UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Spatial, temporal and behavioral patterns of marine protists

Patrons espaials, temporals i de comportament dels protistes marins

Caterina Rodríguez Giner

Spatial, temporal and behavioral patterns of marine protists

Patrons espaials, temporals i de comportament dels protistes marins

Caterina Rodríguez Giner

Barcelona, 2017

Tesi presentada per a obtenir el títol de Doctora per la
Universitat Politècnica de Catalunya

Programa de doctorat en Ciències del Mar
Departament d'Enginyeria Hidràulica, Marítima i Ambiental

La Doctoranda	El Director	El co-director
Caterina Rodríguez Giner	Dr. Ramon Massana	Dr. Ramiro Logares

Als meus pares Carmen i Ildefons,
a la meva àvia Lola

L’essencial és invisible als ulls

Antoine de Saint-Exupéry

Contents

Resum/ Resumen/ Summary 9

Glossary 13

General Introduction 15

Aims and Outline 31

Chapter 1 Seasonal and long-term community patterns in temperate marine planktonic protists 35

Chapter 2 Global changes in activity and community structure of marine picoeukaryotes through the water column 75

Chapter 3 Chemotactic response of natural protists communities towards various stimuli 107

Chapter 4 Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes 129

General Discussion 159

Conclusions 167

General References 171

Acknowledgements 181

RESUM

Els oceans són ecosistemes dominats per microorganismes. Entre ells, els protistes (organismes eucariotes unicel·lulars) tenen un paper fonamental en diverses funcions imprescindibles per al sistema marí, principalment com a productors primaris i depredadors, i per tant juguen un paper clau en les xarxes tròfiques marines. Degut a la seva importància ecològica, es necessària la caracterització de les comunitats de protistes al llarg del temps i l'espai per tal d'entendre la seva dinàmica així com els seus patrons biogeogràfics. Això ens permetrà inferir els processos que governen l'estructura de les comunitats de protistes. En aquest context, la present tesi aborda la variabilitat de les comunitats naturals de protistes fent èmfasi en la seva distribució i resposta a diferents nivells. Primerament, ens hem centrat en la dimensió temporal, analitzant les dinàmiques de les comunitats de pico- i nanoeucariotes amb la finalitat de caracteritzar les diverses estratègies utilitzades pels diferents membres de la comunitat, així com també predir la recurrència dels patrons observats. S'ha identificat que, en conjunt, la comunitat de protistes segueix un patró anual, i també s'ha trobat un patró semblant per a algunes de les espècies, generalment les més abundants. Posteriorment, hem adreçat la dimensió espaial, avaluant com canvien les comunitats i l'activitat dels grups taxonòmics al llarg de la columna d'aigua en 13 estacions repartides pels diferents oceans de tot el món. S'ha observat que la comunitat està marcada per l'estratificació vertical i, a més, que la capa mesopelàgica és la regió on la majoria dels grups taxonòmics semblen estar més metabòlicament actius. Un altre nivell d'estudi d'aquesta tesi ha estat la resposta dels protistes davant la presència de senyals químics. L'oceà és un ecosistema complex on els nutrients estan distribuïts de manera heterogènia en diferents àrees, i això provoca l'existència de gradients químics que poden desencadenar diverses respostes dels microorganismes. Per tal d'identificar les respostes individuals dels protistes es van realitzar experiments de quimiotaxi. Els resultats experimentals van demostrar l'existencia d'una preferència d'alguns atractants químics com els exudats de bacteris, envers uns altres. Finalment, degut a que la majoria d'estudis de diversitat microbiana estan basats en abundàncies relatives de les diferents espècies presents a la mostra, s'han relacionat aquestes abundàncies amb les abundàncies reals. En resum, aquesta tesis profunditza en els patrons temporals i espaials de les comunitats de protistes, així com en les preferències quimiotàctiques dels diferents membres, contribuint a ampliar el nostre coneixement sobre els processos que estructuren les comunitats de protistes en la seva dimensió temporal, espaial i conductual.

