1742/1776	98,1%	Clade E	Acanthocoepsis unguiculata	L10823	1642/1781	92,2%	Marine
MO010_1.00039	GQ382463	544	BL001221.16	AY426868	535/544	98,3%	BL001221.16	AY426868	535/544	98,3%	Clade E	Calliacantha sp. CEE-2003	AF272000	499/549	90,9%	Marine
MO010_1.00077	GQ382493	609	BL001221.16	AY426868	600/609	98,5%	BL001221.16	AY426868	600/609	98,5%	Clade E	Calliacantha sp. CEE-2003	AF272000	564/614	91,9%	Marine
MO010_1.00078	GQ382494	581	BL001221.16	AY426868	572/581	98,5%	BL001221.16	AY426868	572/581	98,5%	Clade E	Calliacantha sp. CEE-2003	AF272000	536/586	91,5%	Marine
UI12G07	EU446321	1558	SCM28C135	AY665096	1432/1562	91,7%	SCM28C135	AY665096	1432/1562	91,7%	Clade F	Acanthocoepsis unguiculata	L10823	1420/1562	90,9%	Marine
UI13H07	EU446354	1557	SCM28C135	AY665096	1434/1566	91,6%	SCM28C135	AY665096	1434/1566	91,6%	Clade F	Acanthocoepsis unguiculata	L10823	1422/1566	90,8%	Marine
cLA12A08	EU446377	1556	DSGM-66	EU446411	1397/1507	92,7%	DSGM-66	EU446411	1397/1507	92,7%	Clade F	Acanthocoepsis unguiculata	L10823	1418/1560	90,9%	Marine
cLA12E05	EU446385	1472	M1_18E10	DQ103820	1357/1485	91,4%	M1_18E10	DQ103820	1357/1485	91,4%	Clade F	Acanthoeca spectabilis	AF084233	1336/1476	90,5%	Marine
cLA14H07	EU446411	1556	SCM28C135	AY665096	1435/1564	91,8%	SCM28C135	AY665096	1435/1564	91,8%	Clade F	Acanthocoepsis unguiculata	L10823	1423/1564	91,0%	Marine
FV23_1A4	DQ310214	1393	Savillea micropora	EU011928	1331/1397	95,3%	OLI11013	AJ402331	1268/1397	90,8%	Clade G	Savillea micropora	EU011928	1331/1397	95,3%	Marine
cLA12G11	EU446388	1541	Acanthocoepsis unguiculata	L10823	1466/1545	94,9%	NPK2_136	EU371175	1423/1545	92,1%	Clade G	Acanthocoepsis unguiculata	L10823	1466/1545	94,9%	Marine
NS371L73	AJ829808	691	NPK2_136	EU371175	675/693	97,4%	NPK2_136	EU371175	675/693	97,4%	Clade H	Diaphanoeca grandis	AF084234	663/689	96,2%	Marine
M1_18E10	DQ103820	1546	NPK2_136	EU371175	1501/1548	97,0%	NPK2_136	EU371175	1501/1548	97,0%	Clade H	Diaphanoeca grandis	AF084234	1489/1559	95,5%	Marine
MD65.31	DQ120004	796	FV36_2D08	DQ310249	783/793	98,7%	FV36_2D08	DQ310249	783/793	98,7%	Clade H	Diaphanoeca grandis	AF084234	781/793	98,5%	Marine
500D3m13	DQ248175	468	NS371L73	AJ829808	459/468	98,1%	NS371L73	AJ829808	459/468	98,1%	Clade H	Diaphanoeca grandis	AF084234	455/471	96,6%	Marine
FV18_1F5	DQ310210	996	Diaphanoeca grandis	AF084234	936/946	98,9%	NOR46.34	DQ120005	882/948	93,0%	Clade H	Diaphanoeca grandis	AF084234	936/946	98,9%	Marine
FV36_2B09	DQ310239	1399	Diaphanoeca grandis	AF084234	1390/1401	99,2%	DB25_BASS	EU154974	1201/1256	95,6%	Clade H	Diaphanoeca grandis	AF084234	1390/1401	99,2%	Marine
FV36_2A12	DQ310248	1398	Diaphanoeca grandis	AF084234	1385/1399	99,0%	DB25_BASS	EU154974	1199/1255	95,5%	Clade H	Diaphanoeca grandis	AF084234	1385/1399	99,0%	Marine
FV36_2D08	DQ310249	1398	Diaphanoeca grandis	AF084234	1385/1393	99,4%	DB25_BASS	EU154974	1204/1253	96,1%	Clade H	Diaphanoeca grandis	AF084234	1385/1393	99,4%	Marine
FV36_2F06	DQ310266	996	Diaphanoeca grandis	AF084234	938/946	99,2%	DB25_BASS	EU154974	782/797	98,1%	Clade H	Diaphanoeca grandis	AF084234	938/946	99,2%	Marine
FV36_CilB9	DQ310285	1331	Diaphanoeca grandis	AF084234	1319/1330	99,2%	NOR46.34	DQ120005	1229/1324	92,8%	Clade H	Diaphanoeca grandis	AF084234	1319/1330	99,2%	Marine
FV36_CilC10	DQ310286	1333	Diaphanoeca grandis	AF084234	1327/1332	99,6%	NOR46.34	DQ120005	1234/1324	93,2%	Clade H	Diaphanoeca grandis	AF084234	1327/1332	99,6%	Marine
FV36_CilC11	DQ310287	1333	Diaphanoeca grandis	AF084234	1326/1332	99,5%	NOR46.34	DQ120005	1233/1324	93,1%	Clade H	Diaphanoeca grandis	AF084234	1326/1332	99,5%	Marine
FV36_CilD7	DQ310289	1333	Diaphanoeca grandis	AF084234	1325/1332	99,5%	NOR46.34	DQ120005	1232/1324	93,1%	Clade H	Diaphanoeca grandis	AF084234	1325/1332	99,5%	Marine
FV36_CilA8	DQ310290	1333	Diaphanoeca grandis	AF084234	1325/1332	99,5%	NOR46.34	DQ120005	1233/1324	93,1%	Clade H	Diaphanoeca grandis	AF084234	1325/1332	99,5%	Marine
FV36_CilA12	DQ310302	1332	Diaphanoeca grandis	AF084234	1321/1332	99,2%	OLI11013	AJ402331	1233/1337	92,2%	Clade H	Diaphanoeca grandis	AF084234	1321/1332	99,2%	Marine
FV36_CilF10	DQ310306	1334	Diaphanoeca grandis	AF084234	1327/1333	99,5%	NOR46.34	DQ120005	1233/1325	93,1%	Clade H	Diaphanoeca grandis	AF084234	1327/1333	99,5%	Marine
FV36_CilH9	DQ310309	1332	Diaphanoeca grandis	AF084234	1325/1332	99,5%	NOR46.34	DQ120005	1231/1323	93,0%	Clade H	Diaphanoeca grandis	AF084234	1325/1332	99,5%	Marine
FV36_CilG10	DQ310312	1333	Diaphanoeca grandis	AF084234	1327/1332	99,6%	NOR46.34	DQ120005	1233/1324	93,1%	Clade H	Diaphanoeca grandis	AF084234	1327/1332	99,6%	Marine
FV36_CilE11	DQ310313	1334	Diaphanoeca grandis	AF084234	1328/1333	99,6%	NOR46.34	DQ120005	1232/1325	93,0%	Clade H	Diaphanoeca grandis	AF084234	1328/1333	99,6%	Marine
FV36_CilH12	DQ310315	1333	Diaphanoeca grandis	AF084234	1324/1332	99,4%	NOR46.34	DQ120005	1233/1324	93,1%	Clade H	Diaphanoeca grandis	AF084234	1324/1332	99,4%	Marine
FV36_CilD8D9D12	DQ310339	1334	Diaphanoeca grandis	AF084234	1326/1332	99,5%	NOR46.34	DQ120005	1237/1326	93,3%	Clade H	Diaphanoeca grandis	AF084234	1326/1332	99,5%	Marine

PD6.09	DQ647529	875	DB25_BASS	EU154974	871/875	99,5%	DB25_BASS	EU154974	871/875	99,5%	Clade H	Diaphanoeca grandis	AF084234	839/882	95,1%	Marine
DB25_BASS	EU154974	1258	NPK2_136	EU371175	1241/1255	98,9%	NPK2_136	EU371175	1241/1255	98,9%	Clade H	Diaphanoeca grandis	AF084234	1214/1264	96,0%	Marine
NPK2_136	EU371175	1774	Diaphanoeca grandis	AF084234	1701/1783	95,4%	M1_18E10	DQ103820	1501/1548	97,0%	Clade H	Diaphanoeca grandis	AF084234	1701/1783	95,4%	Marine
UI11E03	EU446305	1544	NPK2_136	EU371175	1486/1544	96,2%	NPK2_136	EU371175	1486/1544	96,2%	Clade H	Diaphanoeca grandis	AF084234	1485/1553	95,6%	Marine
UI13A05	EU446337	1541	NPK2_136	EU371175	1490/1548	96,3%	NPK2_136	EU371175	1490/1548	96,3%	Clade H	Diaphanoeca grandis	AF084234	1487/1557	95,5%	Marine
cLA12B02	EU446378	1547	Diaphanoeca grandis	AF084234	1496/1561	95,8%	NPK2_136	EU371175	1492/1548	96,4%	Clade H	Diaphanoeca grandis	AF084234	1496/1561	95,8%	Marine
cLA14G03	EU446410	1544	NPK2_136	EU371175	1488/1544	96,4%	NPK2_136	EU371175	1488/1544	96,4%	Clade H	Diaphanoeca grandis	AF084234	1487/1553	95,8%	Marine
CS050L11	FJ169749	940	DB25_BASS	EU154974	934/940	99,4%	DB25_BASS	EU154974	934/940	99,4%	Clade H	Diaphanoeca grandis	AF084234	900/950	94,7%	Marine
MO010_42.00292	GQ382897	640	DB25_BASS	EU154974	628/641	98,0%	DB25_BASS	EU154974	628/641	98,0%	Clade H	Diaphanoeca grandis	AF084234	608/649	93,7%	Marine
m311	AY331788	1218	SA1_4D9	EF526904	794/812	97,8%	SA1_4D9	EF526904	794/812	97,8%	Clade I	Stephanoeca diplocostata	AF084235	1085/1123	96,6%	Marine
M1_18A02	DQ103821	1556	NIF_2H2	EF526731	1545/1556	99,3%	NIF_2H2	EF526731	1545/1556	99,3%	Clade I	Stephanoeca diplocostata	AY149899	1487/1569	94,8%	Marine
UI13C07	EU446341	1537	Stephanoeca diplocostata	AY149899	1476/1558	94,7%	M1_18A02	DQ103821	1449/1518	95,5%	Clade I	Stephanoeca diplocostata	AY149899	1476/1558	94,7%	Marine

Bicosoecids

Name	Accession Number	read	Closest match	Accession Number	similarity	%	Closest environmental match	Accession Number	similarity	%	Phylogenetic group	Closest cultured match	Accession Number	similarity	%	Environment
Sey077	AY605212	802	PSH9SP2005	EU162647	764/795	96,1%	PSH9SP2005	EU162647	764/795	96,1%	Bicosoeca	Bicosoeca petiolata	AY520444	741/808	91,7%	Freshwater
H05_SE4A	FJ592448	1083	Bicosoeca vacillans	AY520445	1440/1578	91,3%	PSH9SP2005	EU162647	1046/1086	96,3%	Bicosoeca	Bicosoeca vacillans	AY520445	996/1109	89,8%	Freshwater
PSH9SP2005	EU162647	1791	Amb_18S_929	EF023669	1641/1807	90,8%	Amb_18S_929	EF023669	1641/1807	90,8%	Bicosoeca	Bicosoeca vacillans	AY520445	1603/1801	89,0%	Freshwater
Amb_18S_1440	EF023971	1863	PSH9SP2005	EU162648	1631/1811	90,1%	PSH9SP2005	EU162648	1631/1811	90,1%	Bicosoeca	Bicosoeca petiolata	AY520444	1736/1805	96,2%	Freshwater
Amb_18S_929	EF023669	1884	PSH9SP2005	EU162647	1641/1807	90,8%	PSH9SP2005	EU162647	1641/1807	90,8%	Bicosoeca	Bicosoeca petiolata	AY520444	1702/1814	93,8%	Freshwater
B60	EF196740	952	LG10-05	AY919718	878/957	91,7%	LG10-05	AY919718	878/957	91,7%	Freshwater A	Tabularia tabulata	AY485475	832/979	85,0%	Freshwater
BA125	EF196695	837	CH1_2A_3	AY821964	794/800	99,3%	CH1_2A_3	AY821964	794/800	99,3%	Freshwater A	Adriamonas peritocrescens	AF243501	724/846	85,6%	Freshwater
CH1_2A_3	AY821964	1587	LG10-05	AY919718	1417/1569	90,3%	LG10-05	AY919718	1417/1569	90,3%	Freshwater A	Adriamonas peritocrescens	AF243501	1409/1583	89,0%	Freshwater
LG05-12	AY919697	1671	Adriamonas peritocrescens	AF243501	1480/1650	89,7%	PSA11SP2005	EU162646	1525/1719	88,7%	Freshwater A	Adriamonas peritocrescens	AF243501	1480/1650	89,7%	Freshwater
LG19-12	AY919748	1677	Adriamonas peritocrescens	AF243501	1480/1649	89,8%	PSA11SP2005	EU162646	1514/1716	88,2%	Freshwater A	Adriamonas peritocrescens	AF243501	1480/1649	89,8%	Freshwater
LG28-12	AY919782	1678	Nerada mexicana	AY520453	1470/1711	85,9%	PSA11SP2005	EU162646	1479/1730	85,5%	Freshwater A	Nerada mexicana	AY520453	1470/1711	85,9%	Freshwater
LG09-12	AY919714	1679	Nerada mexicana	AY520453	1479/1707	86,6%	PSA11SP2005	EU162646	1481/1719	86,2%	Freshwater A	Nerada mexicana	AY520453	1479/1707	86,6%	Freshwater
LG36-05	AY919808	1679	Nerada mexicana	AY520453	1477/1708	86,5%	PSA11SP2005	EU162646	1479/1720	86,0%	Freshwater A	Nerada mexicana	AY520453	1477/1708	86,5%	Freshwater
LG60-06	AY919822	1679	Nerada mexicana	AY520453	1480/1710	86,5%	PSA11SP2005	EU162646	1482/1719	86,2%	Freshwater A	Nerada mexicana	AY520453	1480/1710	86,5%	Freshwater
LG12-12	AY919726	1680	Adriamonas peritocrescens	AF243501	1486/1648	90,2%	PSA11SP2005	EU162646	1466/1640	89,4%	Freshwater A	Adriamonas peritocrescens	AF243501	1486/1648	90,2%	Freshwater
LG15-12	AY919737	1683	Nerada mexicana	AY520453	1482/1713	86,5%	PSA11SP2005	EU162646	1494/1734	86,2%	Freshwater A	Nerada mexicana	AY520453	1482/1713	86,5%	Freshwater
LG02-05	AY919683	1685	Adriamonas peritocrescens	AF243501	1483/1652	89,8%	PSA11SP2005	EU162646	1525/1719	88,7%	Freshwater A	Adriamonas peritocrescens	AF243501	1483/1652	89,8%	Freshwater
LG25-12	AY919774	1686	Adriamonas peritocrescens	AF243501	1489/1653	90,1%	PSA11SP2005	EU162646	1519/1720	88,3%	Freshwater A	Adriamonas peritocrescens	AF243501	1489/1653	90,1%	Freshwater
LG30-01	AY919785	1686	Nerada mexicana	AY520453	1561/1711	91,2%	PSA11SP2005	EU162646	1555/1722	90,3%	Freshwater A	Nerada mexicana	AY520453	1561/1711	91,2%	Freshwater
LG10-05	AY919718	1688	PSA11SP2005	EU162646	1554/1660	93,6%	PSA11SP2005	EU162646	1554/1660	93,6%	Freshwater A	Nerada mexicana	AY520453	1518/1657	91,6%	Freshwater
B4	EF196790	712	EBF23.97	FJ410660	667/681	97,9%	EBF23.97	FJ410660	667/681	97,9%	Freshwater B	Siluania monomastiga	AF072883	675/715	94,4%	Freshwater
BI14	EF196802	951	P34.6	AY642710	892/952	99,7%	P34.6	AY642710	892/952	99,7%	Freshwater B	Adriamonas peritocrescens	AF243501	888/956	92,9%	Freshwater
B480	EF196795	971	P34.6	AY642710	911/972	93,7%	P34.6	AY642710	911/972	93,7%	Freshwater B	Adriamonas peritocrescens	AF243501	908/976	93,0%	Freshwater
P34.6	AY642710	1346	PSA11SP2005	EU162646	1282/1350	95,0%	PSA11SP2005	EU162646	1282/1350	95,0%	Freshwater B	Adriamonas peritocrescens	AF243501	1266/1335	94,8%	Freshwater
CH1_5A_8	AY821966	1592	Adriamonas peritocrescens	AF243501	1480/1596	92,7%	PSA11SP2005	EU162646	1509/1717	87,9%	Freshwater B	Adriamonas peritocrescens	AF243501	1480/1596	92,7%	Freshwater
CH1_2B_3	AY821965	1627	Adriamonas peritocrescens	AF243501	1515/1636	92,6%	PSA11SP2005	EU162646	1460/1588	91,9%	Freshwater B	Adriamonas peritocrescens	AF243501	1515/1636	92,6%	Freshwater
LG33-04	AY919797	1688	Paramonas globosa	AY520452	1666/1695	98,3%	PSA11SP2005	EU162646	1554/1719	90,4%	Freshwater B	Paramonas globosa	AY520452	1666/1695	98,3%	Freshwater
LG20-12	AY919753	1700	Adriamonas peritocrescens	AF243501	1548/1719	90,1%	PSA11SP2005	EU162646	1543/1724	89,5%	Freshwater B	Adriamonas peritocrescens	AF243501	1548/1719	90,1%	Freshwater
LG21-12	AY919758	1700	Adriamonas peritocrescens	AF243501	1548/1719	90,1%	PSA11SP2005	EU162646	1542/1724	89,4%	Freshwater B	Adriamonas peritocrescens	AF243501	1548/1719	90,1%	Freshwater
PSE8SP2005	EU162645	1761	Adriamonas peritocrescens	AF243501	1648/1787	92,2%	LG30-01	AY919785	1535/1725	89,0%	Freshwater B	Adriamonas peritocrescens	AF243501	1648/1787	92,2%	Freshwater
PSA11SP2005	EU162646	1768	Adriamonas peritocrescens	AF243501	1641/1778	92,3%	LG10-05	AY919718	1554/1717	90,5%	Freshwater B	Adriamonas peritocrescens	AF243501	1641/1778	92,3%	Freshwater
M1_18B12	DQ103774	1579	OR000415.17	AY381181	1390/1601	86,8%	OR000415.17	AY381181	1390/1601	86,8%	Bicosoeca	Rhizidiomyces apophysatus	AF163296	1380/1607	85,9%	Marine
M1_18G05	DQ103786	1545	UI11D07	EU446304	1344/1539	87,3%	UI11D07	EU446304	1344/1539	87,3%	Bicosoeca	Rhizidiomyces apophysatus	AF163295	1350/1574	85,8%	Marine
M2_18B03	DQ103795	1696	Bicosoeca vacillans	AY520445	1513/1686	89,7%	PSH9SP2005	EU162647	1478/1701	86,9%	Bicosoeca	Bicosoeca vacillans	AY520445	1513/1686	89,7%	Marine