RESUMEN

Los océanos son ecosistemas dominados por microorganismos. Entre estos, los protistas (organismos eucariotas unicelulares) tienen un papel fundamental en diversas funciones imprescindibles para el ecosistema, principalmente como productores primarios y depredadores, pero también tienen un papel clave en la red trófica microbiana. Debido a su importancia ecológica, es necesaria la caracterización de las comunidades de protistas a través de gradientes temporales y espaciales, para comprender su dinámica y distribución biogeográfica. Esto nos permitirá inferir los procesos que determinan la estructura de las comunidades de protistas. En dicho contexto, la presente tesis doctoral se centra en el estudio de las comunidades naturales de protistas con el objetivo de aumentar nuestro conocimiento sobre su distribución y respuesta a distintos estímulos ambientales. Primeramente, nos hemos centrado en la dimensión temporal, analizando las dinámicas de las comunidades de pico- y nanoeucariotas con la finalidad de caracterizar las diversas estrategias usadas por los distintos miembros de la comunidad. Se ha identificado que en conjunto, la comunidad de protistas sigue un patrón anual, y que diversas especies muestran un patrón estacional. Posteriormente, hemos estudiado la dimensión espacial, evaluando los cambios en la comunidad y en la actividad de los distintos grupos taxonómicos a lo largo de la columna de agua en 13 estaciones distribuidas en los distintos océanos de todo el mundo. Nuestros resultados muestran que la comunidad tiene una marcada segregación vertical, y además la capa mesopelágica es la región con mayor actividad metabólica. Otro nivel de estudio de esta tesis ha sido la respuesta de los protistas a distintos gradientes químicos. El océano es un ecosistema complejo, donde los nutrientes están distribuidos de manera heterogénea en distintas áreas, cosa que provoca la existencia de gradientes químicos que dan lugar a distintas respuestas de los distintos microorganismos. Con la finalidad de identificar respuestas comportamentales de los protistas, se realizaron experimentos de quimiotaxis. Los resultados experimentales demostraron preferencia por algunos atractantes, principalmente hubo una respuesta significativa hacia los exudados de bacterias. Finalmente, debido a que la mayoría de estudios de diversidad microbiana están basados en abundancias relativas de las diferentes especies presentes en la muestra, se relacionaron dichas abundancias con el número total de células presentes en la misma muestra. En resumen esta tesis caracteriza patrones temporales y espaciales de las comunidades de protistas marinos, así como también las preferencias quimiotácticas de los distintos miembros de la comunidad, ampliando nuestro conocimiento sobre los procesos que estructuran las comunidades de protistas a través de la dimensión temporal, espacial y conductual.

SUMMARY

The oceans are microbial-dominated ecosystems, where protists (single-celled eukaryotes) play fundamental roles performing multiple functions as primary producers, consumers, decomposers as well as trophic linkers in aquatic food webs. Due to their ecological relevance, it is important to characterize protist communities along temporal and spatial gradients to understand their biogeography and dynamics. This will allow us to infer the processes that drive the assembly of protist communities. In this dissertation, we focused in the study of natural protists communities aiming to understand their distributions and responses within different dimensions. We first focused in the temporal dimension, analyzing the dynamics of pico- and nanoeukaryotic communities, aiming to characterize the diverse strategies of their members and to determine seasonality. We identified an annual seasonal pattern in the protist community as well as in several composing taxa. Then, we addressed the spatial dimension, analyzing the changes in community composition along the water column in 13 stations distributed in the global ocean, exploring also the vertical variation in the relative metabolic activity of different taxa. Our results showed a clear vertical stratification of the community and indicated that the mesopelagic layer is the region where most taxonomic groups were more metabolically active. Another aim of this thesis was to explore the response of natural protists to chemical signals. The ocean is a complex ecosystem with nutrients heterogeneously distributed along several patches, so there may be a lot of chemical gradients promoting diverse responses on marine microbes. For that purpose, we performed chemotactic experiments to identify individual responses towards different attractants within a natural protist community. The experimental results showed a preferential response towards some of the tested chemical cues, mainly bacterial exudates. Finally, since studies on microbial diversity generally use the relative abundances of phylotypes in a given sample, we explored the relationship between the latter relative abundances to cell abundances in several chosen taxa. In sum, this dissertation determines temporal and spatial patterns in protists communities as well as chemotactic preferences in different taxa, contributing to broaden our understanding of the structuring processes operating across temporal, spatial and behavioral dimensions in the protist world.