500B1m13	DQ248194	481	P6X1b-1	AY789790	447/483	92,5%	P6X1b-1	AY789790	447/483	92,5%	Bicosoeca	Bicosoeca vacillans	AY520445	442/482	91,7%	Marine
500C7	DQ248193	485	P6X1b-1	AY789790	444/477	93,1%	P6X1b-1	AY789790	444/477	93,1%	Bicosoeca	Bicosoeca vacillans	AY520445	440/476	92,4%	Marine
500A3	DQ248197	485	M2_18B03	DQ103795	482/485	99,4%	M2_18B03	DQ103795	482/485	0,9938144	Bicosoeca	Bicosoeca vacillans	AY520445	443/490	90,4%	Marine
500G2m13	DQ248195	499	M3_18D02	DQ103804	497/499	99,6%	M3_18D02	DQ103804	497/499	99,6%	Bicosoeca	Bicosoeca vacillans	AY520445	460/502	91,6%	Marine
UI11D07	EU446304	1552	Nerada mexicana	AY520453	1391/1580	88,0%	ME1-24	AF363207	1373/1575	87,2%	Bicosoeca	Nerada mexicana	AY520453	1391/1580	88,0%	Marine
Biosope_T123.046	FJ537381	759	Boroka karpovii	DQ220718	721/754	95,6%	Bicosoecida sp. SL204	EF432537	720/753	95,6%	Boroka	Boroka karpovii	DQ220718	721/754	95,6%	Marine
FV18_2D1	DQ310274	1563	Bicosoeca petiolata	AY520444	1702/1814	93,8%	SA2_2G12	EF527176	1506/1576	95,6%	Boroka	Boroka karpovii	DQ220718	1542/1568	98,3%	Marine
He001206_O15	AJ965066	547	IND58.32	EF620528	545/547	99,6%	IND58.32	EF620528	545/547	99,6%	Caecitellus	Caecitellus paraparvulus	DQ220717	545/547	99,6%	Marine
He001206_T30	AJ965067	547	IND58.32	EF620528	547/547	100,0%	IND58.32	EF620528	547/547	100,0%	Caecitellus	Caecitellus paraparvulus	DQ220717	547/547	100,0%	Marine
He000427_23	AJ965008	548	IND58.32	EF620528	538/548	98,2%	IND58.32	EF620528	538/548	98,2%	Caecitellus	Caecitellus paraparvulus	DQ220717	538/548	98,2%	Marine
CS123L38	FJ169703	760	OC4.14	EF620524	760/760	100,0%	OC4.14	EF620524	760/760	100,0%	Caecitellus	Caecitellus paraparvulus	AY520446	760/760	100,0%	Marine
Biosope_T39.110	FJ537556	791	Caecitellus parvulus	AF174368	775/791	98,0%	CS_E045	AY046666	771/791	97,5%	Caecitellus	Caecitellus parvulus	AF174368	775/791	98,0%	Marine
Biosope_T39.003	FJ537533	802	Caecitellus parvulus	AF174368	792/802	98,8%	CS_E045	AY046666	788/802	98,3%	Caecitellus	Caecitellus parvulus	AF174368	792/802	98,8%	Marine
IND58.26	EU561896	816	Caecitellus paraparvulus	DQ220717	860/860	100,0%	He001005.33	EF050072	815/816	99,9%	Caecitellus	Caecitellus paraparvulus	DQ220717	816/816	100,0%	Marine
IND60.71	EU561987	843	Caecitellus paraparvulus	DQ220717	841/843	99,8%	He001005.33	EF050072	840/843	99,6%	Caecitellus	Caecitellus paraparvulus	DQ220717	841/843	99,8%	Marine
IND33.70	EU561867	856	Caecitellus paraparvulus	DQ220717	921/922	99,9%	He001005.33	EF050072	853/856	99,6%	Caecitellus	Caecitellus paraparvulus	DQ220717	854/856	99,8%	Marine
IND60.07	EU561939	860	Caecitellus paraparvulus	DQ220717	841/843	99,8%	He001005.33	EF050072	859/860	99,9%	Caecitellus	Caecitellus paraparvulus	DQ220717	860/860	100,0%	Marine
IND70.14	EU562012	860	Caecitellus paraparvulus	DQ220717	860/860	100,0%	He001005.33	EF050072	859/860	99,9%	Caecitellus	Caecitellus paraparvulus	DQ220717	860/860	100,0%	Marine
CS123S28	FJ775663	905	OC4.14	EF620524	903/905	99,8%	OC4.14	EF620524	903/905	99,8%	Caecitellus	Caecitellus paraparvulus	AY520446	903/905	99,8%	Marine
IND60.10	EU561941	907	Caecitellus paraparvulus	DQ220717	860/860	100,0%	He001005.33	EF050072	906/907	99,9%	Caecitellus	Caecitellus paraparvulus	DQ220717	907/907	100,0%	Marine
IND58.41	EU561904	922	Caecitellus paraparvulus	DQ220717	907/907	100,0%	He001005.33	EF050072	920/922	99,8%	Caecitellus	Caecitellus paraparvulus	DQ220717	921/922	99,9%	Marine
CS123S26	FJ775662	923	OC4.14	EF620524	921/923	99,8%	OC4.14	EF620524	921/923	99,8%	Caecitellus	Caecitellus paraparvulus	AY520446	921/923	99,8%	Marine
CS123L22	FJ169698	928	OC4.14	EF620524	925/928	99,7%	OC4.14	EF620524	925/928	99,7%	Caecitellus	Caecitellus paraparvulus	AY520446	925/928	99,7%	Marine
CS123S51	FJ775664	933	OC4.14	EF620524	914/933	98,0%	OC4.14	EF620524	914/933	98,0%	Caecitellus	Caecitellus paraparvulus	AY520446	914/933	98,0%	Marine
CS_E045	AY046666	1132	Caecitellus parvulus	AF174368	1125/1132	99,4%	Biosope_T39.110	FJ537321	1045/1054	0,9914611	Caecitellus	Caecitellus parvulus	AF174368	1125/1132	99,4%	Marine
Biosope_T39.110	FJ537321	1530	Caecitellus parvulus	AF174368	1521/1532	99,3%	OC4.14	EF620524	1465/1537	95,3%	Caecitellus	Caecitellus parvulus	AF174368	1521/1532	99,3%	Marine
IND58.06	EF620527	1544	Caecitellus paraparvulus	AY520446	1540/1544	99,7%	He001005.33	EF050072	1524/1545	98,6%	Caecitellus	Caecitellus paraparvulus	AY520446	1540/1544	99,7%	Marine
OC4.7	EF620523	1695	He001005.33	EF050072	1693/1695	99,9%	He001005.33	EF050072	1693/1695	99,9%	Caecitellus	Caecitellus paraparvulus	DQ220715	1679/1680	99,9%	Marine
OC4.14	EF620524	1695	Caecitellus paraparvulus	AY520446	1692/1695	99,8%	He001005.33	EF050072	1673/1695	0,9870206	Caecitellus	Caecitellus paraparvulus	AY520446	1692/1695	99,8%	Marine
IND33.38	EF620526	1695	Caecitellus paraparvulus	AY520446	1691/1695	99,8%	He001005.33	EF050072	1672/1695	98,6%	Caecitellus	Caecitellus paraparvulus	AY520446	1691/1695	99,8%	Marine
IND58.32	EF620528	1695	He001005.33	EF050072	1691/1695	99,8%	He001005.33	EF050072	1691/1695	99,8%	Caecitellus	Caecitellus paraparvulus	DQ220715	1677/1680	99,8%	Marine
He001005.33	EF050072	1712	OC4.7	EF620523	1693/1695	99,9%	OC4.7	EF620523	1693/1695	99,9%	Caecitellus	Caecitellus parvulus	AY827847	1681/1684	99,8%	Marine
I1A1	AM041082	465	BS16_E11	FN598363	461/465	99,1%	BS16_E11	FN598363	461/465	99,1%	Cafeteria	Cafeteria roenbergensis	FJ032655	461/465	99,1%	Marine
I1D1	AM041087	465	BS16_E11	FN598363	461/466	98,9%	BS16_E11	FN598363	461/466	98,9%	Cafeteria	Cafeteria roenbergensis	FJ032655	461/466	98,9%	Marine
I2B3	AM041090	465	BS16_E11	FN598363	459/465	98,7%	BS16_E11	FN598363	459/465	98,7%	Cafeteria	Cafeteria roenbergensis	FJ032655	459/465	98,7%	Marine
D2A3	AM041080	482	BS16_E11	FN598363	474/481	98,5%	BS16_E11	FN598363	474/481	98,5%	Cafeteria	Cafeteria roenbergensis	FJ032655	474/481	98,5%	Marine
I1A3	AM041119	486	BS16_E11	FN598363	477/486	98,1%	OC4.19	EF620525	477/486	98,1%	Cafeteria	Cafeteria roenbergensis	FJ032655	477/486	98,1%	Marine
D2B6	AM041109	515	BS16_E11	FN598363	480/488	98,4%	BS16_E11	FN598363	480/488	98,4%	Cafeteria	Cafeteria roenbergensis	FJ032655	480/488	98,4%	Marine
He001005_203	AJ965039	550	Cafeteria roenbergensis	FJ032655	546/550	99,3%	OC4.19	EF620525	546/550	99,3%	Cafeteria	Cafeteria roenbergensis	FJ032655	546/550	99,3%	Marine
OC4.2	EF620522	1716	Cafeteria sp. CAFSW0510	AM493687	1709/1713	99,8%	Zeuk71	AY916583	1572/1640	95,9%	Cafeteria	Cafeteria sp. CAFSW0510	AM493687	1709/1713	99,8%	Marine
OC4.19	EF620525	1716	Cafeteria sp. CAFSW0510	AM493687	1709/1713	99,8%	Zeuk71	AY916583	1572/1640	95,9%	Cafeteria	Cafeteria sp. CAFSW0510	AM493687	1709/1713	99,8%	Marine
OC4.1	EF620521	1717	Cafeteria sp. CAFSW0510	AM493687	1710/1713	99,8%	Zeuk71	AY916583	1573/1640	95,9%	Cafeteria	Cafeteria sp. CAFSW0510	AM493687	1710/1713	99,8%	Marine

Suplementary table 3. List of studies that report the total number of clones and their phylogenetic affiliations. For each study sampling information and clone libray composition is given. In each library, clones were assigned to putative phototrophic protist groups (PP) such as prasinophytes, dinoflagellates or haptophytes, to other heterotrophic protists groups (OHP) such as ciliates or fungi, and to putative heterotrophic flagellate groups (HF). Sequences within HF were assigned to chrysophytes (Chrysos), choanoflagellates (Choanos), bicosoecids (Bicos), MAST and MALV and other heterotrophic flagellates (OHF).

Year	Author	Hábitat	Size fraction (µm)	Code	lib	Total	PP	OHP	HF	Chrysos	Choanos	Bicos	MAST	MALV	OHF	Reference
2005	Luo	Freshwater	Whole	Zeuk	3	288	55	69	164	18	0	81	0	0	65	Appl Environ Microbiol **71**:6175-6184
2005	Slapeta	Freshwater	Whole	CH1	3	90	61	20	9	0	0	3	0	0	6	Proc Roy Soc Lond B **272**: 2073-2081
2005	Slapeta	Freshwater	Whole	CV1	4	199	5	170	24	2	2	0	0	0	20	Proc Roy Soc Lond B **272**: 2073-2081
2008	Chen	Freshwater	0.8 - 20	EB /GHB/ WLB / MLB / LC / WB	6	528	199	167	162	48	3	8	6	0	97	Microb Ecol **56**: 572-583
2008	Lepere	Freshwater	0.2 - 5	B / BA	2	486	88	250	148	15	12	30	0	0	91	Appl Environ Microbiol **74**:2940-2949
2001	Díez	Marine	0.2 - 1.6	ANT	2	125	78	9	38	2	0	0	34	0	2	Appl Environ Microbiol **68**:4554-4558
2001	Díez	Marine	0.2 - 2	NA	2	37	18	2	17	1	0	0	7	7	2	Appl Environ Microbiol **68**:4554-4558
2001	Díez	Marine	0.2 - 5	ME1	1	62	30	5	27	3	0	0	12	12	0	Appl Environ Microbiol **68**:4554-4558
2001	Moon-van der Staay	Marine	0.2 - 3	OLI	1	51	11	17	23	0	2	0	7	6	8	Nature **409**:607-610
2002	Edgcomb	Marine	Whole	A / C / CS	7	269	53	98	118	7	0	1	17	49	44	Proc Natl Acad Sci USA **99**:7658-7662
2004	Massana	Marine	0.2 - 3	BL	4	340	115	24	201	1	5	0	34	139	22	Appl Environ Microbiol **70**:3528-3534
2004	Romari	Marine	0.2 - 3	RA / RD	8	367	139	64	164	0	6	1	43	60	54	Limnol Oceanogr **49**: 784-798
2006	Lovejoy	Marine	0.2 - 3	NOR / MD /NW	8	236	74	43	119	1	3	0	45	7	63	Appl Environ Microbiol **72**:3085-3095
2006	Medlin	Marine	0.2 - 3	He / Or	7	563	139	47	377	10	1	14	30	296	26	Microb Ecol **52**: 53-71
2006	Medlin	Marine	0.2 - 3	Or	1	81	12	38	31	0	0	0	14	5	12	Microb Ecol **52**: 53-71
2007	Countway	Marine	0.2 - 200	ENVP	6	923	166	322	435	0	1	0	34	311	89	Environ Microbiol **9**: 1219-1232
2007	Not	Marine	0.2 - 2	SSRP / Q2	6	225	9	46	170	3	1	0	12	144	10	Environ Microbiol **9**: 1233-1252
2008	Not	Marine	0.2 - 3	IND	8	510	113	101	296	19	0	13	60	183	21	Deep Sea Res Part I **55**: 1456-1473
2009	Not	Marine	0.6 - 3	RFM	1	62	7	24	31	2	0	0	5	24	0	PLoS ONE **4**:e7143
2009	Amacher	Marine	Whole	101 / 102 / 104 / 108 / 111 / 115	2	291	124	94	73	2	1	0	17	52	1	Deep Sea Res Part I **56**: 2206-2215
					82	5733	1496	1610	2627	130	36	151	355	1219	632	

3

Una nova mirada al voltant de la diversitat del picoplàncton eucariòtic marí

En l'última dècada els estudis independents de cultiu basats en biblioteques de clons d'ADNr 18S han revelat nombroses noves seqüències d'alt rang taxonòmic. Aquesta nova diversitat ha alterat significativament la nostra visió de les xarxes tròfiques microbianes i de l'evolució dels eucariotes. No obstant això, el panorama actual de la biodiversitat d'eucariotes marins es pot veure alterat significativament pels biaixos d'amplificació de PCR, per la presència de gens d'ADNr multicòpia en una sola cèl·lula i la capacitat de l'ADN per persistir com a material extracel·lular. En aquest estudi es va realitzar una anàlisi de les dades metagenòmiques procedents de l'expedició Global Ocean Survey (GOS), a la recerca de signatures ribosòmiques eucariòtiques. Aquest mètode independent de PCR no revela patrons filogenètics massa diferents als de les biblioteques ambientals de clons, el que suggereix que la PCR no imposa biaixos importants en la descripció de la diversitat per mitjà de les tècniques moleculars que en depenen. L'anàlisi de les diferents fraccions de cèl·lules al GOS mostra una imatge diferent en funció de la mida. L'alta diversitat en radiolaris i ciliats a la fracció de menys de 0,8 µm (i la seva absència en la fracció de 0,8 a 3 µm), suggereixen que la major part de l'ADN d'aquesta fracció prové de material extracel·lular de les cèl·lules més grans. A més, es van comparar els patrons filogenètics de biblioteques de clons de l'ADNr 18S i els obtinguts per mitjà de transcripció reversa de l'ARNr, a partir de la mateixa mostra recollida al Mar Mediterrani. Les biblioteques van revelar grans diferències, amb taxons com pelagofítes o picobilifítes que només es detecten a la biblioteca de ARNr 18S. Els MAST (Estramenòpils Marins) van aparèixer com a bacterívors potencialment importants i es va observar una disminució significativa en la contribució de les seqüències d'alveolats i radiolaris, que dominen les biblioteques d'ADNr. Les biblioteques d'ARNr semblen estar menys afectades pel nombre de còpies de cada tàxon i representen millor la diversitat de protists actius dins els cicles biogeoquímics marins que no pas les biblioteques d'ADNr.

OPEN ACCESS Freely available online

PLoS one

New Insights into the Diversity of Marine Picoeukaryotes

Fabrice Not[1,2]*, Javier del Campo[1], Vanessa Balagué[1], Colomban de Vargas[2], Ramon Massana[1]

1 Institut de Ciències del Mar, CSIC, Barcelona, Spain, **2** Station Biologique de Roscoff, UMR7144, Evolution du Plancton et PaléoOcéans (EPPO), Centre National de la Recherche Scientifique (CNRS) et Université Pierre et Marie Curie (UPMC), Place George Teissier, Roscoff, France

Abstract

Over the last decade, culture-independent surveys of marine picoeukaryotic diversity based on 18S ribosomal DNA clone libraries have unveiled numerous sequences of novel high-rank taxa. This newfound diversity has significantly altered our understanding of marine microbial food webs and the evolution of eukaryotes. However, the current picture of marine eukaryotic biodiversity may be significantly skewed by PCR amplification biases, occurrence of rDNA genes in multiple copies within a single cell, and the capacity of DNA to persist as extracellular material. In this study we performed an analysis of the metagenomic dataset from the *Global Ocean Survey* (GOS) expedition, seeking eukaryotic ribosomal signatures. This PCR-free approach revealed similar phylogenetic patterns to clone library surveys, suggesting that PCR steps do not impose major biases in the exploration of environmental DNA. The different cell size fractions within the GOS dataset, however, displayed a distinct picture. High protistan diversity in the $<$0.8 μm size fraction, in particular sequences from radiolarians and ciliates (and their absence in the 0.8–3 μm fraction), suggest that most of the DNA in this fraction comes from extracellular material from larger cells. In addition, we compared the phylogenetic patterns from rDNA and reverse transcribed rRNA 18S clone libraries from the same sample harvested in the Mediterranean Sea. The libraries revealed major differences, with taxa such as pelagophytes or picobiliphytes only detected in the 18S rRNA library. MAST (Marine Stramenopiles) appeared as potentially prominent grazers and we observed a significant decrease in the contribution of alveolate and radiolarian sequences, which overwhelmingly dominated rDNA libraries. The rRNA approach appears to be less affected by taxon-specific rDNA copy number and likely better depicts the biogeochemical significance of marine protists.

Citation: Not F, del Campo J, Balagué V, de Vargas C, Massana R (2009) New Insights into the Diversity of Marine Picoeukaryotes. PLoS ONE 4(9): e7143. doi:10.1371/journal.pone.0007143

Editor: Ryan L. Earley, University of Alabama, United States of America

Received February 25, 2009; **Accepted** August 11, 2009; **Published** September 29, 2009

Funding: This work was funded by the CANARAD project (CTM2007-28745-E, MEC) to R.M. and the ESUMAST Marie Curie fellowship (MEIF-CT-2005-025000) to F.N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: not@sb-roscoff.fr

Introduction

In the last decade, 18S rDNA clone libraries have been considered as the gold standard approach for conducting molecular surveys of marine protist diversity in the environment [1,2]. These investigations, almost exclusively performed on the picoplanktonic size fraction (0.2–3 μm), have unveiled high rank taxa such as the so-called MALV (marine alveolates, [3]), MAST (marine stramenopiles, [4]), and picobiliphytes [5], many of which have become cornerstone taxa for microbial ecologists. Diversity surveys of picoplanktonic protists in different marine regions have generated broadly similar patterns [2,6], with dominance of non-photosynthetic groups, including tiny parasites [7] and grazers [8]. In contrast, epifluorescence microscopy typically reveals a dominance of photosynthetic or mixotrophic cells over heterotrophic cells (ca 80% vs 20%, respectively) in the oceans [9]. This suggests that 18S rDNA clone libraries may give a significantly biased view of diversity. Several technical limitations inherent to culture-independent explorations of microbial diversity have been highlighted [10,11]. Among these, biases during DNA extraction and PCR amplification steps [12], primer selectivity, multiple rDNA gene copy number [13], and the existence of pseudogenes [14] or extracellular DNA [15], are particularly relevant.

Alternative approaches focused on photosynthetic protists have recently been developed to overcome the apparent bias towards heterotrophic cells. These include the construction of clone libraries from flow cytometry sorted populations [16], studies specifically targeting plastid genes [17], and the use of taxon-specific primers [18]. However, PCR biases, rDNA copy number, and extracellular DNA remain as potentially problematic issues with these approaches. A promising alternative which does not require PCR steps is the metagenomic approach, based on direct cloning and shotgun sequencing of environmental DNA.

This strategy was recently used to study prokaryotic life on a worldwide scale (Sorcerer, Global Ocean Survey expedition, [19]). Studies that compared metagenomic and 16S rDNA PCR-based clone libraries demonstrated that these two approaches were complementary for bacterial community analysis [20,21]. With respect to eukaryotic microbes, phylogenetic information present in metagenomic libraries has thus far received very little attention [22]. Another perspective to investigate microbial diversity is to target directly the 18S rRNA (i.e. the ribosomes themselves) as a proxy for both diversity and metabolic activity of cells [23], and to avoid the problems induced by differences in rDNA copy number and the perturbation from dissolved DNA. This approach has been proven to be effective on prokaryotic communities [11,24,25], but to date has only been applied on protist communities in an oxygen depleted environment [26].

In the present study we performed an in-depth analysis of the metagenomic dataset from the GOS expedition, seeking eukary-

otic signatures through the presence of 18S rDNA genes. We also compared the protist diversity assessed by 18S rDNA libraries prepared from both environmental DNA and RNA extracted from the same water sample collected in the Mediterranean Sea. We show that overall the PCR induced biases do not appear to impact significantly diversity surveys. Rather we argue that rDNA copy number and extracellular DNA (partially by-product of the size fractionation) are major issues that introduce biases in current studies of protist diversity. Environmental 18S rRNA clone libraries appear to represent a promising means to minimize these biases and thereby offer new perspectives in the study of the diversity and function of marine protist.

Results

Taxonomic composition in 18S rDNA clone libraries versus the metagenomic dataset

Taxonomic affiliation of sequences retrieved from PCR amplified 18S rDNA clones libraries performed on the picoplankton size fraction (0.2 to 3 μm) of samples collected in the photic zone around the globe [2] was compared to that of 18S rDNA sequences found in the <3 μm size fraction of the GOS metagenomic dataset (Figure 1A). Despite the large differences in the number of sequences analyzed for both datasets, random sub-sampling of the larger dataset demonstrated that the range of expected averaged distributions on a smaller number of sequences matched closely to the distribution observed (Figure S1). This shows that looking at a limited number of sequences does not affect the diversity observed at the taxonomic level we considered. The clonal representation of the different taxonomic groups in both datasets was significantly correlated (slope 0.78; $R^2 = 0.39$; $p = 0.0165$), indicating that both integrated datasets yielded comparable results.

In the clone libraries, out of the 2175 sequences reviewed by Massana and Pedrós-Alió, alveolates dominated the assemblages with 50.3% of the sequences retrieved (most of which were MALV, marine alveolates: 19.2% MALV-II and 16.7% MALV-I). The second most represented taxon was the stramenopiles, accounting for 20% of the eukaryotic sequences (of which 10.9% were MAST, marine stramenopiles). Prasinophytes and radiolarians accounted for 12.1% and 4.1% of the sequences, respectively. Cryptophytes, haptophytes and picobiliphytes represented 2.9%, 2.4%, and 1.1% of the sequences. The category "other", accounting for 7.2% of the sequences, was mainly composed of cercozoans, choanoflagellates and unassigned alveolates. Out of 116 sequences extracted from the GOS metagenomic dataset, the most represented groups were the alveolates (40.5%, of which 24.1% belonged to MALV-II and 5.2% to MALV-I), radiolarians (18.1%), stramenopiles (16.4%, including 12.9% of MAST), prasinophytes (11.2%), and haptophytes (6.9%). Sequences

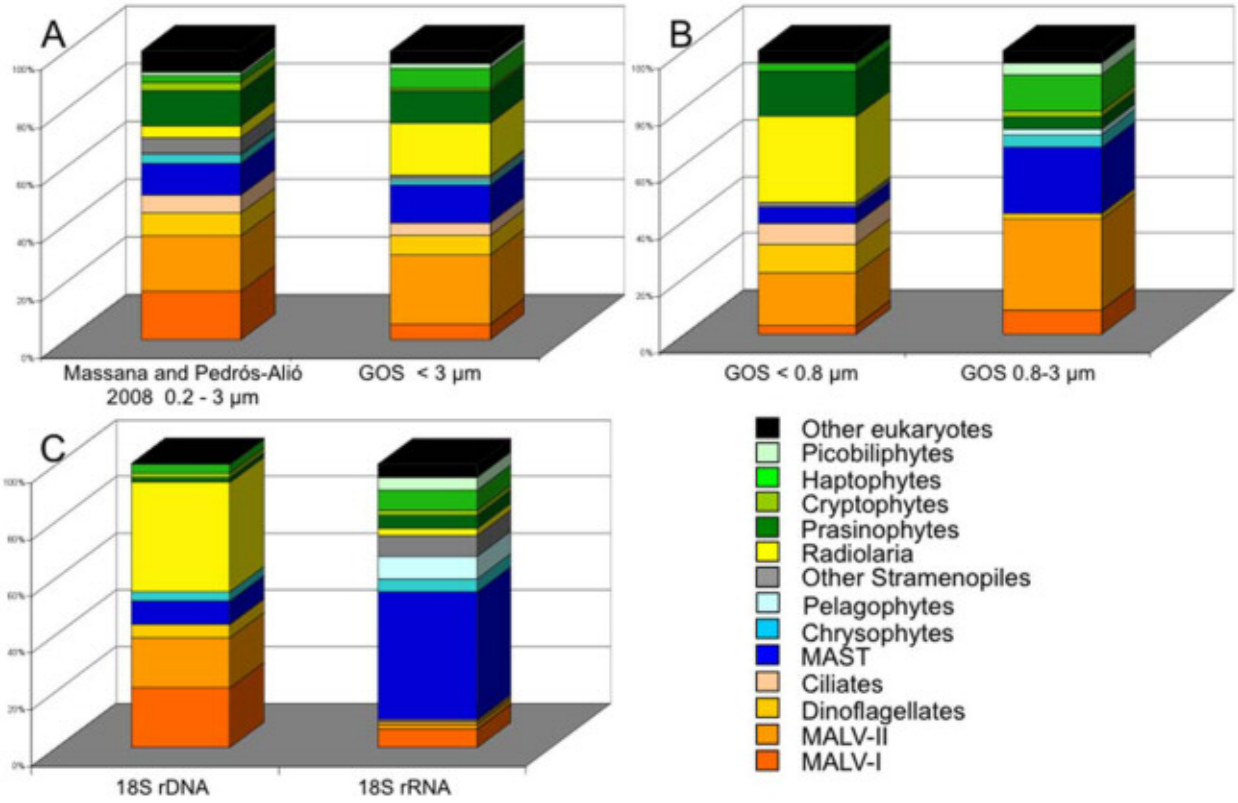


Figure 1. Relative contribution of different taxonomic groups from 18S rDNA sequences obtained from the picoplankton fraction of marine samples. A. Comparison of data obtained through PCR-based clone libraries as presented in [2] *versus* the metagenomic data retrieved from < 3 μm size fraction of the GOS dataset. **B.** Detail of the metagenomic GOS dataset obtained from two different size fractions <0.8 μm and 0.8 to 3 μm. **C.** Comparison of clone libraries performed on the same sample from the Mediterranean Sea (0.6 to 3 μm size fraction) after DNA extraction (62 sequences) and RNA extraction (111 sequences). Actual numbers of sequences affiliated to each taxonomic group used to prepare these graphs are shown in Table S1.
doi:10.1371/journal.pone.0007143.g001

affiliated to the picobiliphytes accounted for 1.7% of total sequences.

The 18S rDNA sequences retrieved from the GOS dataset had an averaged similarity of 96.0% with sequences deposited in the GenBank database. The most dissimilar sequences affiliated with marine alveolates (e.g. 80.6% similarity), whereas some were identical to GenBank entries (mostly prasinophytes and the haptophyte *Chrysochromulina*) and many were >99% similar to GenBank sequences. Closest matches for most GOS sequences corresponded to environmental clones obtained from PCR-based studies (Tables S4 and S5)

Taxonomic analysis of distinct size fractions among the metagenomic dataset

Of the 116 18S rDNA sequences identified in the metagenomic dataset from the GOS expedition, 69 derived from the <0.8 μm fraction and 47 from the 0.8–3 μm size fraction. Clearly, both size fractions were capturing a distinct fraction of picoeukaryotic diversity (Figure 1B), and the percentage of taxonomic groups observed in the two size fractions did not correlate at all (slope 0.18; $R^2 = 0.03$; $p = 0.5523$). Considering the smaller size fraction (<0.8 μm), radiolarians contributed 30.4% and stramenopiles 7.2% of the sequences (with 5.8% MAST). The overall contribution of alveolates was 41.9% of the sequences, including 18.8% of MALV-II, 10.1% of dinoflagellates and 7.2% of ciliates. Prasinophytes contributed 15.9% and haptophytes 2.9%. No picobiliphyte sequences were detected. In the larger size fraction (0.8–3 μm) the overall contribution of alveolates remained similar, but there was an increase of MALV-II (31.9% of sequences) and a decrease of dinoflagellates (2.1%) and ciliates (not detected). The contribution of stramenopiles increased drastically to 29.8% (of which 23.4% were MAST) while not a single radiolarian sequence was identified. Prasinophytes decreased to 4.3%, whereas the contributions of haptophytes, chrysophytes, and picobiliphytes increased to 12.8%, 4.3%, and 4.3%, respectively.

18S rDNA clones libraries prepared from DNA and RNA extracts

18S rDNA environmental clone libraries were constructed from DNA and RNA extracts (followed by a reverse transcription) obtained from the same seawater sample (Figure 1C). Considering the limited number of clones sequenced and previous knowledge for marine samples, the libraries were explored in numbers far from saturation. Nevertheless, obvious patterns could be distinguished and the distribution of diversity observed for the 18S rRNA library is well outside the range of expected values for 18S rDNA libraries. Again, there was no correlation among the clonal percentage of taxonomic groups in the two libraries (slope -0.02; $R^2 = 0.00$; $p = 0.9539$). Among the 62 sequences from the DNA based library, 43.5% affiliated to alveolates, 38.7% to radiolarians, and 11.3% to stramenopiles. Most alveolate sequences affiliated with MALV-I (21.0%) or MALV-II (17.7%). Most of the stramenopiles belonged to MAST (i.e. 8.1% of the sequences). Chrysophytes, haptophytes, prasinophytes, and cryptophytes were detected but with a low clonal representation. In the rRNA based library, the diversity observed for the 111 sequences analyzed was drastically different. The contribution of alveolates decreased to 9.9% and the contribution of stramenopiles increased to 64.8% including 45.0% MAST. The contribution of sequences affiliated to haptophytes and prasinophytes increased to 7.2% and 4.5%, respectively. In contrast, the contribution of radiolarians sharply decreased down to 2.7%. The pelagophytes and picobiliphytes, which were not detected in the DNA survey, contributed 8.1% and 4.5% of sequences in the RNA survey, respectively. Also only detected in the RNA-based library, dictyochophytes made up half of the "other stramenopiles" category and *Telonemia* the major fraction of the "other eukaryotes" (data not shown).

In each library, Operational Taxonomic Units (OTUs) were defined using a 99% identity threshold (Table 1). Of the 62 and 111 sequences from the DNA and RNA based libraries, 34 and 52 OTUs were identified, respectively. Only 2 OTUs were present in both libraries, one affiliated to MALV-I, and the other to MAST-4. Using a 98% identity threshold, 29 and 46 OTUs were identified for the DNA and RNA based libraries respectively, but only one additional OTU (belonging to chrysophytes) was common to the two libraries. Statistical comparisons performed with LIBSHUFF found a significant difference between the two libraries ($p<0.001$).

Discussion

18S rDNA clone libraries and metagenomic surveys give similar diversity patterns

Our analyses of the 18S rDNA sequences retrieved from the metagenomic dataset from the GOS expedition did not reveal substantial differences as compared to the PCR-based environmental clone libraries (Figure 1A). Both datasets were obtained from a similar size fraction (<3 μm) and correspond to compilations of sequences from various sampling locations and thus represent a reasonable integration of the photic layer in the marine environment. Eukaryotic microbial diversity assessed by means of environmental clone libraries of the 18S rDNA gene has been reported from a variety of ecosystems over the last decade [2,6]. This approach has led to the discovery of eukaryotic taxa such as the MALV and MAST groups that often dominate the

Table 1. Number of sequences and OTUs (Operational Taxonomic Units) defined at 99% identity threshold in different taxonomic groups from both DNA- and RNA-based libraries.

	DNA		RNA	
	# seq.	OTU 99%	# seq.	OTU 99%
MALV-I	13	8	7	2
MALV-II	11	8	2	2
Dinoflagellates	3	2	1	1
Ciliates	0	0	1	1
MAST	5	3	50	20
Chrysophytes	2	1	5	2
Pelagophytes	0	0	9	1
Other Stramenopiles	0	0	8	6
Radiolarians	24	8	3	2
Prasinophytes	1	1	5	3
Cryptophytes	1	1	2	1
Haptophytes	2	2	8	6
Picobiliphytes	0	0	5	1
Telonema	0	0	3	2
Other	0	0	2	2
TOTAL	**62**	**34**	**111**	**52**
Ratio OTUs / # seq.	**0.55**		**0.47**	

doi:10.1371/journal.pone.0007143.t001

community in terms of clonal abundance. Among the technical issues usually invoked to lead to biases in 18S rDNA clone libraries there is the PCR step before the cloning procedure [10,12]. Metagenomic approaches directly clone and shotgun sequence the DNA from a given sample, without prior PCR. The similarity in diversity patterns between the PCR cloning and metagenomic approaches suggests little impact of the PCR step on the outcome of clone libraries in terms of sequence diversity and relative contribution of specific taxa. Our observation is consistent with similar studies on 16S rDNA bacterial diversity that did not find significant differences at high phylogenetic levels between metagenomic and PCR-based libraries [20].

Analysis of GOS size fractions refines our view of actual community composition

Separate analysis of the two size fractions from the GOS dataset revealed clear differences in terms of taxonomic composition (Figure 1B). As the smallest eukaryotic organism known so far has a cell diameter of 0.8 μm [27], some of the 18S rDNA signatures observed in the <0.8 μm fraction might indeed derive from very small eukaryotes (like the prasinophytes that appeared mostly in this small fraction, Table S4), but many sequences most likely derive from cell debris or extracellular DNA from larger cells. This is likely the case for radiolarians, dinoflagellates, and ciliates, groups known to contain relatively large nano- and microplanktonic cells, and for which sequences were prominent in the <0.8 μm fraction and nearly absent from the 0.8–3 μm fraction. Among these groups, the radiolarians were the most intriguing, since these relatively large exoskeleton bearing protists typically represent a significant fraction of 18S rDNA sequences in diversity surveys of marine picoeukaryotes (Figure 1A). These radiolarian sequences appear highly diverse [28], and most likely derived from larger organisms for which molecular data are not yet available, highlighting the extent of both the unknown diversity in this taxonomic group and filtration artifacts which affect environmental surveys of the smallest size fractions. As suggested in a recent study that investigated the eukaryotic diversity of the <0.8 μm size fraction in a subset of the GOS dataset (i.e. Sargasso Sea samples) [22], future environmental surveys should target the 0.8–3 μm fraction, which may actually better represent the picoeukaryote diversity.

Several studies have proved the occurrence of extracellular DNA (particulate or dissolved) in water or sediments [29–31]. Among this DNA pool, a substantial portion contains high molecular weight molecules and is thought to be derived from eukaryotic organisms [29]. This extracellular DNA is prone to PCR amplification, and genes such as the one coding for the rbcL enzyme have been successfully amplified from particle-free water samples [15]. It is very likely that a fraction of the extracellular DNA is retained onto 0.2 μm filters, through collection of aggregates or molecular adsorption. Consequently, we believe that it is important to consider the interference of extracellular DNA when assessing the diversity of eukaryotic microbes in ecological perspectives.

The RNA approach gives complementary perspectives on marine protist diversity

Diversity assessed by means of libraries prepared after reverse transcription of extracted RNA led to a drastically different view of the community as compared to the classical DNA-based approach (Figure 1C). It is generally recognised that 18S rDNA diversity surveys are not quantitative with respect to cell abundance [32,33]. Besides PCR biases, the contributions of specific taxa are related to the number of rDNA copies within cells of the taxa. Although rDNA copy number is usually assumed to be correlated with cell size [13,34], for a limited size range (e.g. picoeukaryotes) this number can vary significantly depending on phylogenetic affiliation and is also suspected to be influenced by life strategies of cells (e.g. parasitic, heterotrophic, autotrophic) [2,35]. The effect of taxon-specific rDNA copy number is avoided when analysing extracted RNA. Moreover, extracellular RNA is much less stable than DNA, minimising the problem of amplification from extracellular material. Ribosome content within a single cell is commonly viewed as a proxy of cellular activity status [23,36]. Therefore, 18S rRNA libraries are intentionally skewed to give insights on both diversity and taxon specific activity within protist assemblages [26]. As a flip side effect we might have expected an over representation of the most active taxa. However, both DNA-based and RNA-based libraries contained a high diversity, with comparable ratios of OTUs/number of sequences (Table 1). We found very little overlap in the sequences retrieved in the DNA and RNA libraries. At the 98% identity threshold, only 3 OTUs (ca. 4%) were detected in both libraries, which is rather low compared to the 27% observed in a similar study on anoxic waters [26]. This discrepancy might be explained by a lower sequencing effort done here but also by the selective nature of anoxic waters that might impose stronger constraints on the communities compared with open ocean conditions, implying a lower diversity and therefore a higher overlap between rDNA and rRNA libraries.

The diversity observed by both approaches is clearly not distributed within the same high level taxa, paralleling observations made on prokaryotes or on eukaryotes in an extreme environment [11,24–26]. Some photosynthetic groups such as pelagophytes and picobiliphytes were not detected in the 18S rDNA based library, whereas they contributed notably to the 18S rRNA library (Figure 1C). The relative contribution of other photosynthetic groups such as the prasinophytes and the haptophytes was also higher in the rRNA library. This might reflect a relatively higher metabolic activity in these photosynthetic taxa at the time of sampling, or may indicate that they have fewer rDNA copies (e.g. *Pelagomonas*, [13]), so they could be diluted in the environmental DNA surveys by cells with a higher rDNA copy number (e.g. alveolates). Among prasinophytes, cells belonging to the genus *Micromonas* were identified as being the most active (Table S3), confirming previous studies showing the significance of this genus in coastal ecosystems [37]. Regarding heterotrophic protists, sequences belonging to MAST-3, -4 and -7 appeared as prominent grazers (Table S3), which together with the widespread distribution of these taxa suggest they might actually be the major protistan predators in the oceans [8]. Finally, the most pronounced divergence between both libraries was the contribution of alveolates and radiolarians, which overwhelmingly dominated DNA-based diversity surveys [2]. This perhaps reflects the high 18S rDNA gene diversity and high copy number matching the parasitic life strategy of MALV [7,38] and further supports the putative presence of extracellular radiolarian 18S rDNA in seawater.

Conclusions

Size fractionation, metagenomics, and 18S rRNA libraries bring new perspectives for the understanding of marine picoeukaryotic diversity. In particular, rRNA libraries reduce significantly two of the major biases of rDNA diversity surveys, the rDNA copy number and the occurrence of extracellular DNA, but are in turn skewed towards the active part of the communities. Considering the relative ease of handling ribosomal RNA molecules, extended

diversity surveys based on environmental rRNA will undoubtedly provide insights into the ecology of uncultured species. Associated with stronger depth of sequencing (e.g. 454 [39]), this approach will probably help to achieve a nearly exhaustive view of protist diversity and to better appreciate the contribution and function of specific organisms in the microbial food web.

Materials and Methods

Mining the GOS dataset using CAMERA

The Global Ocean Survey (GOS) covered a variety of oceanic regions from Nova Scotia to South Africa across the Caribbean, the Panama Channel, the Pacific and the Indian Ocean [19] and data is accessible through the CAMERA database [40]. For the purpose of our analysis, and to compare waters of similar characteristics, only samples from offshore and coastal photic zones were used, whereas samples from environments such as hypersaline lagoons or mangroves were discarded. Seventy two sampling sites, representing a sequencing effort of 14000 Mb, were analyzed for the <0.8 µm fraction, whereas only 8 sampling sites (850 Mb) were analyzed for the 0.8–3 µm fraction. This demonstrates the primary focus on prokaryotes of the GOS expedition. The fraction <3 µm recorded in our analysis corresponds to the sum of data retrieved from the two size fractions. We searched for 18S rDNA genes using the eukaryotic specific primers EukA and EukB [41], 528f [42], 336f and 1209f [43] as *in silico* probes. Sequences were then assigned to specific taxonomic groups after the results of BLAST searches [44]. Chimeras were detected by doing BLAST with different regions of the sequence. Metazoans, marine euryarchaeote group II sequences (obtained with EukA primer), and short (<100 bp) sequences were discarded. We ended up with a total of 116 eukaryotic sequences from this metagenomic survey, with 69 and 47 sequences in <0.8 µm and 0.8–3 µm size fractions, respectively.

Sampling procedures for the DNA vs RNA clone libraries

Seawater samples were harvested on November, 15th 2007 in the Mediterranean Sea off Villefranche sur Mer (France). Water was collected with a 12L Niskin bottle deployed successively at 40, 60, 80, 100, 120, and 140 meter depths. After a pre-filtration through a 1000 µm mesh, equal volumes of water from each depth were mixed together in order to obtain an integration of the communities throughout the water column. Then water was gently sieved through 63 µm and 20 µm meshes and filtered through a 3 µm pore size 47 mm diameter polycarbonate filter. For DNA and RNA libraries, around 4 liters of the fraction below 3 µm were filtered onto 0.6 µm pore size 47 mm diameter polycarbonate filters at a rate of 90 ml min^{-1}. Finally the filters were flash frozen in liquid nitrogen and stored at −80°C until further analysis.

Nucleic acid extractions and clone library construction

For DNA extraction, the filter was cut in small pieces with a sterile razor blade and placed in a 2 ml microcentrifuge tube. Liquid nitrogen was added to the tube and the frozen sample material was disrupted manually with a disposable pellet pestle (Fisher Bioblock), repeating this step four more times. This disruption procedure was followed by DNA extraction with a DNeasy Plant Mini kit (Qiagen) following the manufacturer's recommendations. DNA extracts were stored at −80°C until analysis. For RNA extraction, filters were immersed in RLT buffer (from a Quiagen RNeasy kit) mixed with an equal amount of 0.1 and 0.5 µm glass beads and subsequently vortexed. Then the RNeasy kit instructions for Plants and Fungi were followed. Quantification of extracted nucleic acids was performed with the Qubit Quantitation platform (Invitrogen). Prior to reverse transcription, a DNase digestion step was performed with DNaseI (Roche Diagnostic) and efficient digestion was controlled by gel electrophoresis. Reverse transcription was performed on pure RNA using the SuperScript II kit (Invitrogen) according to the manufacturer's instructions. The eukaryotic 18S specific EUKB primer [41] was used for the reverse transcription.