GLOSSARY

Alpha-diversity: Diversity present at a single site, usually expressed as the number of species (richness) or an index (such as Shannon's) that incorporates also the eveness.

Autotroph: Organism that synthetizes organic carbon from the fixation of inorganic carbon, generally by photo- or chemosynthesis (i.e using the energy of light or of inorganic chemical reactions).

Beta-diversity: Differences in community composition (species) between different sites.

BLAST: Basic Local Alignment and Search Tool. A computer program for finding sequences in a database which are similar to a query sequence. It is the most widely used search tool.

Chemotaxis: The response of an organism driven by the detection of a chemical gradient. It can be positive, when the response elucidates a movement towards the chemical cue, or negative, if the movement goes to the opposite direction of the stimulus.

Eutrophic: Aquatic environment with high concentration of nutrients and minerals.

Heterotroph: Organism that bases the acquisition of metabolic energy and carbon by the consumption of living or dead organic matter (particulate or dissolved organic matter).

Mixotroph: Organism that is part autotrophic and part heterotrophic.

Oligotrophic: Aquatic environment that is poor in nutrients and primary production.

OTU: Operational Taxonomic Unit. This is generally used as a pragmatical proxy of species. It contains sequences that share a given level of similarity. In eukaryotes usually 99% similarity is used.

Primary production: Fixation of CO_2 to produce organic matter using light energy.

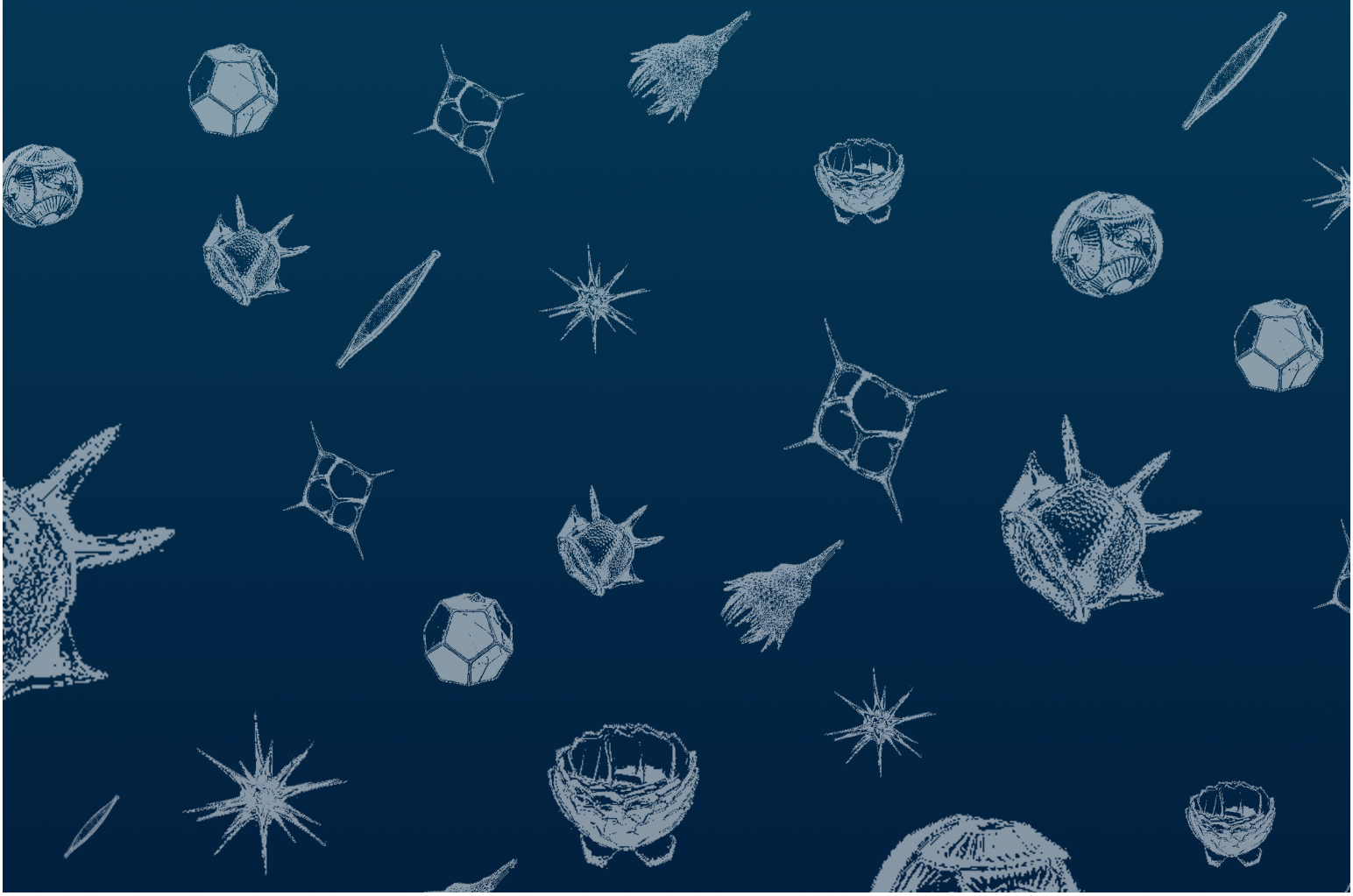
Protists: Term used for referring to single-celled eukaryotes. It includes all eukaryotes that are not plants, animals or fungi, so it does not have a proper taxonomic meaning.