Both 18S rDNA genes and 18S ribosomal cDNA were PCR amplified using the same set of primers, 528f [42] and EUKB [41]. Approximately 10 ng of DNA were used as a template in a 50 µl PCR mixture containing 200 µM of each dNTP, 1.5 mM $MgCl_2$, 0.5 µM of each primer and 1.25 units of *Taq* DNA polymerase (Promega) with the PCR buffer supplied with the enzyme. Reactions were carried out in a thermocycler with the following cycle: an initial denaturing step at 94°C for 3 min, 35 cycles of denaturing at 94°C for 45 s, annealing at 55°C for 1 min and extension at 72°C for 3 min, and a final extension step at 72°C for 10 min. In order to check the quality of the RNA extraction, we used the RNA extract digested by DNase as a PCR template. Negative results confirmed the lack of remnant DNA after digestion which could have interfered with the results obtained for the cDNA libraries. PCR products were used for clone library construction. In both cases, three separate fresh PCR products (50 µl) were pooled and cleaned with the Qiagen PCR Purification kit and cloned using the TOPO-TA cloning kit (Invitrogen). Putative positive clones were checked by PCR amplification using the same primer set. PCR reactions showing the right insert size were purified and sequenced with the 528f primer on an ABI Prism 3100 sequencer (Applied Biosystems) at the Station Biologique de Roscoff sequencing facility.

Taxonomic affiliation of the 18S rDNA sequences obtained in this study (between 800 and 950 bp length) and putative chimeras were identified by using BLAST as explained before (data shown in Tables S2 and S3). Among the 113 cDNA clones sequenced 2 were chimeras leaving 111 sequences for further analysis. Sixty seven rDNA clones were sequenced, 2 chimeras were identified, and 3 metazoan sequences (Appendicularia and copepods) were discarded, leaving 62 sequences for further analysis. Operational Taxonomic Units (OTU) at 99% identity threshold were identified and compared among libraries using the DOTUR and SONS programs [45,46]. Statistical comparisons of the two libraries were performed with the webLIBSHUFF tool [47]. Sequences have been deposited in GenBank under accession numbers GQ344621 to GQ344796.

Statistical analysis

Considering the small number of sequences retrieved from our analysis, we wanted to make sure that comparisons between datasets were meaningful. Using R software we calculated the expected distribution of sequences from small size samples compared to a larger reference dataset. The random sub-sampling procedure of 62 and 47 sequences was replicated 1000 times from the Massana and Pedrós-Alió (2008) dataset (2175 sequences) and the GOS dataset (116 sequences), respectively. Standard deviations were calculated for each taxonomic group considered and comparisons between observed and expected datasets were plotted (Figure S1).

Correlations were performed with the statistical package JMP 5.0.1a to evaluate the degree of divergence between paired datasets and estimate the impact of PCR approaches (Figure 1A), size fractionation (Figure 1B), and 18S rDNA versus 18S rRNA clones libraries (Figure 1C), on environmental diversity surveys.

Supporting Information

Figure S1 Taxonomic distribution of observed diversity compared to expected distribution in a sample of smaller size. A) Histogram showing the observed distribution of sequences in the Massana and Pedrós-Alió 2008 dataset (Black) and the average and standard deviation of expected distribution after random sub-sampling of 62 sequences, replicated 1000 times (Red). B) Histogram showing the observed distribution of sequences in the GOS < 3µm dataset (Black) and the average and standard deviation of expected distribution after random sub-sampling of 47 sequences, replicated 1000 times (Red).
Found at: doi:10.1371/journal.pone.0007143.s001 (3.02 MB TIF)

Table S1 Number of sequences for each taxonomic group found in the analyzed dataset
Found at: doi:10.1371/journal.pone.0007143.s002 (0.05 MB DOC)

Table S2 List of closest blast results for the RNA based clone library
Found at: doi:10.1371/journal.pone.0007143.s003 (0.26 MB DOC)

Table S3 List of closest blast results for the DNA based clone library
Found at: doi:10.1371/journal.pone.0007143.s004 (0.16 MB DOC)

Table S4 Closest blast hits on sequences retrieved from the GOS < 0.8µm dataset
Found at: doi:10.1371/journal.pone.0007143.s005 (0.10 MB DOC)

Table S5 Closest blast hits on sequences retrieved from the GOS 0.8 - 3 µm dataset
Found at: doi:10.1371/journal.pone.0007143.s006 (0.08 MB DOC)

Acknowledgments

We are indebted to Dr. John Dolan for having provided us lab space and sampling facilities in Villefranche sur Mer. We thank Sarah Romac for sequencing assistance, Yoshiyuki Ishitani for help with DNA extractions, and Marta Ribes, Gildas le Corguillé and Océane Dahan for support with statistical analysis. We are also grateful to Daniel Vaulot and Ian Probert for comments on the manuscript and English improvements.

Author Contributions

Conceived and designed the experiments: FN JdC CdV RM. Performed the experiments: FN JdC VB. Analyzed the data: FN JdC CdV RM. Wrote the paper: FN JdC VB CdV RM.

References

1. Epstein S, Lopez-Garcia P (2008) "Missing" protists: a molecular prospective. Biodiversity and Conservation 17: 261–276.
2. Massana R, Pedros Alió C (2008) Unveiling new microbial eukaryotes in the surface ocean. Current Opinion in Microbiology 11: 213–218.
3. López García P, Rodriguez Valera F, Pedros Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409: 603–607.
4. Massana R, Castresana J, Balague V, Guillou L, Romari K, et al. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. Applied and Environmental Microbiology 70: 3528–3534.
5. Not F, Valentin K, Romari K, Lovejoy C, Massana R, et al. (2007) Picobiliphytes: A marine picoplanktonic algal group with unknown affinities to other eukaryotes Science 315: 252–254.
6. Worden AZ, Not F (2008) Ecology and diversity of picoeukaryotes. In: Kirchman DL, ed. Microbial Ecology of the Ocean. 2nd edition ed. New York: Wiley-Liss. pp 159–196.
7. Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, et al. (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (Alveolata). Environmental Microbiology 10: 3349–3365.
8. Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. Environmental Microbiology 8: 1515–1522.
9. Jürgens K, Massana R (2008) Protist grazing on marine bacterioplankton. In: Kirchman DL, ed. Microbial Ecology of the Oceans. 2nd edition ed. New York: Wiley-Liss. pp 383–424.
10. Wintzingerode Fv, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. Fems Microbiology Review 21: 213–229.
11. Moeseneder MM, Arrieta JM, Herndl GJ (2005) A comparison of DNA- and RNA-based clone libraries from the same marine bacterioplankton community. Fems Microbiology Ecology 51: 341–352.
12. Suzuki MT, Giovannoni SJ (1996) Bias Caused by Template reannealing in the Amplification of Mixtures of 16S rRNA Genes by PCR. Applied and Environmental Microbiology 62: 625–630.
13. 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 Microbial Ecology 52: 79–92.
14. Thornhill DJ, LaJeunesse TC, Santos SR (2007) Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. Molecular Ecology 16: 5326–5340.
15. Paul JH, Cazares L, Thurmond J (1990) Amplification of the rbcL Gene from Dissolved and Particulate DNA from Aquatic Environments. Applied and Environmental Microbiology 56: 1963–1966.
16. Shi X, Marie D, Vaulot D (submitted) Novel photosynthetic lineages uncovered in the South East Pacific Ocean from flow cytometry sorted picoeukaryote populations..
17. Fuller NJ, Campbell C, Allen DJ, Pitt FD, Zwirglmaier K, et al. (2006) Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using a PCR primer biased towards marine algal plastids. Aquatic Microbial Ecology 43: 79–93.
18. Liu H, Probert I, Uitz J, Claustre H, Aris-Brossou S, et al. (2009) Extreme diversity in noncalcifying haptophytesexplains a major pigment paradox in open oceans. Proceedings of the National Academy of Sciences in press.
19. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. PloS Biology 5: e77.
20. Cottrell MT, Waidner LA, Yu L, Kirchman DL (2005) Bacterial diversity of metagenomic and PCR libraries from the Delaware River. Environmental Microbiology 7: 1883–1895.
21. Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Applied and Environmental Microbiology 69: 2684–2691.
22. Piganeau G, Desdevises Y, Derelle E, Moreau H (2008) Picoeukaryotic sequences in the Sargasso Sea metagenome. Genome Biology 9: R5.
23. Poulsen LK, Ballard G, Stahl DA (1993) Use of rRNA Fluorescence In Situ Hybridization for Measuring the Activity of Single Cells in Young and Established Biofilms. Applied and Environmental Microbiology 59: 1354–1360.
24. Gentille G, Giuliano L, D'Auria G, Smedile F, Azzaro M, et al. (2006) Study of bacterial communities in Antarctic coastal waters by a combination of 16S rRNA and 16S rDNA sequencing. Environmental Microbiology 8: 2150–2161.
25. Mills HJ, Martinez RJ, Story S, Sobecky PA (2005) Characterization of Microbial Community Structure in Gulf of Mexico Gas Hydrates: Comparative Analysis of DNA- and RNA-Derived Clone Libraries. Applied and Environmental Microbiology 71: 3235–3247.
26. Stoeck T, Zuendorf A, Breiner H-W, Behnke A (2007) A molecular approach to identify active microbes in environmental eukaryote clone libraries. Microbial Ecology 53: 328–339.
27. Courties C, Vaquer A, Trousselier M, Lautier J, Chrétiennot-Dinet M-J, et al. (1994) Smallest eukaryotic organism. Nature 370: 255.
28. Not F, Gausling R, Azam F, Heidelberg JF, Worden AZ (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. Environmental Microbiology 9: 1233–1252.
29. Jiang SC, Paul JH (1995) Viral Contribution to Dissolved DNA in the Marine Environment as Determined by Differential Centrifugation and Kingdom Probing. Applied and Environmental Microbiology 61: 317–325.
30. Dell'Anno A, Danovaro R (2005) Extracellular DNA Plays a Key Role in Deep-Sea Ecosystem Functioning. Science 309: 2179.
31. Vlassov VV, Laktionov PP, Rykova EY (2007) Extracellular nucleic acids. BioEssays 29: 654–667.
32. Kirchman DL (2002) The ecology of Cytophaga-Flavobacteria in aquatic environments. Fems Microbiology Ecology 39: 91–100.
33. Not F, Latasa M, Scharek R, Viprey M, Karleskind P, et al. (2008) Phytoplankton diversity across the Indian Ocean: A focus on the picoplanktonic size fraction. Deep - Sea Research Part I - Oceanographic Research Papers 55: 1456–1473.
34. Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, et al. (2008) Quantifying diatom and dinoflagellate biomass in coastal marine seawater samples by real-time PCR. Applied and Environmental Microbiology.

35. Vaulot D, Romari K, Not F (2002) Are autotrophs less diverse than heterotrophs in marine picoplankton? Trends in Microbiology 10: 266–267.
36. Buckley BA, Szmant AM (2004) RNA/DNA ratios as indicators of metabolic activity in four species of Caribbean reef-building corals. Marine Ecology - Progress Series 282: 143–149.
37. Not F, Latasa M, Marie D, Cariou T, Vaulot D, et al. (2004) A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. Applied and Environmental Microbiology 70: 4064–4072.
38. Chambouvet A, Morin P, Marie D, Guillou L (2008) Control of toxic marine dinoflagellate blooms by derial parasitic killers. Science 322: 1254–1257.
39. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.
40. Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: A Community Resource for Metagenomics. PloS Biology 5: e75.
41. Medlin LK, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like r RNA-coding regions. Gene 71: 491–499.
42. Elwood HJ, Olsen GJ, Sogin ML (1985) The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. Molecular Biology and Evolution 2: 399–410.
43. Giovannoni SJ, DeLong EF, Olsen GJ, Pace NR (1988) Phylogenetic group-specific oligonucleotide probes for identification of single microbial cells. Journal of Bacteriology 170: 720–726.
44. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, David J. Lipman (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25: 3389–3402.
45. Schloss PD, Handelsman J (2005) Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. Applied and Environmental Microbiology 71: 1501–1506.
46. Schloss PD, Handelsman J (2006) Introducing SONS, a Tool for Operational Taxonomic Unit-Based Comparisons of Microbial Community Memberships and Structures. Applied and Environmental Microbiology 72: 6773–6779.
47. Singleton DR, Furlong MA, Rathbun SL, Whitman WB (2001) Quantitative Comparisons of 16S rRNA Gene Sequence Libraries from Environmental Samples. Applied and Environmental Microbiology 67: 4374–4376.

4

Enfrontant-se al biaix de cultiu en flagel·lats heteròtrofs marins mitjançant incubacions d'aigua de mar enriquida

Els flagel·lats heterotròfics exerceixen un paper important en els sistemes aquàtics. La diversitat d'aquest conjunt heterogeni de protists es basa generalment en soques cultivades, en les quals s'han realitzat els corresponents estudis ultraestructurals, fisiològics i moleculars. Cal destacar que els estudis per mitjà de tècniques moleculars de mostres ambientals han revelat nous llinatges eucariotes que fins ara no han pogut ser cultivats, amagant possiblement característiques evolutives i ecològiques fonamentals. El biaix de cultiu apareix sovint com un dubte raonable darrere dels estudis ecològics, però poques vegades ha estat degudament tractat. Aquest estudi va ser dissenyat per tal de desentrellar aquest fenomen, analitzant l'efecte de la matèria orgànica sobre una comunitat natural de flagel·lats heterotròfics. Vam establir diversos microcosmos amb una concentració creixent de matèria orgànica d'origen divers. La dinàmica de creixement va ser seguida per mitjà de microscòpia d'epifluorescència i la diversitat analitzada per DGGE i biblioteques de clons, mostrant una clara substitució de la comunitat, que diferia cada vegada més de la mostra inicial a mesura que la matèria orgànica anava augmentant. Dins d'aquest gradient també hi va haver un augment de les seqüències relacionades amb organismes cultivats, principalment crisofícies, així com una disminució en els índexs de diversitat. Vam arribar a la conclusió que el biaix de cultiu és el resultat de la utilització de matèria orgànica en el procés d'aïllament, que impulsa un canvi en la comunitat cap a condicions més a prop dels clàssic cultius de laboratori. Només l'ús de mètodes d'aïllament alternatius permetrà l'accés als protists que constitueixen els tàxons més abundants a l'ecosistema i que tenen un paper actiu en el flux de carboni i energia.

Facing Culturing Bias in Marine Heterotrophic Flagellates Through Seawater Enrichment Incubations

Javier del Campo[a], Vanessa Balagué[a], Irene Forn[a], Itziar Lekunberri[a, b], Ramon Massana[a]

[a]Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar, CSIC. Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Catalonia, Spain.

[b]Department of Marine Biology, University of Vienna. Althanstrasse 14, 1090 Vienna, Austria.

Heterotrophic flagellates play an important role in aquatic systems. The diversity of this heterogeneous protist assemblage is generally based on cultivated strains, on which ultrastructural, physiological and molecular studies have been performed. In top of that, environmental molecular surveys have unveiled novel eukaryotic lineages that have refused cultivation so far, which might hide fundamental evolutionary and ecological features. Culturing bias appears often as a reasonable doubt behind ecological studies but has seldom been directly addressed. This study was designed to deal with this by analyzing the effect of organic matter in a confined community of natural heterotrophic flagellates. Several microcosms were established with an increasing amount of organic matter and different organic matter sources. Growth dynamics followed by epifluorescence microscopy and diversity analyzed by DGGE and clone libraries showed a clear substitution in the community, which differed more and more from the initial sample as the organic matter increased. Within this gradient there was also an increase of sequences related to cultured organisms, mostly chrysophytes, as well as a decrease on diversity indices. We conclude that culturing bias is partly explained by the use of organic matter in the isolation process, which drives a shift on the community to conditions closer to laboratory cultures. Only the use of alternative isolation methods will allow to access to the heterotrophic flagellates that constitute the abundant taxa of the original assemblage and have an active role in carbon and energy flow.

Key words: 18S rDNA; culture bias; chrysophytes; enrichment; heterotrophic flagellates; unamended incubations; DGGE; diversity.

Introduction

Marine Heterotrophic Flagellates (HF) perform key roles in microbial food webs and global biogeochemical cycles as trophic linkers and nutrient remineralizers (Sherr and Sherr 2002). HF are distributed all along the world oceans at concentrations between 10^2 and 10^4 cells ml^{-1}, representing 10-30% of protist cells in upper marine waters (Jürgens and Massana 2008). HF cells are often phagotrophs (Pernthaler 2005) but also may include dispersal stages of parasites of other marine organisms (Guillou et al. 2008). This central role in marine ecosystems has been translated into a great interest in studying these organisms under controlled conditions in the laboratory. Cultures have been essential for physiological and phylogenetic studies but the ecological relevance of cultured strains is not clear.

Bacterivorous HF have been cultivated using seawater supplemented with cereal grains or yeast extract that promote the growth of bacteria as food (Cowling 1991). In this rich media, bacteria are much larger and more abundant than at *in situ* conditions. This strategy retrieves mainly the same pools of species such as *Cafeteria* spp., *Paraphysomonas* spp., or *Bodo* spp. (Fenchel 1982b; Arndt et al. 2000, Scheckenbach et al. 2005), which are considered to be generally rare in the marine plankton (Jürgens and Massana 2008). In contrast, some of the most abundant and representative HF in the environment refuse cultivation (Massana et al. 2004a, Massana et al. 2006) reflecting what has been named culturing bias. This bias is also affecting the output of enrichment experiments. Thus, HF species that were found at very low abundance in the environment, such as *Paraphysomonas imperforata* (Lim et al. 1999) and *Cafeteria roenbergensis* (Massana et al. 2007), where the ones that overwhelmingly dominated in organic matter supplemented enrichments.

Although culturing bias and organic matter effect on enrichments is a well-known (and often

controversial) topic in prostistogy, there are no published studies where the effect of the quantity and quality of the added substrates on the microbial community is comprehensively analyzed using different techniques. Previous studies adressing the protists dynamics in microbial amended or unamended incubations (Lim, 1999, Countway 2005, Massana et al. 2006) focused on the evolution and properties of the incubated community along time, but were not designed to face the culturing bias conundrum.

The aim of this study was to report the effects of organic matter enrichments to microbial dynamics and HF community structure. We considered that an increase of organic matter in the enrichment would approach the system to classical culturing conditions, so we would expect the development of HF closely related to cultured ones. In the opposite side there were the unamended incubations, where we would expect protists similar to the ones dominating in the environment (Massana et al. 2006). The dynamics of bacteria and HF abundances were followed by epifluorescence microscopy. The diversity and identity of the HF proliferating in these enrichments were analyzed by DGGE fingerprinting and 18S rDNA clone libraries done at the eukaryotic peaks. The main goal of this study was to put together ideas and concepts related to the culturing bias that had been assumed or refused by protistologists and microbial ecologists but never specifically addressed. The enrichment effect on the abundance and composition of HF assemblages was obvious and consistent with the culturing bias scenario.

Results

We aimed to observe the differences on the community composition of Heterotrophic Flagellates (HF) among incubations differing in the amount and origin of organic matter and started with the same initial community of small protists (<3 µm). With this objective we performed 5 different microbial incubations (2 replica for each of them): one with no organic matter at all (O), what has been named unamended incubation; two with an increasing proportion of yeast extract: 0.01% (L) and 0.1% (M); and two other with 0.1% of an alternative source of organic matter, rice extract (R) and a known proportion of nutrients with glucose as the main carbon source (P). The objective of the first three incubations was to analyze the effect of the increase of the same source of organic matter in the incubated community. The last two had the objective of determining the effect of alternative organic matter sources added at the same concentration. To achieve our objectives we have used different analytical tools such as cell counts by epifluorescence microscopy, image analysis, and molecular techniques (DGGE, clone libraries and phylogenetic trees).

Table 1. Growth rate (µ), doubling time (DT), and flagellate yield of each incubation (calculated as the ratio of flagellates appeared and the decrease of 10^3 bacteria) based on cell abundance dynamics.

	µ (day-1)	DT (hours)	Yield (f/10^3b)
OA	1,38	12,10	4.5
OB	1,43	11,67	6.3
LA	2,22	7,49	8.2
LB	2,75	6,05	13.3
MA	3,40	4,89	18.5
MB	4,12	4,04	21.5
RA	3,02	5,52	16.4
RB	3,06	5,43	27.1
PA	4,13	4,02	11.3
PB	3,72	4,47	8.5

Heterotrophic flagellates and bacterial abundance dynamics

In all incubations we detected first a bacterial peak of abundance occurring the first 1-2 days, which was followed by a peak of HF that typically consumed bacterial cells (Fig. 1). All treatments showed a second bacterial peak occurring after 4-6 days of incubation. When we increased the organic matter added we observed an increase of bacterial and HF cells at the peaks and a delay in the occurrence of both peaks. In all cases, replicated treatments exhibited very coincident and reproducible dynamics. The abundance of phototrophic flagellates decreased during the 8 days of the experiment, typically becoming a very low percentage (<1%) of eukaryotes at the moment of the HF peak (data not shown).

In the O treatment bacterial and HF peaks appeared earlier than in the other treatments, at days 1 and 2.5, respectively, and exhibited the lower abundance, 2.5 x 10^6 bacteria mL^{-1} and 12 x 10^3 HF mL^{-1}. The second bacterial peak occurred at day 5.5 and was rather minor. Microbial peaks appeared later in the L treatment, on day 1.5 for bacteria and day 4 for HF. There was also a delay between both peaks, which occurred more separated than in the O treatment and had higher cell abundance: 10 x 10^6 bacteria mL^{-1} and 50-100 x 10^3 HF mL^{-1}. In this treatment the second bacterial peak appeared at the same time than in O, being half of the first peak. In treatment M there were also 2 bacterial peaks, the first one on day 2.5 and the second one on day 5.5, but in this case both were of comparable abundance. Interestingly, the HF peak did not appear until day 6 or 7 (depending on the replica) and seemed to

be feeding from the second bacterial peak. Treatment M exhibited the highest cell abundance at the peaks: 40 x 10^6 bacteria mL^{-1} and 400 x 10^3 HF mL^{-1}. In treatment R the timing of the peaks and its cell abundance were remarkably similar to L treatment, which had ten times less organic matter in the form of yeast extract. At P treatment the timing of the peaks (including the second bacterial peak) was very similar to the M treatment, so here HF were also feeding on the second bacterial peak, whereas the attained cell abundances were similar to treatments L and R. An overview of measured growth rates of HF assemblages indicated faster growth when increasing organic matter (Table 1). We also observed how the ratio between flagellate appeared and bacteria consumed (yield) was lower in the unamended incubation and P treatment than in the other three treatments.

In order to describe the morphology of food prey for HF, we analyzed bacterial cell sizes and the percentage of bacterial cells in aggregates at two time points along the incubations (Fig. 2). Before the HF peaks, both bacterial size and percentage of bacteria in aggregates generally increased in incubations richer in organic matter. The effect of HF grazing on bacterial assemblages was apparent by the analysis of samples after HF peaks, which in general showed a smaller bacterial cell size and a lower percentage of bacteria in aggregates. Nevertheless, there were some noticeable exceptions. In terms of cell size, the L treatment present slightly smaller cells than the O treatment, and HF in the P and R treatments could not reduce bacterial size. In addition, HF in the R treatment were not able to reduce bacterial aggregation.

Fingerprinting analysis of eukaryotic community structure

The DGGE image displayed a fingerprint of the protist composition in the initial assemblage and in samples at the HF peaks (Fig. 3 A). The cluster analysis revealed that O samples grouped with the initial sample (T0) whereas the remaining samples with organic matter additions clustered together (Fig. 3 B). Among these, samples that received yeast extract were related, as well as samples that received alternative organic matter additions.

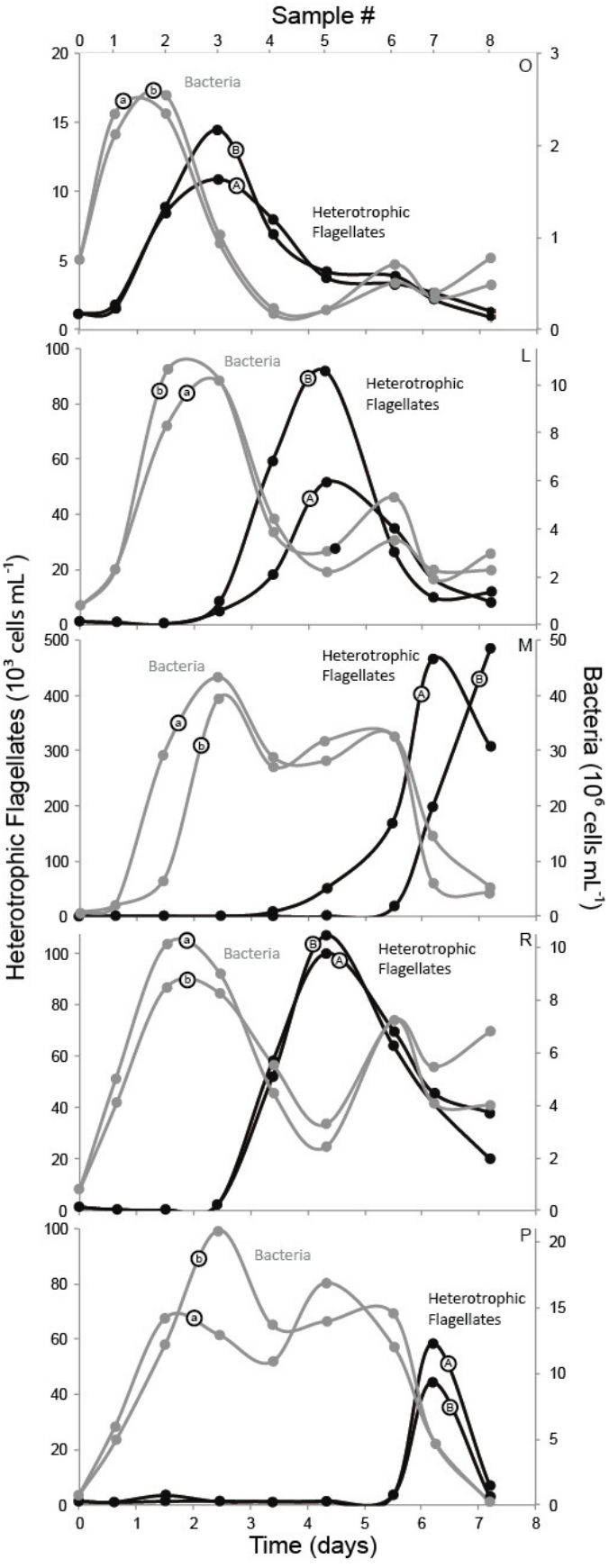


Figure 1. Growth dynamics of Heterotrophic Flagellates (black dots) and Bacteria (grey dots) in the five organic matter enrichments followed by epifluorescence microscopy. Treatment O corresponds to an unamended incubation, L to 0.01% yeast extract addition, M to 0.1% yeast extract addition, R to 0.1% rice extract addition and P to 0.1% glucose based enrichment. There are two replicas for each treatment (HF: A and B, bacteria: a and b). Sample numbers are shown in the top of the graphs.

Effect of organic matter on HF diversity

We prepared clone libraries for samples at the HF peaks in order to determine the phylogenetic affiliation of the communities developing in each incubation. For this analysis we chose only one microcosm of each treatment (replica A), given that DGGE fingerprints revealed a high reproducibility among replicas. We calculated the number of OTUs (clustered at a 99% similarity criterion) in order to determine the diversity in each sample. Chrysophytes and MAST were the dominant groups in the different treatments, and by far most clones affiliated within these two stramenopile groups (Table 2). For the unamended incubation sample (OA3) we sequenced 32 clones and identified 15 different OTUs, 9 corresponding to chrysophytes, 4 to MAST, 1 to prasynophytes and 1 to dinoflagellates. In the L treatment (LA5) we sequenced 71 clones and found 15 different OTUs. In this case 11 OTUs corresponded to chrysophytes, 5 to MAST and 1 to radiolaria. In the M clone library (MA7; 45 clones sequenced) there was a striking decrease in the number of OTUs, only 6 (5 chrysophytes and 1 MAST). In the R and P treatments (RA5 and PA7) again we went down in the number of OTUs, as the 14 and 21 clones sequenced only yielded 3 or 4 OTUs, respectively. Although the sequencing effort in R5A and PA7 was lower than in the other libraries, the pattern of low diversity was clear enough. In the R sample 2 OTUs affiliated to chrysophytes, 1 to MAST and 1 to bolidophytes, whereas in the P treatment 3 OTUs corresponded to chrysophytes and 1 to MAST. The lower diversity suggested by the lower number of OTUs when increasing organic matter was corroborated by the Shannon diversity Indices (Shannon 1948) based on the number of clones belonging to each OTU. This index decreased from 2.5 at the unamended incubation to values less than 1 in the 0.1% organic matter treatments (Fig. 4).

Trying to define the possible role of organic matter altering the community and driving culturing bias, we classified the sequences (and OTUs) in two groups, those with similarity values above or below 94% to its Closest Cultured Match (CCM) in GenBank (Table 2). This classification has been done under the hypothesis that organic matter will increase the number of sequences similar to cultured protists (over 94% to its CCM). In O treatment most of the defined OTUs (10 out of 15) have similarity values under 94% to CCM. An intermediate situation occurred in the medium amended treatment (L), in which 9 of 15 OTUs had similarity values over 94% to CCM. For the 0.1% amended treatments (M, R and P) almost all OTUs had similarity values over 94% to CCM. So, the shift from "uncultured" to "cultured" HF by increasing organic matter was obvious (Fig. 4). This shift was also clear when analyzing the number of sequences in each library.

The cloning and sequencing analysis revealed that most sequences belonged to chrysophytes. So we constructed a chrysophyte ML phylogenetic tree in order to go deeper in the phylogenetic affiliation of these sequences, compare chrysophyte sequences in different treatments, and investigate if they belong to cultured and uncultured clades (Fig. 5). Most chrysophyte sequences from the unamended incubation belonged to novel clades such as Clade H and Clade I, whereas sequences obtained from enriched treatments belonged mainly to Clades F1 and J, which contain the well-known *Paraphysomonas* and *Oikomonas* species.

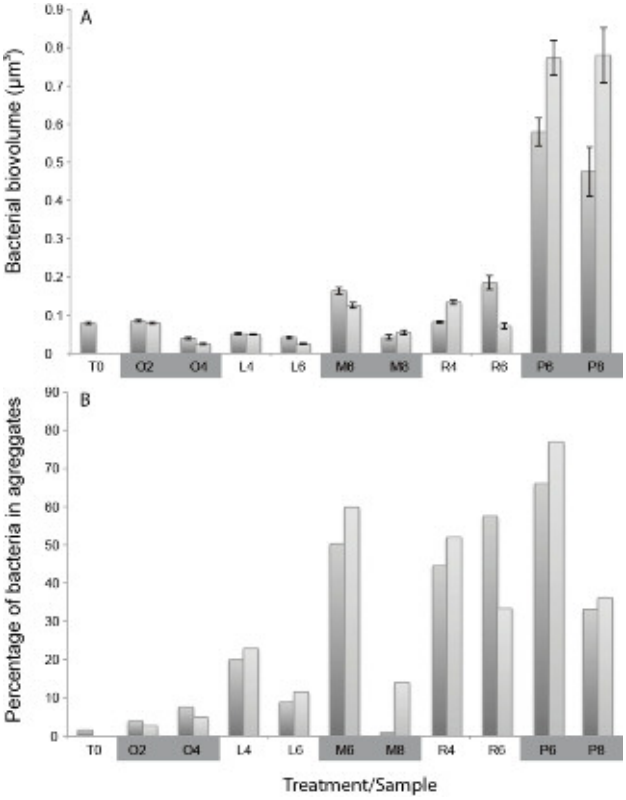


Figure 2. A. Averaged biovolume of bacterial cells in the original sample (T0) and in the two replicas of the five treatments before and after HF peaks. Error bars show SE of individual measurements. **B.** Percentage of bacterial cells forming aggregates in the same samples.

Discussion

To determine the effect of organic matter enrichments on the heterotrophic component of the microbial loop (organic matter - heterotrophic bacteria - heterotrophic flagellates) we analyzed the original community at time 0 and its evolution under different conditions. By using a multi-technique analysis we have made an effort to give an integrated view of the organic matter generated bias in enrichments and its direct link with culturing bias in protistology

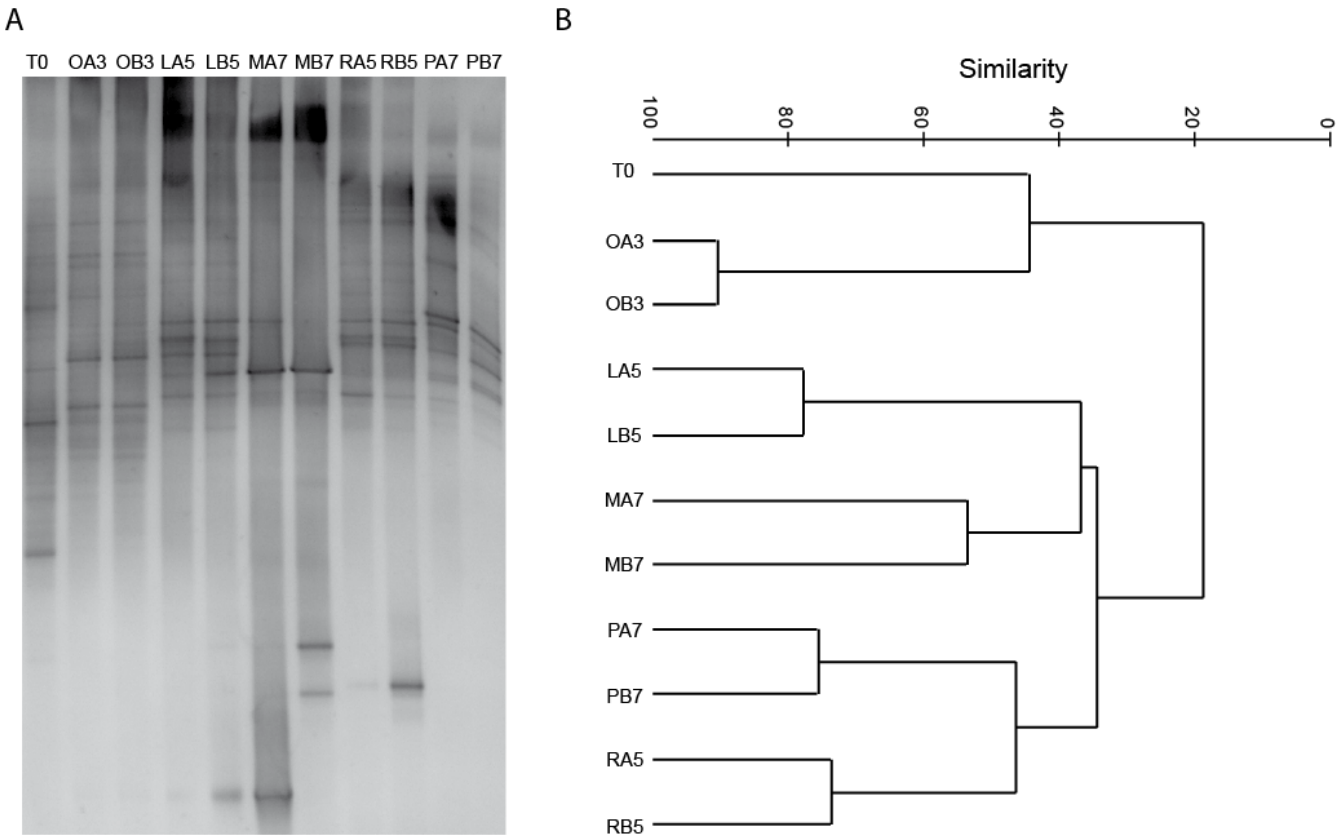


Figure 3. A. 18S rDNA DGGE fingerprint of eukaryotic assemblages obtained from the original sample (T0) and the enrichments at HF peaks. **B.** Dendrogram relating the samples calculated on the basis of Bray-Curtis dissimilarity.

In incubations where we added the same organic matter source at different concentrations (O, L and M), the peak abundance of bacteria and HF increased from O to M and this was accompanied by a delay in the apparition of both peaks. There was also a larger gap between the timing of the bacterial and HF peaks. The delay in the bacterial peak could be due to the time needed by the community to adapt to the enrichment conditions. In unamended incubations (O bottle) the original populations found the right conditions to start growing. On the other hand, in the enriched L and M incubations, the dominant populations adapted to lower nutrient concentrations could be inhibited to grow (Rappé et al. 2002), and some minoritary population, well-adapted to high nutrient concentrations, could fastly develop and adopt a dominant role in the community (Eilers et a. 2000). So this bacterial community will derive from a lower inoculum size, delaying the apparition of the peak. Another factor that could delay the peak is the number of bacterial cells in the peak, typically one order of magnitude higher in the enriched samples, which would then need more time of exponential growth. Similar reasons can be invoked for the delay in the appearance of HF peaks. In particular HF communities in high organic matter incubations will need to adapt to high bacterial concentrations, larger bacteria, and in some cases the formation of aggregates. Most likely, the original dominant HF species were not prepared for these enriched conditions. The differences in growth rate are consistent with different species growing in different incubations (Table. 1), with fast growing populations in enriched conditions. Again, these fast growing opportunistic HF species were likely in low concentrations in the original oligotrophic sample (Pedrós-Alió 2006), and needed an extra time to growth from this low inocula and attain the high observed abundances.

All treatments had two bacterial peaks, being the second more important at increased organic matter. The second bacterial peak was not necessary a result of the consumption of the first peak by HF but could be a substitution in the bacterial community (Allers et al. 2007). These authors found that in similar microcosm experiments the first peak was formed mostly by *Alteromonadaceae* and the second peak moslty by *Rhodobacteraceae*. We must highlight that HF in the M treatment are eating bacteria from the second peak. This significant delay in HF growth could be explained by the large bacteria and the aggregates appearing at high concentrations of organic matter (Fig. 2). Consuming large or aggregated bacteria is more difficult than eating free living bacteria (Jürgens & Güde 1994), and only some specialized species can do it and proliferate in these conditions. In the M treatment the dominating *Paraphysomonas* spp. and *Oikomonas* spp. are known to be easily isolated

from marine snow (Davis et al. 1984). Those flagellates were clearly able to consume this bacterial field, composed by large free-living cells (mean volumes of 0.2 μm^3, three times larger than *in situ* sizes) and up to 60% of bacteria in aggregates. The consumption of large bacteria is also reflected on the yield data, because the number of bacteria necessary for the generation of one flagellate decreases along the increase of organic matter.

In R and P treatments, which had an alternative source of organic matter (at the same concentration than M) we identified similar dynamics with respect the timings of the peaks than in L and M, respectively. Hence, HF in the P treatment were also feeding on the second bacterial peak. In addition, the bacterial and HF abundance at the peaks was similar between R and P and much lower than the abundance attained in M. This clearly highlights the differences in HF developments in function of the type of organic matter.

The organic matter source in the R treatment, rice extract, is rich in starch. The use of these large molecules requires production of starch hydrolyzing exoenzymes, and not all bacteria have this enzymatic machinery. So starch is less accessible and less desirable for most bacteria than yeast extract. This could explain why the R treatment had bacterial abundances closer to L treatment, which had one order of magnitude less organic matter but used a more attractive source. In fact, microbial dynamics in R and L treatments were strikingly similar. In the case of the P treatment, although glucose is an accessible carbon source and was properly supplemented with N and P, this treatment might lack some oligoelements (such as aminoacids or vitamins) that are present in yeast extract. This could again explain the lower bacterial numbers as compared with the M treatment. Moreover, the very large bacteria (0.6-0.8 μm^3) in the first peak, unique in this treatment (Fig. 2), seemed to become a grazing refuge that avoided HF exploitation. As previously said large bacteria can be a protection strategy developed by bacterial populations against predation (Hahn et al. 2000, Jürgens & Güde 1994, Simek et al. 2001). Flagellate yield values for the P treatment does not reflect the consumption of very large cells and deviates from the consistent values detected in the other four treatments (Table 1). This suggests that other mechanisms might be involved in the crash of the bacterial peak in the P treatment.

A general analysis of HF diversity by DGGE fingerprinting revealed that the unamended treatments grouped with the original sample and the rest of the enriched treatments grouped together in another cluster. This agrees with previous studies that showed that unamended incubations promoted the growth of HF present in the natural assemblage and prevented a great modification of the community structure (Massana et al. 2006). On the contrary the enrichments are promoting the mergence of other populations not very abundant in the original sample (Lim et al 1999, Massana et al. 2007).

Table 2. Sequences obtained in 18S rDNA clone libraries for the different samples at the HF peak, classified based on phylogeny and depending on the similarity to its CCM (94% as the boundary). Number within brackets display the number of OTUs (grouped at a 99% similarity criteria) of each group of sequences.