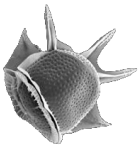
Rare biosphere: Taxa present at low relative abundance in a natural ecosystem.

Richness: The total number of species present in a community.

18S rDNA: Gene encoding the RNA of the small subunit of the ribosome. This gene is found in all eukaryotes, often in many copies per genome, and is widely used as marker to identify eukaryotic microorganisms.

General Introduction





THE MARINE ECOSYSTEM AND THE PLANKTON, AN OVERVIEW

The marine environment is the largest contiguous habitat on Earth. Ocean waters cover approximately 70% of the Earth's surface and play a relevant role in the functioning of the planet, having a significant influence on the Earth's climate, especially due to the ocean circulation. The thermohaline circulation redistributes heat and chemical elements from the tropics to the Polar Regions, exerting a key influence on many biogeochemical cycles. Besides, the oceans contain different habitats with remarkable changes in depth, light or temperature. The deep ocean and sea bottom, where no light arrives, is clearly different from the epipelagic ocean (<200 m depth) where there is penetration of sunlight. Furthermore, marine habitats closer to land are influenced by the input of terrigenous materials (such as sediments, freshwater, organic carbon and nutrients), which increase the habitat variability within these areas and with respect to the open ocean. On the contrary, the bathypelagic ocean (>1000 m depth) is characterized by a relatively constant physical and chemical environment (Arístegui *et al.* 2009). In particular, the bathypelagic zone contains variable, but non-limiting, concentrations of nutrients and oxygen. Thus, despite we normally consider the ocean as a single ecosystem, we have to take into account that it encompasses a wide spectrum of living conditions.

The oceans harbor an enormous diversity of organisms, in particular microorganisms, that are called the '*unseen majority*', comprising 10^{29} prokaryotic cells and 10^{30} virus (Whitman *et al.* 1998; Suttle 2007). Marine microorganisms, including Bacteria, Archaea and Protist, inhabit all marine ecosystems, from the tropics to the sea ice, and from the surface waters to the deep abyssal depths. They carry out key functions in the ocean, such as transducing solar energy and catalyzing biogeochemical transformations of nutrients, being a crucial link in the ocean's carbon cycle (Karl 2007). Thus, marine microbes are the major drivers of biogeochemical processes in the oceans and crucial components of marine ecosystems (Sherr & Sherr 2000; Azam & Malfatti 2007; Falkowski *et al.* 2008). However it has been only during the last 40 years that we started to appreciate the importance of these microscopic forms in the oceans.

Is in the plankton (from the Greek '*planktos*' meaning '*errant*' or '*drifter*') where the importance of microbial life is more apparent. The plankton contains diverse organisms that live in the water column that are not capable to swim against a current, which includes bacteria, archaea, and eukaryotic organisms like algae, protozoa or jellyfish. The size range of plankton comprises about 9 orders of magnitude (0.02 μm to 2 m), yet most organisms are microscopic. Planktonic microbes are usually categorized in three different classes according to their size: picoplankton

(from 0.2-2μm), nanoplankton (from 2-20μm) and microplankton (from 20-200 μm). The two larger size classes are mostly composed by eukaryotes, whereas the picoplankton was originally thought to be composed of prokaryotes, but the existence and abundance of small eukaryotes that fits the picoplankton size was soon detected and the definition was modified to include bacteria, archaea and eukaryotes (Johnson & Sieburth 1982).

Plankton can also be divided into two main functional groups: (1) the phytoplankton (from the Greek terms '*phyton*' or plant and '*planktos*'), which comprise autotrophic organisms that use solar energy to convert CO_2 to organic carbon through the photosynthesis, and (2) heterotrophic organisms that live on previously synthesized organic matter, which they take by grazing, osmotrophy or parasitism. Marine phytoplankton is composed by a variety of organisms including Cyanobacteria (prokaryotic organisms) and eukaryotic algae that range from less than 2 μm to more than 100 μm (Johnson & Sieburth 1982). Yet, it is generally dominated by the smaller forms. Besides, marine phytoplankton, despite representing <1% of the photosynthetic biomass of Earth, contributes to almost half of the net Earth's primary production (Field *et al.* 1998; Falkowski 2012). Specifically, the picoplankton is the size-range that contributes more significantly to primary production, with estimates ranging from 35 to 73% of the total production (Li 1994; Partensky *et al.* 1996).

MARINE PROTISTS, KEY PLAYERS IN THE OCEANS

Protists were first described by Anton van Leeuwenhoek in the 17th century. The term protist refers to single-celled eukaryotic organisms, although many of them could form colonies that exhibit coordinated behavior (Caron *et al.* 2009). They account for the majority of eukaryotic diversity and are present in all branches of the eukaryotic tree of life (Baldauf 2003). Albeit most of them are microscopic, they can span more than five orders of magnitude in size and display multiple morphologies (Fig. 1). Based on their size, they are also classified in the three categories explained before: picoeukaryotes (0.2-2 μm), nanoeukaryotes (2-20 μm) and microeukaryotes (20-200 μm), although the picoeukaryotes often extends to organisms passing through a 3 μm pore size filter (Massana 2011). For instance, the smallest eukaryote known to date, *Ostreococcus tauri,* measures only 0.8 μm (Chrétiennot-Dinet *et al.* 1995) being smaller than some bacteria.