		Total	Chrysophytes	MAST	Prasynophytes	Bolidophytes	Dinophytes	Radiolaria
	All	32 (15)	18 (9)	11 (4)	2 (1)	-	1 (1)	-
OA3	>94%	8 (5)	5 (3)	1 (1)	2 (1)	-	-	-
	<94%	24 (10)	13 (6)	10 (3)	-	-	1 (1)	-
	All	71 (15)	65 (11)	5 (3)	-	-	-	1 (1)
LA5	>94%	61 (9)	61 (9)	-	-	-	-	-
	<94%	10 (6)	4 82)	5 (3)	-	-	-	1 (1)
	All	45 (6)	42 (5)	3 (1)	-	-	-	-
MA7	>94%	42 (5)	42 (5)	-	-	-	-	-
	<94%	3 (1)	-	3 (1)	-	-	-	-
	All	14 (3)	12 (2)	1 (1)	-	1 (1)	-	-
RA5	>94%	13 (2)	12 (2)	-	-	1 (1)	-	-
	<94%	1 (1)	-	1 (1)	-	-	-	-
	All	21 (4)	20 (3)	1 (1)	-	-	-	-
PA7	>94%	20 (3)	20 (3)	-	-	-	-	-
	<94%	1 (1)	-	1 (1)	-	-	-	-

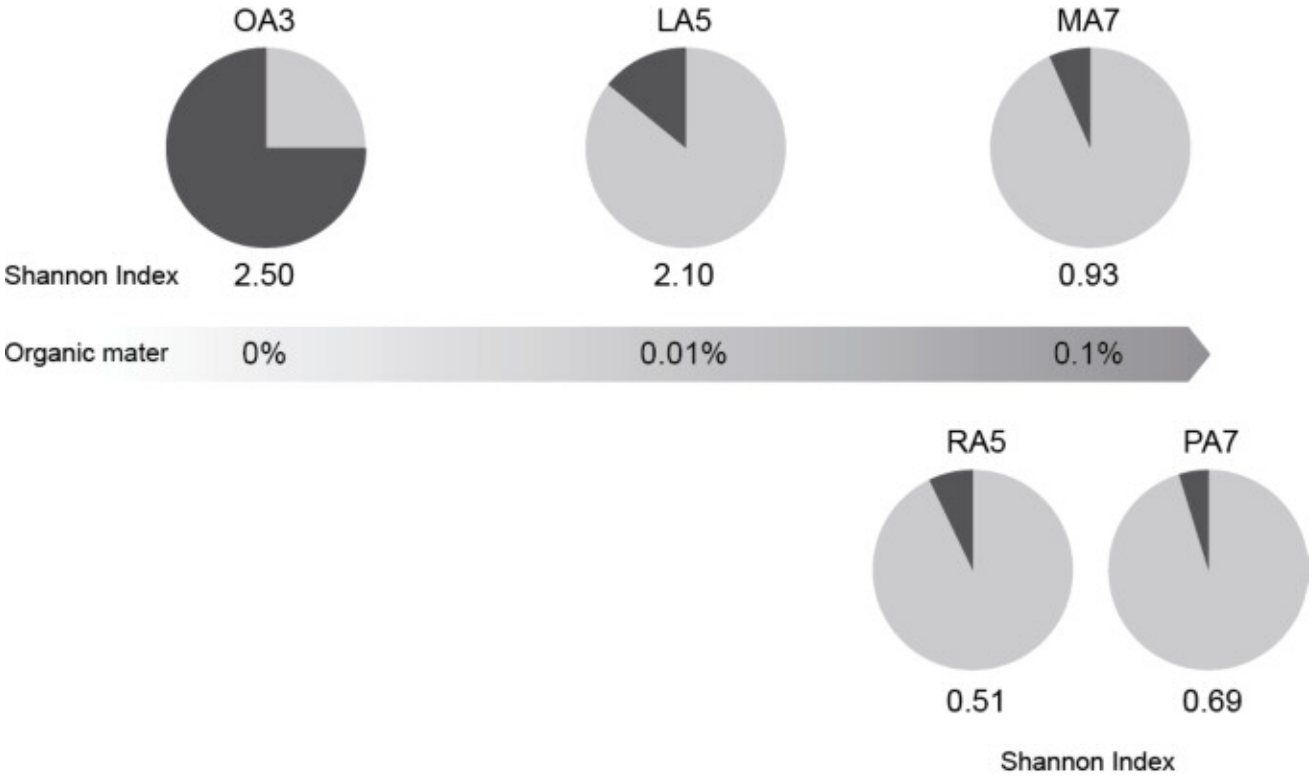


Figure 4. Pie charts representing the proportion of clones in five 18S rDNA libraries affiliating to uncultured (similarity against its CCM <94%; light grey) or cultured (similarity against its CCM >94%; dark grey) protists. The Shannon diversity index is also shown for each library under its corresponding pie chart.

A more detailed phylogenetic analysis highlights clearly the bias effect caused by the organic matter. While in OA3 clone library there was a predomination of uncultured protists, the increase of organic matter reversed this trend, and cultured protists became clearly dominant in M, R and P treatments, being L at an intermediate stage. This fact is even more obvious when we look at the chrysophyte tree and see how the sequences belonging to enriched samples belong to clades with a high representation of cultured organism and the ones from the unamended incubation belong to novel clades (del Campo and Massana 2011).

The selective and homogenizer role of the organic matter was confirmed by a clear decrease of the diversity when increasing organic matter concentration (Fig. 4). Whereas the Shannon index in OA3 library was 2.5, this index decreased to less than 1 in MA7, RA5 and PA7 (being again LA5 in an intermediate stage). Organic matter enrichments resulted in communities not only dominated by cultured organisms but also less diverse, here dominated by *Paraphysomonas* spp. and *Oikomonas* spp. Those species have a great capacity to eat bacteria (Fenchel 1982a, Caron et al. 1985, Eccleston-Parry et al. 1994) and potentially were able to outcompete the organisms that were originally dominant in the oligotrophic initial sample.

Culturing bias is an important obstacle for protistologists that intend to obtain a complete picture of the eukaryotic tree of life and for microbial ecologists that intent a better understanding of marine ecosystem functioning. There have been many indirect signs of culturing bias, for example that different organisms are retrieved by using culturing independent approaches (Díez et al. 2001a, López-García et al. 2001, Moon-van der Staay et al. 2001) than by using culturing dependent ones (Patterson et al. 1993, Vørs et al. 1995, Tong 1997, Tong et al. 1997, Tikhonenkov and Mazei 2006, Tikhonenkov et al. 2006). More direct evidences derive from the fact the enriched species are often found at very low abundance in the original sample (Lim et al. 1999, Massana et al. 2007). In this study we aimed at adressing the culturing bias starting with the same initial assemblage and following HF developing with different sources and amounts of organic matter. We have showed how the increase in organic matter modifies the dynamics of the community due to a substitution process within both bacterial and HF members of the initial community. Also we observed a dramatic decrease on diversity and a gradual increase of cultured species. Our data highlights that classical culturing techniques based on enrichments need to be complemented with novel culturing strategies in order to really catch the protists responsible for most processes in the sea. Culturing efforts done up to now have been extremely important, but a new culturing impulse is needed to advance on our understanding of protist evolution and its ecological role

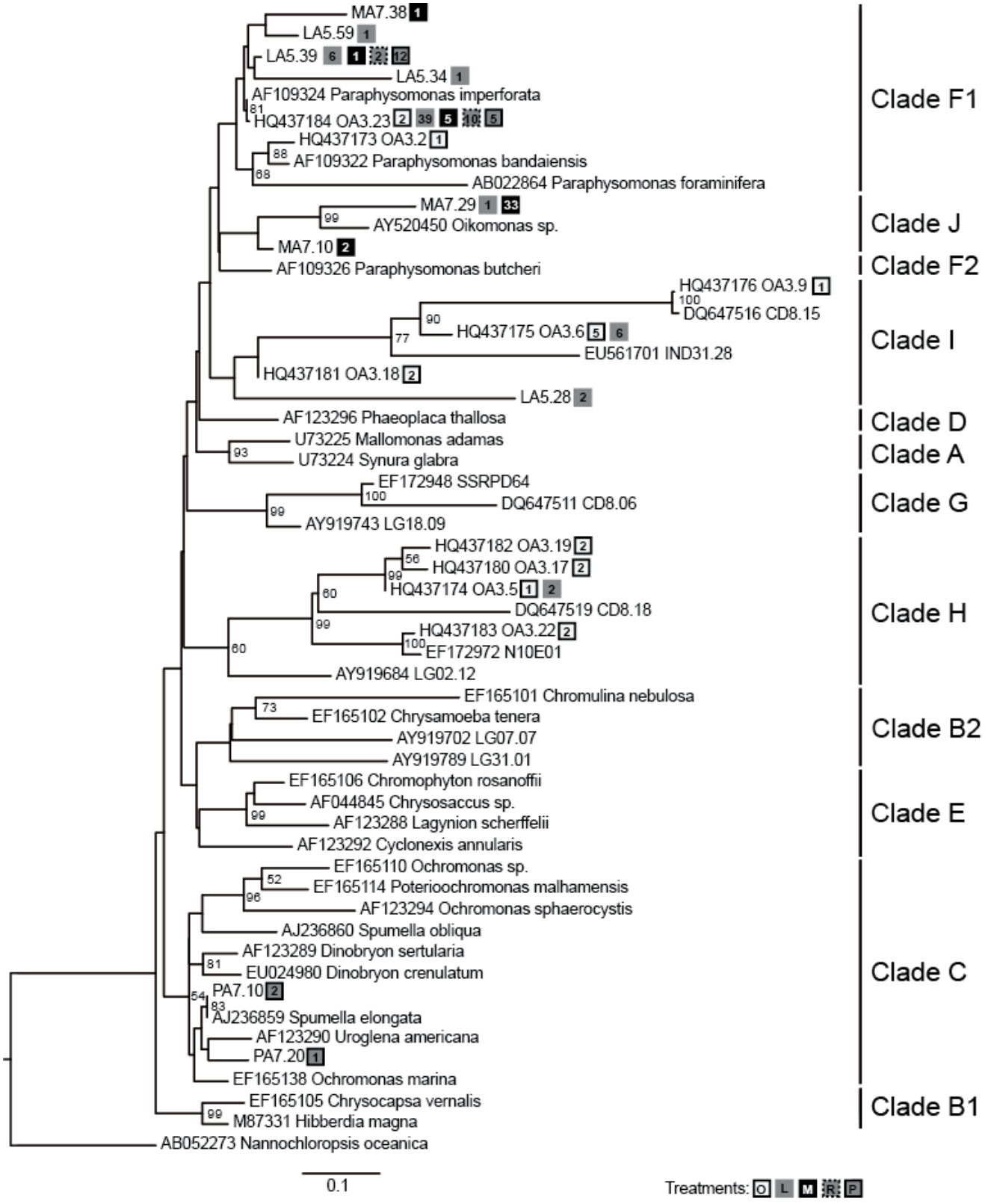


Figure 5. Maximum Likelihood phylogenetic tree of chrysophytes constructed with 38 complete and 16 partial 18S rDNA sequences (1444 informative positions). ML bootstrap values over 50% are shown. Only one sequence representative of each OTU (delineated at 99% similarity) is shown in the tree. The number of sequences from different libraries assigned to each OTU appears after the corresponding sequence name. The scale bar indicates 0.1 substitutions per position.

Methods

Incubations: Water samples from the Blanes Bay Microbial Observatory (NW Mediterranean Sea) were collected 800 m offshore in March 6th 2006. Surface seawater was filtered by gravity first through a nylon mesh of 200 µm and later through polycarbonate filters (Poretics) of 3 µm pore size. Five treatments were prepared with two replicas each: Unamended Incubation (O condition), 0.01 and 0.1% (w/v) of yeast extract (L and M respectively), 0.1% (w/v) of rice extract (R) and 0.1% (w/v) of nutrients in known proportion (P). Rice extract was obtained after boiling rice grains, discarding rice grains and liofilizing the remaining water. The product obtained after liofilization was used as substrate for the R enrichment. In P treatment C:N:P were added as glucose, KH_2PO_4 and NH_4Cl in a C:N:P = 106:16:1 ratio. For each treatment 5 L of the filtered seawater (< 3 µm) were dispensed into 8 L containers and incubated in the dark at 20ºC in a laboratory chamber. Bottles were sampled every day. Subsamples for epifluorescence microscopy were fixed with cold glutaraldehyde (1% final concentration), stained with DAPI and filtered onto 0.2 or 0.6 µm pore-size black polycarbonate filters (Poretics) for numeration of bacteria and flagellates. Bacterial counts included both free-living and aggregated cells. Subsamples of 100 ml were filtered onto 0.2 µm pore-size Durapore filters, submerged in lysis buffer (40 mM EDTA, 50 mM Tris-HCl and 0.75 M sucrose) and kept frozen at –80°C until DNA extraction.

Image analysis of bacterial cell biovolume. Filters for microscopy 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 (Olympus DP72; Olympus Europa GmbH, Hamburg) 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). The percentage of bacterial cells forming aggregates was estimated by direct manual counts from the same pictures.

DGGE. Cell lysis was performed by digestion with lysozyme followed by proteinase K and SDS treatments. DNA was purified by phenol/chloroform extraction and concentrated with a Centricon-100 (Millipore) as described previously (Díez et al. 2001a). About 10 ng of extracted DNA was used as template in a PCR in which eukaryotic 18S ribosomal DNA (rDNA)-specific primers Euk1A and Euk516r-GC were used to amplify a fragment approximately 560 bp long. Details of the primers and PCR conditions are explained elsewhere (Díez et al. 2001a). DGGE was performed with a DGGE-2000 system (CBS Scientific Company) as described previously (Muyzer et al. 1997) using described settings (Díez et al. 2001b). The resulting gel was visualized with UV radiation by using a Fluor-S MultiImager and the MultiAnalyst imaging software (Bio-Rad, USA). Digitized DGGE images were analyzed with the Chemidoc software (Bio-Rad) to obtain the relative contribution (in percentage) of each DGGE band to the total band signal in each lane of the gel. Bands occupying the same position in the different samples were identified. A matrix was then constructed taking into account the presence and relative intensity of individual bands in each sample. Based on this matrix (arcsin transformed), we produced a Bray-Curtis dissimilarity matrix and samples were grouped based on a dendogram (CLUSTER analysis, PRIMER, Plymouth, UK).

18S rDNA cloning and sequencing. PCR was performed with primers EukA and EukB as detailed in Díez et al. 2001a. The PCR product was used to construct a clone library with a TA cloning kit (Invitrogen). The presence of an 18S rDNA insert in positive clones was confirmed by PCR reamplification with the same primers. 183 clones were partially sequenced with the primer 528f by the MACROGEN Genomics Sequencing Services. Sequences were identified and inspected for chimeras by BLAST (Altschul et al. 1997) and KeyDNATools (Guillou et al. 2008), yielding 183 sequences (accession numbers xxxxxxxx – xxxxxxxx). Clone sequences where blasted again in order to determine their similarity against their closest cultured match, CCM (del Campo and Massana 2011). For convenience, sequences with a similarity value against its CCM over 94% are considered that derive from cultured protists, whereas those below 94% are considered from uncultured protists.

Phylogenetic analyses: 18S rDNA sequences were aligned using MAFFT (Katoh et al. 2002) using a close relative as out-group. Alignments were checked with Seaview 3.2 (Galtier et al. 1996) and highly variable regions of the alignment were removed using Gblocks (Castresana 2000). Maximum likelihood (ML) phylogenetic trees were constructed with the fast ML method RAxML (Stamatakis 2006) using the evolutionary model GTRMIXI. Phylogenetic analyses were done in the freely available University of Oslo Bioportal (www.bioportal.uio.no). Repeated runs on distinct starting trees were carried out to select the tree with the best topology (the one having the best Likelihood of 1000 alternative trees). Bootstrap ML analysis was done with 1000 pseudo-replicates and the consensus tree was computed with MrBayes (Huelsenbeck and Ronquist 2001). Trees were edited with FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

References

Allers E, Gómez-Consarnau L, Pinhassi J, Gasol JM, Simek K, and Pernthaler J (2007) Response of Alteromonadaceae and Rhodobacteriaceae to glucose and phosphorus manipulation in marine mesocosms. Environ Microbiol **9**: 2417-2429

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

Arndt H, Dietrich D, Auer B, Cleven EJ, Gräfenhan T, Weitere M, Mylnikov AP (2000) Functional Diversity of Heterotrophic Flagellates in Aquatic Ecosystems. In Leadbeater BSC and Green JC (eds) The Flagellates: Unity, Diversity and Evolution. Taylor & Francis Press, London, pp 240-268

Caron DA, Goldman JC, Andersen OK, Dennett MR (1985) Nutrient cycling in a microflagellate food chain: II. Population dynamics and carbon cycling. Mar Ecol Prog Ser **24**: 243-254

Countway PD, Gast RJ, Savai P, Caron DA (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. J Eukaryot Microbiol **52**: 95-106

Cowling AJ (1991) Free-living heterotrophic flagellates: methods of isolation and maintenance, including sources of strains in culture. In Patterson DJ and Larsen J (eds) The biology of free-living heterotrophic flagellates. Oxford University Press, Oxford, pp 477-492

Davis PG, Sieburth JM (1984). Estuarine and oceanic microflagellate predation of actively growing bacteria : estimation by frequency of dividing-divided bacteria. Mar Ecol Prog Ser **19**: 237-246

del Campo J, Massana R (2011) Emerging Diversity within Chrysophytes, Choanoflagellates and Bicosoecids Based on Molecular Surveys. Protist **162**: 435-448

Díez B, Pedrós-Alió C, Massana R (2001a) 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

Díez B, Pedrós-Alió C, Marsh TL, Massana R (2001b) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. Appl Environ Microbiol **67**: 2942-2951

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

Eilers H, Pernthaler J, Amann R (2000) Succession of pelagic marine bacteria during enrichment: a close look on cultivation-induced shifts. Appl Environ Microbiol **66**: 4634-4640

Fenchel T (1982a) Ecology of heterotrophic microflagellates. II. Bioenergetics and growth. Mar Ecol Prog Ser **9**: 225-231

Fenchel T (1982b) Ecology of heterotrophic microflagellates. IV. Quantitative occurrence and importance as bacterial consumers. Mar Ecol Prog Ser **9**: 35-42

Galtier N, Gouy M, Gautier C (1996) SeaView and Phylo_Win, two graphic tools for sequence alignment and molecular phylogeny. Comput Applic Biosci **12**: 543-548

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R, Scanlan DJ, Worden AZ (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). Environ Microbiol **10**: 397-408

Hahn MW, Moore ERB, Höfle MG (2000) Role of microcolony formation in the protistan grazing defense of the aquatic bacterium Pseudomonas sp. MWH1. Microb Ecol **39**: 175-185

Jürgens K, Güde H (1994) The potential importance of grazing-resistant bacteria in planktonic systems. Marine Ecology Progress Series **112**: 169-188

Jürgens K, Massana R (2008) Protistan Grazing on Marine Bacterioplankton. In Kirchman DL (ed) Microbial Ecology of the Oceans, Second Edition. John Wiley & Sons, Inc., New York, pp 383-441

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

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

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

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

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

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

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

Muyzer G, Brinkhoff T, Nübel U, Santegoeds C, Schäfer H, Wawer C (1997) Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In Akkermans ADL, van Elsas JD, de Bruijn FJ (ed) Molecular microbial ecology manual, vol. 3.4.4. Kluwer Academic Publishers, Dordrecht, pp 1-27

Patterson DJ, Nygaard K, Steinberg G, Turley CM (1993) Heterotrophic flagellates and other protists associated with oceanic detritus throughout the water column in the mid North Atlantic. J Mar Biol Assoc UK **73**: 67-95

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol **14**: 257-63

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

Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature **418**: 630-633

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J **27**: 379-656

Scheckenbach F, Wylezich 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.

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

Simek K, Pernthaler J, Weinbauer MG, Hornak K, Dolan JR, Nedoma J, Masin M, Amann R (2001) Changes in bacterial community composition and dynamics and viral mortality rates associated with enhanced flagellate grazing in a mesoeutrophic reservoir. Appl Environ Microbiol **67**: 2723–2733

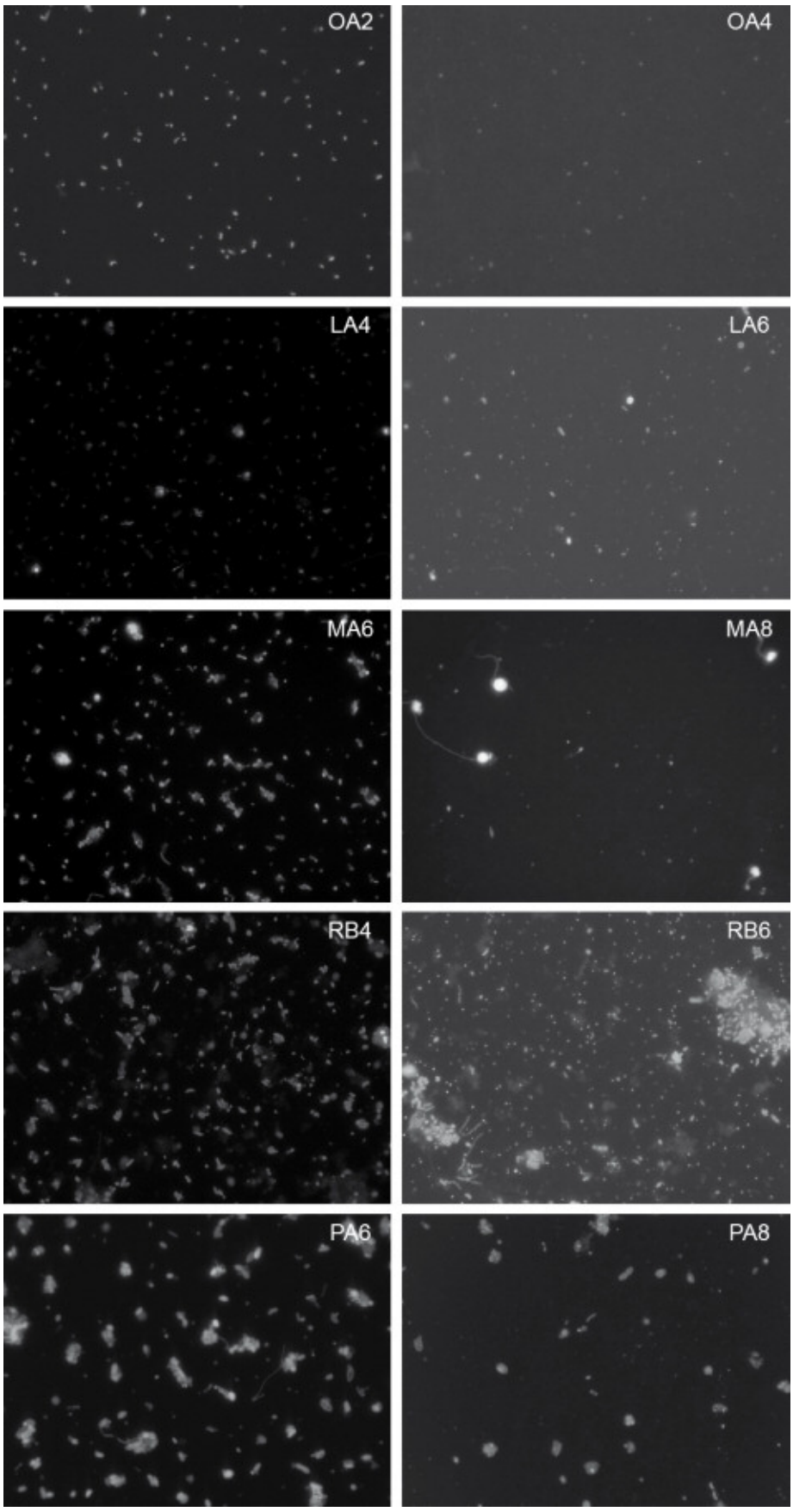
Tikhonenkov DV, Mazei YA (2006) Distribution of heterotrophic flagellates at the littoral of estuary of Chernaya River (Kandalaksha Bay, White Sea). Russ J Mar Biol **32**: 276-283.