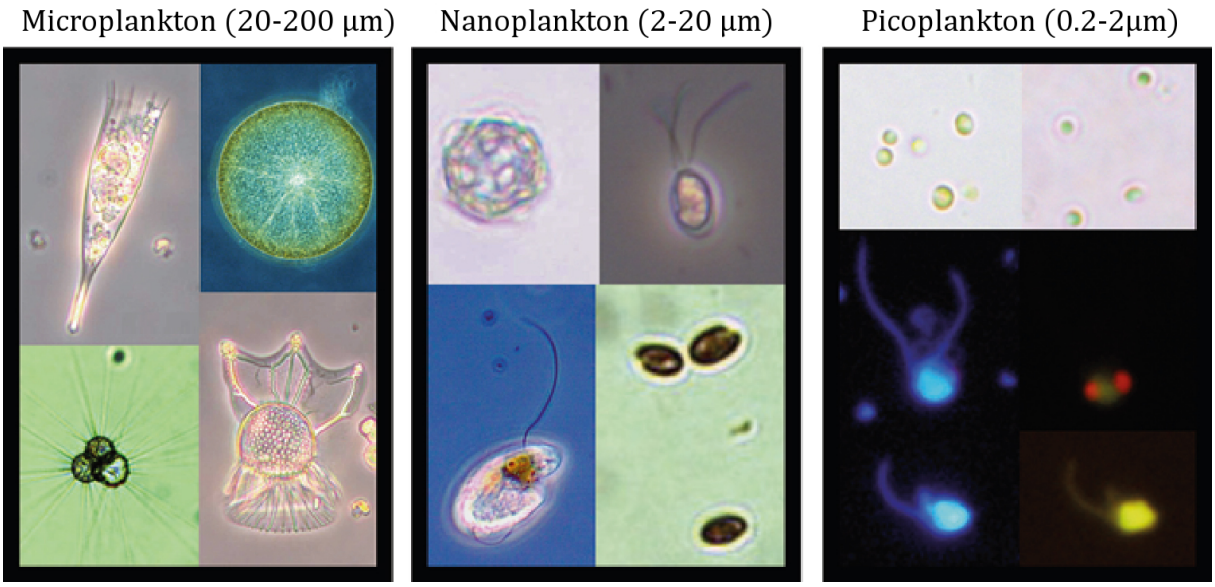


Fig. 1. Microscopy pictures of marine protists according to their size ranges. From Massana, 2015.

In marine environments, protists play a wide range of key ecological roles, such as primary producers, predators, osmotrophs, thus contributing to maintain biogeochemical cycles. It is interesting to focus in the last point, as protists are trophic linkers in marine food webs. In 1974 Pomeroy described the marine food web (Pomeroy 1974) and few years later, in 1983, Azam (Azam *et al.* 1983) presented the *microbial loop* (Fig. 2) which highlights the importance of bacteria as they reabsorb dissolved organic carbon (DOM) from the environment and reintroduce it into the classic food chain when they are grazed by heterotrophic flagellates. In the microbial loop, phytoplanktonic cells largely carry out primary production, whereas heterotrophic cells play a key role in the marine food webs consuming primary producers or their byproducts, and at the same time they are preyed by larger zooplankton. Heterotrophic nanoflagellates (mostly ranging between 2-5µm) are the main bacterivores of the microbial food web (Sherr & Sherr 2002).

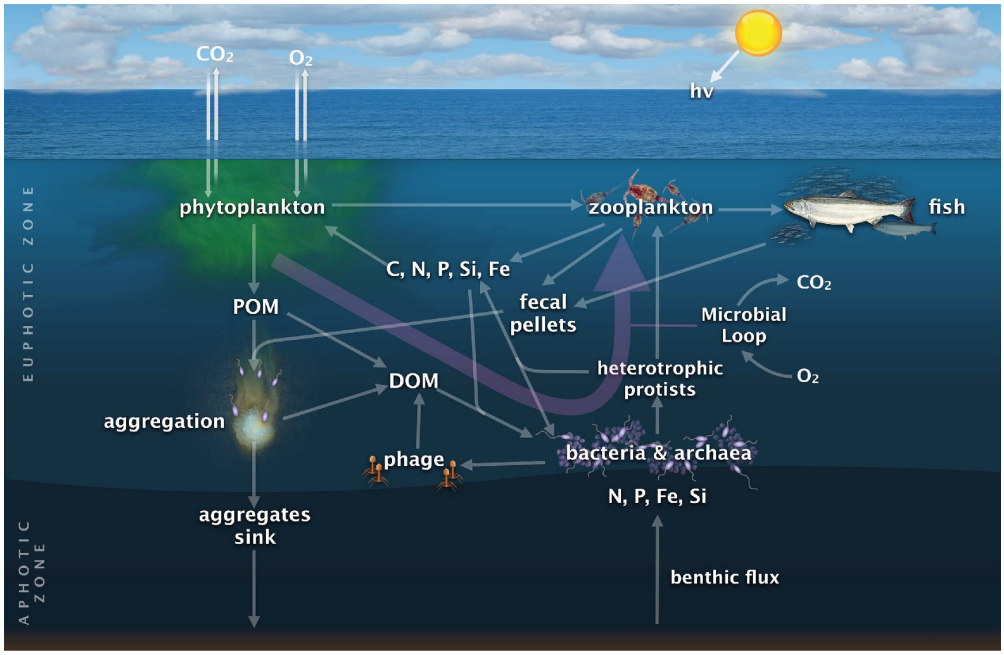


Fig. 2. Schematic representation of marine microbial food webs. The microbial loop is represented by a purple arrow. DOM (Dissolved Organic Matter), POM (Particulated Organic Matter). From Worden *et al.* 2015.

Protist bacterivory and the detection of chemical signals

Predation in aquatic microbial food web is dominated by phagotrophic protists. For phagotrophy we operationally consider protists that ingest living prey as a part of their source of energy. This function is performed by strictly heterotrophic taxa but also by mixotrophic taxa, which base their nutrition by a combination of phagotrophy and photosynthesis (Sanders 1991; Jones 2000). Mixotrophic species are widespread along the major taxonomic lineages e.g. within the chrysophytes, dinoflagellates, haptophytes, euglenophytes, and cryptomonads. The contribution of mixotrophic flagellates to bacterivory in marine sistems was shown to be larger than expected (Sanders *et al.* 2000; Unrein *et al.* 2007; Zubkov & Tarran 2008), accounting for about half the total bacterivory both in coastal (Unrein *et al*. 2007, Fig. 3) and in offshore systems (Zubkov & Tarran 2008).

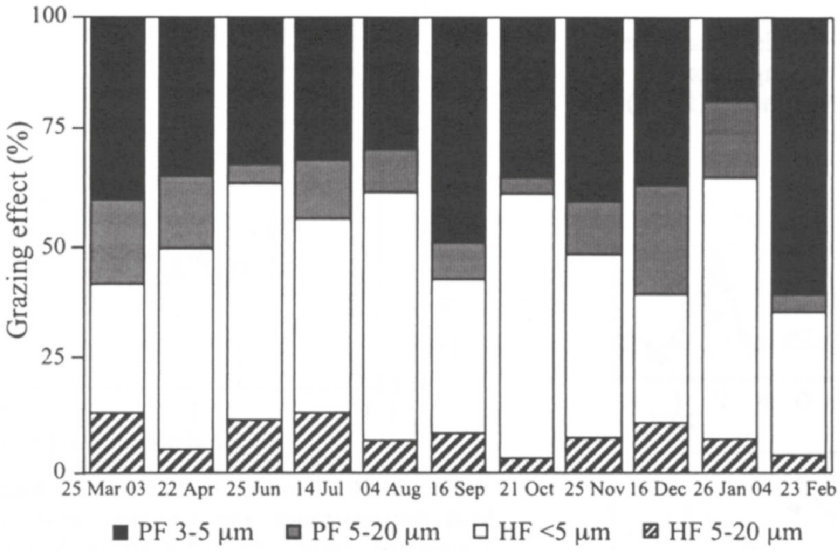


Fig. 3. Contribution of heterotrophic (HF) and mixotrophic (PF) flagellates, each separated into two size classes, to total bacterivory in Blanes Bay during a seasonal cycle. From Unrein et al. (2007).

It is known that protists have a selective grazing behavior determined by prey size, prey motility, cell surface characteristics, biogeochemical composition, and the release of dissolved chemical cues (Jürgens & Massana 2008; Montagnes *et al.* 2008). This strong selection makes prey populations strongly influenced by protistant grazers, thus protists may influence and control the size, taxonomic composition and morphology of prey assemblages (Jürgens & Matz 2002; Sherr & Sherr 2002; Montagnes *et al.* 2008). However the encounter between bacteria and protists may depend on the chemical comunication between prey and predator. It is known that there are chemical interactions between them, for instance through bacterial exudates or other chemical cues, that operate as attractants or repellents in chemoreception prey location for protists. Thus, some phagotrophic protists are able to chemically sense and to accumulate at aggregations of their prey (Fenchel & Blackburn 1999). However, this chemical communication and response to chemical stimuly (chemotaxis) is still poorly studied in protists, in contrast with bacteria (Stocker & Seymour 2012).