Tikhonenkov DV, Mazei YA, Mylnikov AP (2006) Species diversity of heterotrophic flagellates in White Sea littoral sites. Europ J Protistol **42**: 191-200.

Tong SM (1997) Heterotrophic flagellates from the water column in Shark Bay, Western Australia. Marine Biology **128**: 517-536.

Tong SM, Vørs N, Patterson DJ (1997) Heterotrophic flagellates, centrohelid heliozoa and filose amoebae from marine and freshwater sites in the Antarctic. Polar Biology **18**: 91-106.

Vørs N, Buck K, Chavez F, Eikrem W, Hansen L, Ostergaard J, Thomsen H (1995) Nanoplankton of the equatorial Pacific with emphasis on the heterotrophic protists. Deep Sea Res Part II **42**: 585-595



Supplementary figure 1. Micrographs of the bacterial assemblages taken from the different treatments before and after the HF peak.

5

Domant els depredadors més petits de l'oceà

Els protists (eucariotes unicel·lulars) representen la major part de la diversitat eucariota i són actors protagonistes dins els processos de la biosfera. La major part de la diversitat de protists descrita i els coneixements que tenim d'ells a nivell ecofisiològic es basen principalment en soques cultivades. No obstant això, els estudis moleculars ambientals han revelat llinatges completament nous que, com els seus homòlegs procariotes, són essencialment organismes que mai s'han cultivat. El biaix de cultiu és sens dubte un dels majors inconvenients que es poden trobar en microbiologia i altres disciplines relacionades i és particularment greu en el cas dels protists heteròtrofs, que depenen de les fonts orgàniques de nutrients per al seu creixement. Aquí mostrem com hem estat capaços d'obtenir en cultiu protists bacterívors d'importància ecològica imitant en el medi de cultiu les condicions del seu ambient natural. Fent servir dilucions seriades o per selecció per mitjà de citometria de flux, cèl·lules individuals van ser inoculades en aigua de mar envellida amb una concentració de bacteris naturals lleugerament superior a la del medi. Hem aconseguit d'aquesta manera aïllar soques pertanyents als llinatges que, fins al moment, només es coneix la seva existència a partir de seqüències d'ADNr 18S ambientals. Entre elles *Minorisa minuta* candidatus forma una nova branca dins els Rhizaria, ocupant una posició clau des d'un punt de vista evolutiu. Amb una mida mitjana de 1,4 micres representa el bacterívor més petit conegut fins a dia d'avui. Té una amplia distribució en sistemes planctònics a nivell mundial i representa un 5% del protists heteròtrofs de les comunitats d'aigües costaneres. Les característiques fisiològiques d'aquesta soca pot explicar en part el seu èxit en el medi ambient. El cultiu de protists ecològicament rellevants, però que fins avui havien refusat el cultiu, poden proporcionar material d'incalculable valor per a l'ecofisiologia, la genòmica, la modelització d'ecosistemes i per la resolució de diverses qüestions evolutives.

Taming the smallest predators of the oceans

Javier del Campo[1], Fabrice Not[1,2], Irene Forn[1], Michael E. Sieracki[3] and Ramon Massana[1]*

[1] *Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC. Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.* [2] *Marine Plankton Group, UMR7144 Station Biologique de Roscoff, CNRS-UPMC. Place Georges Teissier, 29680 Roscoff, France.* [3] *Bigelow Laboratory for Ocean Sciences. 180 McKown Point Road. 04575-0475 West Boothbay Harbor, Maine, USA.*

Protists (unicellular eukaryotes) arguably account for most eukaryotic diversity and are central players of the biosphere. Known protist diversity and biology is largely based on cultured strains. Yet, environmental molecular surveys have unveiled entirely novel lineages that, as their prokaryotic counterparts, are essentially uncultured. Culture bias is arguably the major drawback for any microbe related science and is particularly severe for heterotrophic protists, which depend on organic food sources for growth. Here we show how ecologically significant bacterivorous protists have been brought into culture by mimicking *in situ* conditions. Single cells sorted by serial dilution or flow cytometry were inoculated into seawater amended with natural bacterial assemblage at nearly *in situ* abundances. Strains belonging to lineages only known so far from environmental sequencing were isolated. Among them, *Minorisa minuta* candidatus forms a novel branch within Rhizaria, holding a key evolutionary position, and with an average size of 1.4 µm represents the smallest bacterial grazer known to date. It has a worldwide planktonic distribution and accounts for 5% of heterotrophic protists communities in coastal waters. Physiological features of this strain can partly explain its success in the environment. Culturing ecologically relevant but elusive protists provide invaluable material for ecophysiology, genomics, ecosystem modeling, and evolutionary issues.

Unicellular organisms are major forces driving our planet ecosystems and are an outstanding reservoir of biological diversity (genes, molecules, metabolic pathways, and cellular processes) yet to be discovered[1]. They are also main actors in macro- and microevolutionary processes for life on Earth[2]. Nowadays, culture independent approaches are regularly applied to investigate the diversity and function of microbes in the environment. Despite the valuable information provided by "omics" environmental studies, culture bias definitely remains as one of the most critical challenges faced by scientists aiming to achieve a full understanding of the ecological role of microbes[3] and is currently a bottleneck in ecosystem studies[4-5]. Environmental DNA surveys demonstrate the extent to which culturing efforts poorly capture *in situ* microbial diversity[6]. Many lineages unveiled in the last few years and holding key phylogenetic positions to understand macroevolutionary patterns among eukaryotes are essentially composed of environmental sequences[7]. It is estimated that as little as 0.1 to 1% of bacterial and protist cells can be easily cultured[8-9]. Ironically, the most represented taxa in the environment refuse culturing while most of the strains represented in culture are very scarce in the environment[10]. Consequently, ultrastructural, physiological and genomic information for many ecologically relevant microorganisms are missing.

This culture bias can be overcome by using original culturing strategies, as demonstrated for *Pelagibacter ubique* and marine crenarchaea[11-12], both initially detected through environmental molecular surveys and later identified as ecologically relevant taxa. *Pelagibacter ubique* was brought into culture by mimicking oligotrophic conditions and marine crenarchaea was cultured in media amended with ammonia once molecular data revealed they were ammonia oxidizers. Similar culturing efforts have seldom been applied to marine protists, even though culture bias is perceived as a major limitation to investigate further the functional role and ecological significance of photosynthetic protists[13], being even more severe for the heterotrophic ones[14].

Marine heterotrophic flagellates perform key processes in microbial food webs as bacterial grazers, trophic linkers, and nutrient remineralizers[15-16]. They exhibit a variety of trophic strategies and constitute a diverse assemblage of poorly identified species[17-18]. Up to now, bacterivorous flagellates have been invariably cultured using rich media composed of seawater supplemented with rice grains or yeast extract that promote the growth of very large bacteria at very high densities. This strategy inexorably retrieves the same species complexes (i.e. *Cafeteria* spp., *Paraphysomonas* spp., *Bodo* spp.) that are known to be rare in marine plankton[14]. Abundant taxa identified by molecular surveys still remain uncultured (e.g. bacterivorous MASTs clades[19]). In order to bring into culture ecologically relevant heterotrophic flagellates, we mimicked oligotrophic marine conditions by amending sterile seawater with a mix of natural bacteria collected from the same sampling site at abundances only slightly higher than *in situ* (5×10^6 cells mL^{-1}). Each pre-culture was initiated with a single cell, obtained by serial dilution or by flow cytometry sorting, and incubated in the dark at *in situ* temperature. The full culturing protocol is outlined in Supplementary Fig. 1.

Pre-cultures based on serial dilution yielded a 5.2% success rate (growth observed in 25 of the 480 inoculated wells). Out of the 25 pre-cultures, 12 were stable and were scaled up to 30 mL culture volumes. Based on their 18S rDNA, 4 pre-cultures were identified as *Paraphysomonas* spp. and the others represented taxa closer to environmental sequences than to any known culture, including a cercozoan, a chlorarachniophyte (two pre-

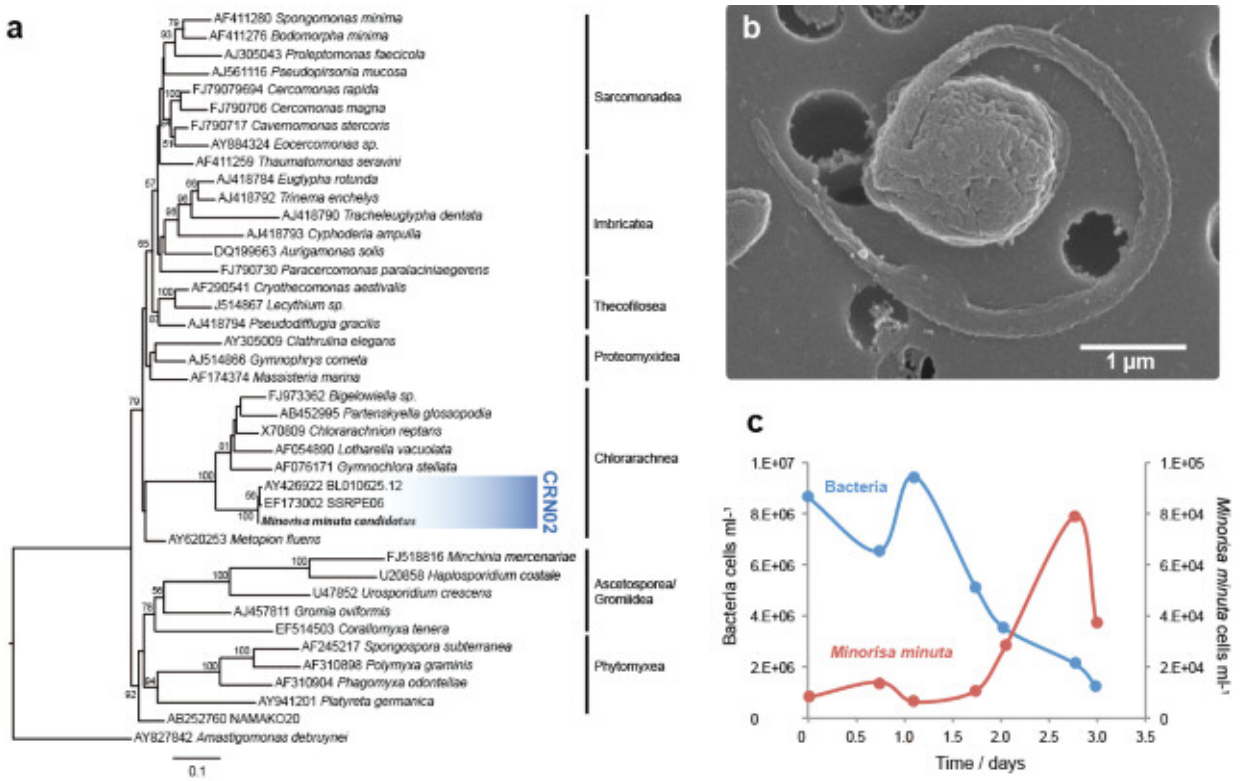


Figure 1. Description of *Minorisa minuta* candidatus. a, Maximum likelihood phylogenetic tree with complete 18S rDNA sequences showing its position within the Cercozoa. The scale bar indicates 0.1 substitutions per position. The coverage of the specific probe CRN02 is shown in blue. **b,** SEM image of a cell of 1.6 µm in size and possessing a single flagellum. **c,** Growth of *M. minuta* candidatus with natural bacteria as prey in a batch culture.

cultures), a choanoflagellate, a chrysophyte, two stramenopiles, and a MAST (Supplementary Table 1). After a second step of single cell inoculation from the latter 8 flasks, some pre-cultures were lost and others evolved to different species. The two resulting, clonal and stable cultures included a rhizaria related to chlorarachniophytes (Fig. 1a) and a stramenopile distantly related to *Developayella* sp. (Supplementary Fig. 2). Single cell sorting by flow cytometry was carried out to avoid time-consuming serial dilution steps. Digestive vacuoles of heterotrophic protists were stained using the vital stain LysoTracker[20] and cells were sorted based on their green fluorescence. From the 400 wells inoculated with single cells, growth was detected in 5 wells (1.25% success rate). All of them were identified as the same rhizarian lineage obtained by the dilution method. Scanning electron microscopy performed on the two stable cultures revealed extremely small cells with little morphological features (Fig. 1b and Supplementary Fig. 3). Rhizarian cells have one flagellum and measure 0.8-2 µm in width (mean of 1.3 µm) and 1.0–2.1 µm in length (mean of 1.5 µm) whereas *Developayella*-like cells have two flagella (one with hairs) and measure 1.2-3 µm in width (mean of 2.1 µm) and 1.5-3.0 µm in length (mean of 2.3 µm).

The rhizarian isolate was distant to any described organism, its 18S rDNA sequence being only 90.6% similar to *Chlorarachnion reptans*. Its basal position within chlorarachniophytes (Fig. 1a) together with its obvious heterotrophic nature (growth in the dark on bacterial food source and lack of observable chloroplasts) suggests it represents an ancient lineage, presumably related to the one that underwent a secondary endosymbiosis event with green algae. The 18S rDNA sequence of our isolate is highly similar to environmental sequences retrieved from the Mediterranean Sea (BL000921.31 and BL010625.12[10]), the Sargasso Sea (SSRPE06[21]), and the English Channel (RA070625T.047[22]). Pending a formal description, this small heterotrophic flagellate has been named *Minorisa minuta* candidatus.

Physiological properties of *M. minuta* candidatus have been assessed from our culture. This flagellate grew relatively fast on natural bacteria (doubling time of 10.6 hours) and reduced bacterial abundances from 10^7 cells mL^{-1} to around 10^6 cells mL^{-1} (Fig. 1c). This growth rate is double of that observed in MAST in unamended incubations[23], and similar to or lower than maximal growth rates observed for other cultured heterotrophic flagellates[24]. The estimated growth efficiency (bacterial biomass converted to protist biomass) of *M. minuta* candidatus in this batch culture was 35%, and its grazing rate was 7 bacteria flagellate^{-1} h^{-1}, again within the range of known heterotrophic flagellates strains. The functional response of *M. minuta* candidatus yields a half-saturation constant much lower than that of other cultured flagellates (below 10^6 cells mL^{-1}; R. Rodríguez-Martínez, unpublished results), suggesting that it is adapted to live at the usual bacterioplankton concentrations in oligotrophic waters (5-10 x 10^5 cells mL^{-1}).

M. minuta candidatus is a significant component of marine heterotrophic flagellates on a global scale, being

both widely distributed and abundant (Fig. 2a). Counts provided by TSA-FISH (Tyramide Signal Amplification-Fluorescent *in situ* Hybridization) using a newly designed specific oligonucleotide probe reveal abundances up to 60 cells mL^{-1} (17 cells mL^{-1} on average) in the Atlantic, Pacific, Indian, and Southern Oceans and the Mediterranean Sea. It accounts for 1.8% of heterotrophic flagellates in these samples, a value that increases up to 5% when considering coastal sites only (Supplementary Table 2). In the same set of samples considered for the present study, *M. minuta* candidatus follows in abundance the uncultured MAST-4, MAST-1C and MAST-1B[19], which is remarkable given that its probe is clearly more specific than MAST probes (roughly species versus family level). *M. minuta* candidatus was detected all year long in a coastal oligotrophic station in the NW Mediterranean Sea (Fig. 2b), ranging from 12 to 120 cells mL^{-1} (52 cells mL^{-1} on average) and accounting for 1 to 12% of heterotrophic flagellates (5% on average). Its abundance was well correlated with bacterial cells number (R^2=0.64; n=11) (Supplementary Fig. 4). Sizing *M. minuta* candidatus cells in natural marine assemblages using microscopy (TSA-FISH) confirms its picoeukaryotic character with cell size varying from 1 to 3 µm (Supplementary Fig. 5).

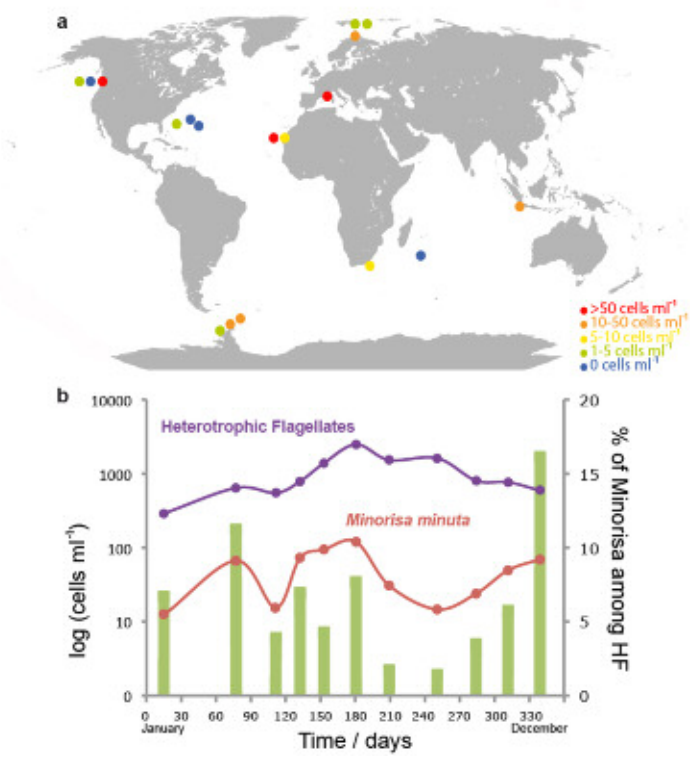


Figure 2. Global marine distribution of *Minorisa minuta* candidatus. a, Abundances at various sampling sites as estimated by specific TSA-FISH counts. **b,** Abundances during a temporal study at the Blanes Bay Microbial Observatory (TSA-FISH counts; year 2007) together with the abundance of heterotrophic flagellates estimated by epifluorescence. Bars represent the contribution of *M. minuta* candidatus cells to heterotrophic flagellates.

Through the culturing approach developed here we isolated several small protists belonging to previously uncultured taxa and from distant lineages within the eukaryotic tree of life (Fig. 1a, Supplementary Fig. 2). When applied at different temporal and spatial scales, this strategy will potentially give access to a wealth of heterotrophic protists in culture. The tiny uniflagellated *M. minuta* candidatus stands up as the smallest bacterivore known so far. Moreover, it represents the only heterotrophic representative within the chlorarachniophyte lineage and is of primary interest to study the transition to secondary plastid endosymbiosis. As for its photosynthetic counterparts in the oceans, the genome analysis of *M. minuta* candidatus will certainly reveal unprecedented cellular, biochemical, and evolutionary pathways[25].

We have shown here that *M. minuta* candidatus has a worldwide marine distribution and is a significant member of heterotrophic flagellate assemblages, particularly in coastal waters. The physiological properties of *M. minuta* candidatus can explain its ecological success and set this species as a good model for dominant marine heterotrophic flagellates, whose parameters could be used to improve ecological models. Getting the environmentally relevant bacteria *Pelagibacter ubique* candidatus in culture[11] led to a leap forward towards a better understating of microbes' function in the oceans and opened up several research directions[26]. Taming small marine predators with ecological relevance holds promise for similar future discoveries.

METHODS SUMMARY

Samples were collected at the Blanes Bay Microbial Observatory on the 25th of September 2007. Forty liters were filtered by gravity through 3 µm and incubated at 20°C in the dark[22]. Seawater from *in situ* and incubated (2 days) was filtered by 0.45 µm and bacteria were concentrated by tangential flow filtration. Bacterial concentrates were quantified by epifluorescence (typically ca 10^8 cells mL^{-1}) and kept frozen. Multi-well plates of 0.1 or 1 mL were filled with aged seawater (3 µm filtered seawater kept in the dark for 2 months, 0.2 µm filtered and autoclaved before use), and an aliquot of the concentrate was added to 5 x 10^6 bacteria mL^{-1}. Heterotrophic flagellates from the same samples were counted by epifluorescence and diluted to add one cell per well. Plates were incubated at 20°C in the dark and inspected for flagellate growth every three days by inverted microscopy. Seawater from a second sampling (30 September 2008) was filtered by 3 µm, and sent to Bigelow Laboratory for Ocean Sciences for cell sorting in a MoFlo™ Flow Cytometer (Dako-Cytomation)[19]. Multi-well plates with a single cell per well were sent back to the Institut de Ciències del Mar. Wells with observed growth were scaled to 30 mL to establish a pre-culture and later a stable culture. Cultures were maintained in 50 mL flasks and transferred every 2 to 4 weeks to fresh media at 1/10 dilution.

For molecular analysis, the whole culture was filtered on 0.6 µm polycarbonate filters, DNA was extracted by standard procedures and PCR was performed with

universal eukaryotic 18S rDNA primers[27]. Sequencing and phylogenetic analyses are detailed in del Campo and Massana (2011)[28], and Scanning Electron Microscopy was performed as in Garcés et al. (2006)[29]. An oligonucleotide probe specific for *M. minuta* candidatus was designed (coverage in Fig. 1a). The probe (CRN02: 5'-TACTTAGCTCTCAGAACC-3') has at least 2 mismatches with non-target sequences. TSA-FISH was performed as described in Not et al. (2002)[30]. Optimized hybridization conditions were 35°C and 30% formamide in the hybridization buffer.

REFERENCES

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320,** 1034-1039 (2008).
2. Cavalier-Smith, T. & Chao, E. E. Phylogeny and megasystematics of phagotrophic heterokonts (Kingdom Chromista). *J. Mol. Evol.* **62,** 388-420 (2006).
3. Giovannoni, S. J., Foster, R. A., Rappé, M. S. & Epstein, S. New cultivation strategies bring more microbial plankton species into the laboratory. *Oceanography*. **20,** 62-69 (2007).
4. Giovannoni, S. J. & Stingl, U. The importance of culturing bacterioplankton in the 'omics' age. *Nature Rev. Microbiol.* **5,** 820-826 (2007).
5. Raes, J. & Bork, P. Molecular eco-systems biology: towards an understanding of community function. *Nature Rev. Microbiol.* **6,** 683-699 (2008).
6. Pedrós-Alió, C. Marine microbial diversity: can it be determined? *Trends Microbiol.* **14,** 257-63 (2006).
7. Epstein, S. & López-García, P. "Missing" protists: a molecular prospective. *Topics in Biodiversity and Conservation* **8,** 27-42 (2008).
8. Amann, R. I., Ludwig, W. & Schleifer, K.H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59,** 143-169 (1995).
9. Caron, D. A., Davis, P. G. & Sieburth, J. McN. Factors responsible for the differences in cultural estimates and direct microscopical counts of populations of bacterivorous microflagellates. *Microb. Ecol.* **18,** 89-104 (1989).
10. Massana, R., Balagué, V., Guillou, L. & Pedrós-Alió, C. Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol. Ecol.* **50,** 231-243 (2004).
11. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418,** 630-633 (2002).
12. Könneke, M. *et al*. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437,** 543-546 (2005).
13. Vaulot, D., Eikrem, W., Viprey, M. & Moreau, H. The diversity of small eukaryotic phytoplankton (≤ 3µm) in marine ecosystems. *FEMS Microbiol. Ecol.* **32,** 795-820 (2008).
14. Jürgens, K. & Massana, R. in *Microbial Ecology of the Oceans, Second Edition* (ed Kirchman, D. L.) 383-441 (John Wiley & Sons, 2008).
15. Sherr, E. B. & Sherr, B. F. Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* **81,** 293-308 (2002).
16. Pernthaler, J. Predation on prokaryotes in the water column and its ecological implications. *Nature Rev. Microbiol.* **3,** 537-546 (2005).
17. Arndt, H. *et al*. in *The Flagellates: Unity, Diversity and Evolution* (eds Leadbeater, B. S. C. & Green, J. C.) 240-268 (Taylor & Francis, 2000).
18. Vaulot, D., Romari, K. & Not, F. Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends Microbiol.* **10,** 266-267 (2002).
19. Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C. Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ. Microbiol.* **8,** 1515-1522 (2006).
20. Heywood, J. L., Sieracki, M. E., Bellows, W., Poulton, N.J. & Stepanauskas, R. Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J.* doi:10.1038/ismej.2010.155 (2011).
21. Not, F., Gausling, R., Azam, F., Heidelberg, J. F. & Worden, A. Z. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ. Microbiol.* **9,** 1233–1252 (2007).
22. Marie, D., Shi, X. L., Rigaut-Jalabert, F. & Vaulot, D. Use of flow cytometric sorting to better assess the diversity of small photosynthetic eukaryotes in the English Channel. *FEMS Microb. Ecol.* **72,** 165-178 (2010).
23. Massana, R., Guillou, L., Terrado, R., Forn, I. & Pedrós-Alió, C. Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat. Microb. Ecol.* **45,** 171-180 (2006).
24. Boenigk, J., Jost, S., Stoeck, T. & Garstecki, T. Differential thermal adaptation of clonal strains of a protist morphospecies originating from different climatic zones. *Environ. Microbiol.* **9,** 593-602 (2007).
25. Worden, A. Z. *et al*. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324,** 268-72 (2009).
26. Tripp, H. *et al*. SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452,** 741-744 (2008).
27. Díez, B., Pedrós-Alió, C. & Massana, R. 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 (2001).
28. del Campo, J. & Massana, R. Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist*. doi:10.1016/j.protis.2010.10.003 (2011).
29. Garcés, E. *et al*. Characterization of NW Mediterranean *Karlodinium* spp. (Dinophyceae) strains using morphological, molecular, chemical, and physiological methodologies. *J. Phycol.* **42,** 1096-1112 (2006).
30. Not, F., Simon, N., Biegala, I. C. & Vaulot, D. Application of fluorescent *in situ* hybridization coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton composition. *Aquat. Microb. Ecol.* **28,** 157-166 (2002).

SUPPLEMENTARY INFORMATION

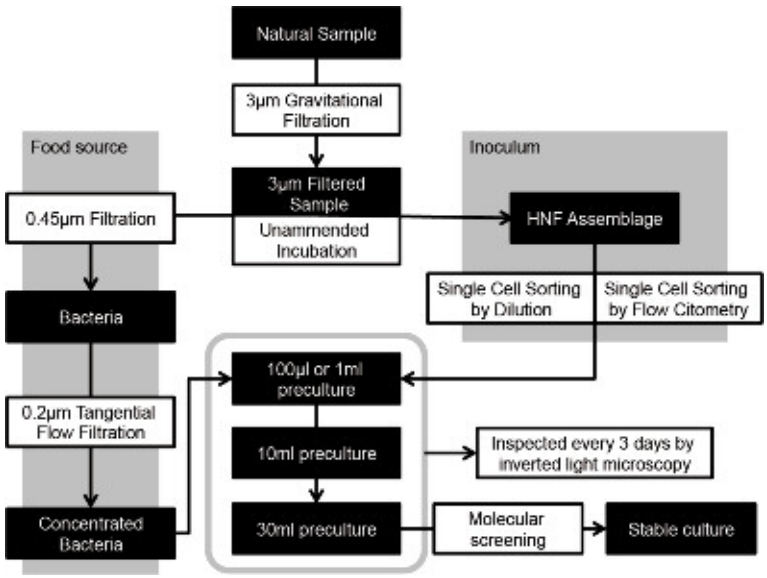


Figure 1. Isolation protocol flow chart.

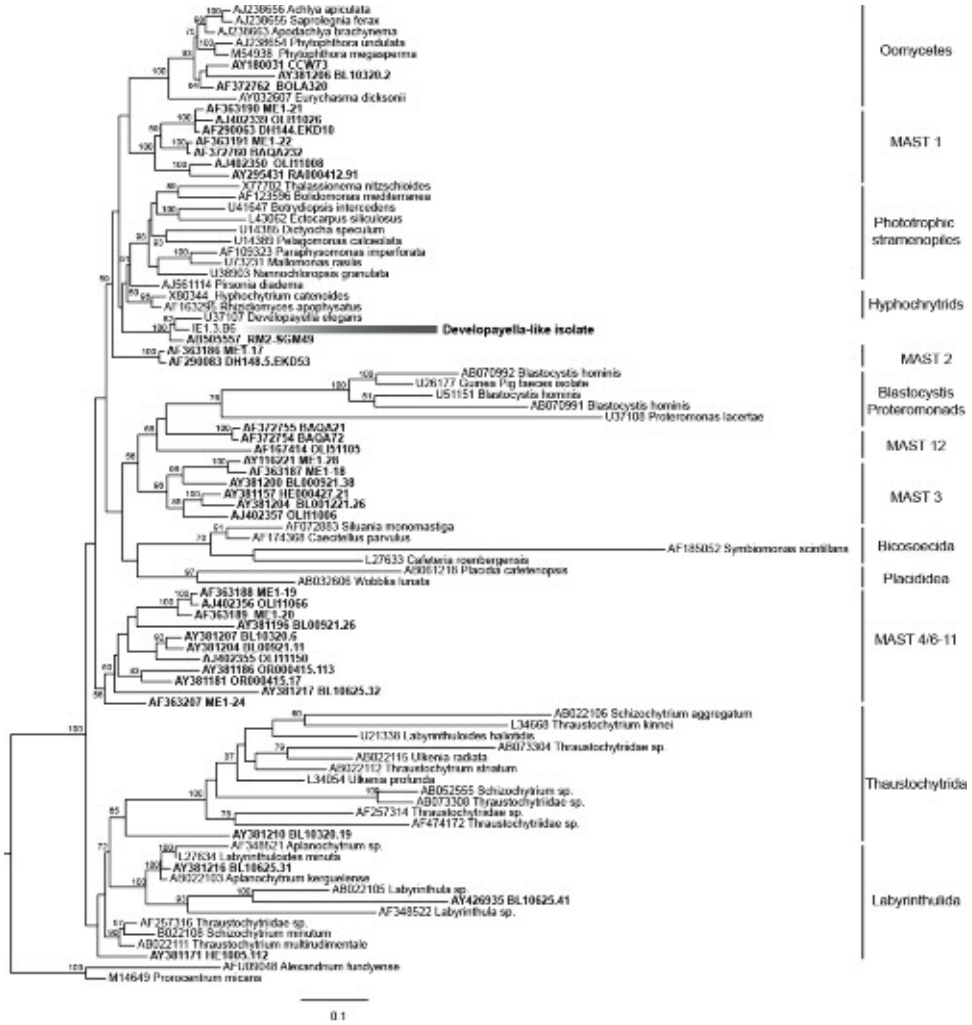


Figure 2. Maximum likelihood phylogenetic tree with complete 18S rDNA sequences showing the position of *Developayella*-like isolate within the stramenopiles. Environmental sequences are in bold. The scale bar indicates 0.1 substitutions per position.

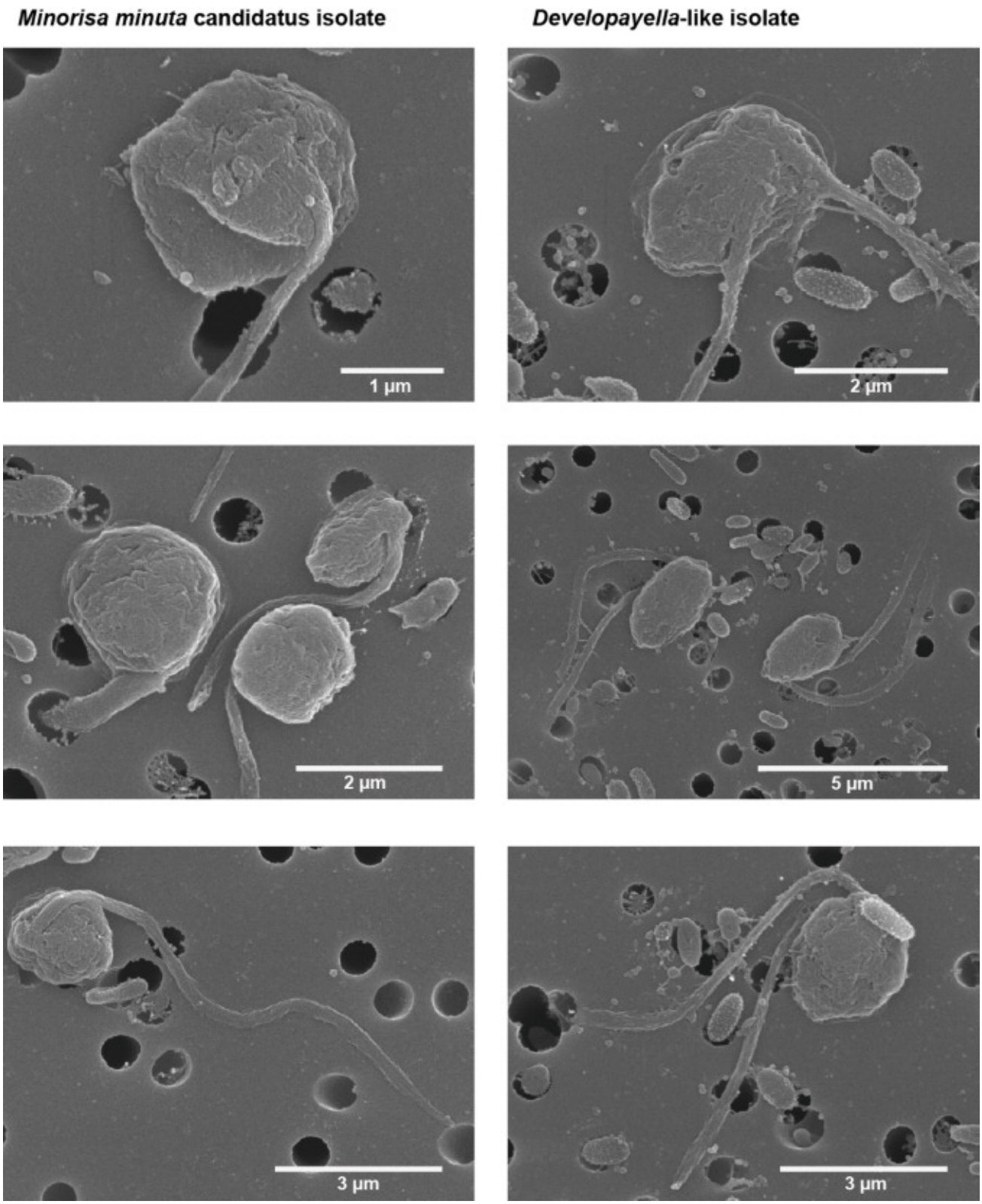


Figure 3. Additional SEM pictures of the two stable cultures. Pictures on left correspond to *Minorisa minuta* candidatus cells and pictures on the right correspond to the *Developayella*-like cells.

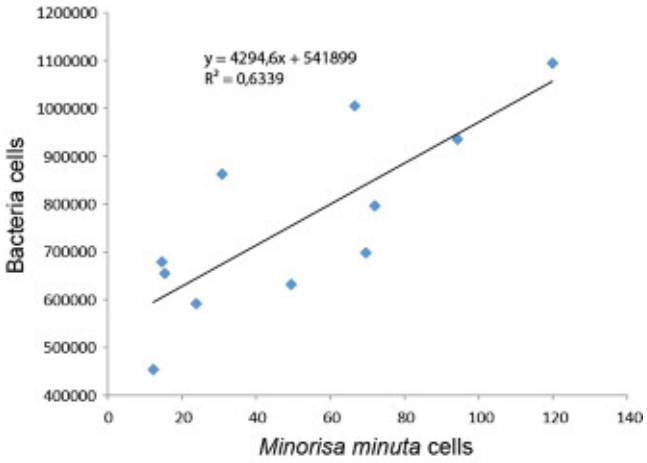


Figure 4. Correlation between the abundances of *Minorisa minuta* candidatus and natural bacteria during the seasonal study (year 2007) at the Blanes Bay Microbial Observatory.

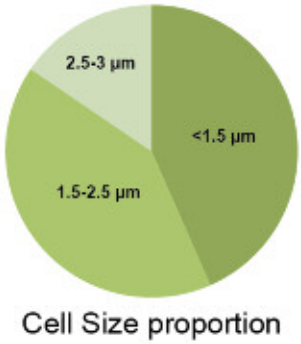


Figure 5. *Minorisa minuta* candidatus cell size distribution in samples from the seasonal study at the Blanes Bay Microbial Observatory (1148 cells measured by TSA-FISH).

Supplementary Table 1. Taxonomic affiliation of the obtained pre-cultures and stable pure cultures.

	Closest match	%	Lineage	Closest cultured match	%
Pre-cultures					
• Rhizaria					
IE1.2.D2	SSRPE06	99.8	chlorarachniophyte	*Chlorarachnion reptans*	90.5
IE1.4.B4	SSRPE06	99.9	chlorarachniophyte	*Chlorarachnion reptans*	90.9
IE4.4.A1*	SA2.4G8	96.8	cercozoan	Cercozoa sp. CC-2009d	94.4
• Opisthokonta					
IE1.4.D1	band 4DB38	97.7	choanoflagellate	*Diaphanoeca grandis*	90.7
• Stramenopiles					
IE1.1.A5	*Paraphysomonas foraminifera*	99.4	chrysophyte		
IE1.3.B3	*Paraphysomonas imperforata*	99.4	chrysophyte		
IE1.3.B6	RM2-SGM49	96.6	stramenopile	*Developayella elegans*	
IE1.4.D5**	Biosope_T123.046	95.7	bicosoecida	*Pseudobodo tremulans*	99.6
IE2.4.A6*	CD8.S17	98.8	chrysophyte	Chrysophyceae sp. CCMP2296	91.5
IE3.4.B1	NIF.1E11	97.0	MAST-3	*Bolidomonas pacifica*	88.3
IE4.6.B5	*Paraphysomonas imperforata*	98.7	chrysophyte		
IE4.8.C6	*Paraphysomonas imperforata*	99.9	chrysophyte		
Stable cultures					
• Rhizaria					
BMO.6.C1* **	SSRPE06	99.6	chlorarachniophyte	*Chlorarachnion reptans*	89.7
• Stramenopiles					
IE1.3.B6	*Developayella elegans*	94.4		stramenopile	

* Sequences with ambiguities.

** Lost after one year of being stable.

*** This working culture was isolated by Flow Cytometry.

Supplementary Table 2. Counts of *Minorisa minuta* (cells ml^{-1}) in a set of samples from diverse oceans and comparison with the counts of heterotrophic flagellates (cells ml^{-1}). Data of HF and codes of samples are from Massana et al. (2006a).

System	Date	*Minorisa minuta*	Heterotrophic flagellates	%-*Minorisa minuta*
ATL1	24-aug-02	4	652	0.7
ATL2	27-aug-02	9	1003	0.9
ATL3	29-aug-02	9	584	1.6
ATL4	23-sep-02	54	1302	4.1
ATL5	30-sep-02	13	814	1.6
ATL6	12-jul-04	9	214	4.1
ATL7	15-jul-04	0	443	0.0
ATL8	30-jul-04	0	593	0.0
Pacific Ocean				
PAC1	09-may-02	12	-	-
PAC2	15-may-02	0.0	-	-
PAC3	18-may-02	60	-	-
Indian Ocean				
INO1	16-may-03	33	687	4.9
INO2	23-may-03	0	626	0.0
INO3	06-jun-03	5	719	0.7
Southern Ocean				
ANT1	03-dec-02	7	1562	0.4
ANT2	05-dec-02	4	1668	0.3
ANT3	11-dec-02	2	2367	0.1
Mediterranean Sea				
BL07	2007*	52	1039	5.0

*Average of a whole year samplig.

1. 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

2. del Campo J, Massana R (2011) Emerging Diversity within Chrysophytes, Choanoflagellates and Bicosoecids Based on Molecular Surveys. Protist 162: 435-448

3. Not F, del Campo J, Balagué V, de Vargas C, Massana R. (2009) New Insights into the Diversity of Marine Picoeukaryotes. PLoS ONE 4: e7143

4. del Campo J, Balagué V, Forn I, Lekunberri I, Massana R (2011) Facing Culturing Bias in Marine Heterotrophic Flagellates Through Seawater Enrichment Incubations. Submitted

5. del Campo J, Not F, Forn I, Sieracki ME, Massana R (2011) Taming the smallest predators of the oceans. Submitted