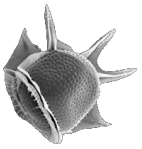
Unraveling protist diversity. A walk through the methodologies

Assessing the species diversity present in the microbial world is more challenging than in the macrobial world, due to a general lack of distinctive morphological characters. Larger protists, usually the ones corresponding to the microplankton, have conspicuous forms that allow

identifying them based on their morphology by microscopic inspection. The most important of them usually belong to: ciliates, dinoflagellates, diatoms or radiolaria. But when moving to the smaller protists, especially the picoeukaryotes, a staining technique and a filtration step is needed. The classical method uses epifluorescence microscopy to observe microbial cells typically stained with DAPI (4,6-diamidino-2-phenylindole), that binds to the DNA of the cells (Porter & Feig 1980). This tool is appropriate to quantify the organisms, but the weak point is that it only gives an idea of the size, broad morphology and general features of the organism as for instance the presence of flagella or chloroplasts. Yet, the organisms cannot be classified into any taxonomic group.

The use of cultures is not a good solution either. Despite culturing is a good approach to obtain model species that can then be used in laboratory experiments to unveil their behavior and functional parameters (e.g. feeding rate or chemotactic responses of the specific taxa), it is well known that many cells do not grow in culture media (del Campo & Massana 2011). For this reason, culturing provides a biased vision of the eukaryotic microbial diversity present in a sample (del Campo *et al.* 2013).

Since approaches independent of culturing and microscopy were needed, little was known about the real diversity of microbes until the application of molecular tools at the beginning of 1990 (Giovannoni *et al.* 1990). These techniques usually use the small subunit ribosomal RNA (SSU rRNA), 18S in eukaryotes and 16S in prokaryotes (Woese & Fox 1977) to determine the diversity present in a sample. The 18S rRNA is part of the small subunit of the ribosome, a complex that is responsible of protein biosynthesis in all living cells. The genes that encode the rRNA are called ribosomal DNA (rDNA) and typically are together forming the rDNA operon. The rDNA operon is located within the cell nucleus, and many copies of this ribosomal gene are distributed in tandem repeats in the genome in a eukaryotic cell. The rRNA is a basic component of all cells (eukaryotic and prokaryotic) and the fact that is universal (i.e. present in all organisms), functionally conserved and highly expressed in ribosomes makes it the most widely used gene in microbial ecology. In addition, ribosomal genes contained conserved and variable regions, sharing a high similarity in their conserved regions among related taxa (used for gene detection), and with large variations in their less conserved regions among different lineages (the hypervariable regions), which are used for phylogeny classification (Woese 1987, Lovejoy *et al.* 2007; Amaral-Zettler *et al.* 2009; Burki 2014). Despite their usefulness, it is not perfect since it is typically a multi-copy gene with copy number varying from 1 to more than 1,000 (Zhu *et al.* 2005; Not *et al.* 2009; Medinger *et al.* 2010), implying that the relative gene abundance of



an organism can deviate from their real abundance obtained by microscopy. In addition, it may well be that not all copies within the same genome are identical.

The first studies sequencing the 18S rDNA from environmental samples revealed a large unknown diversity of eukaryotes and specifically in the picoeukaryotes (Díez *et al.* 2001a; López-García *et al.* 2001; Moon-van der Staay *et al.* 2001). These studies were based on extracting the DNA of all microbial cells contained in the sample, amplify the 18S rDNA genes using PCR (polymerase chain reaction) with general eukaryotic primers and cloning and sequencing the PCR products. This seminal cloning and sequencing approach had a limited sequence output (usually between 100-500 sequences per library) and therefore was insufficient for the total description of picoeukaryotic diversity, as these techniques captured only the most abundant microbes leaving outside the large amount of rare taxa (Pedrós-Alió 2006). Notwithstanding this was revolutionized with the appearance of the 'Next Generation Sequencing'.

During the last 10 years the popularization of High Throughput Sequencing (HTS) has changed the way of studying microbial diversity, increasing the sequencing power and thus allowing a deeper exploration of microbial communities. While with Sanger sequencing around 10^2 sequences were obtained per run, with HTS (454 pyrosequencing or Illumina) 10^6-10^9 sequences per run are generated (Scholz *et al.* 2012). Nevertheless, these technologies are evolving very fast, i.e. 454 had a short life having a peak and almost disappearing during the course of this thesis, and nowadays is Illumina the main sequencing platform used in microbial ecology. But we have to remember that when combining HTS with previous amplification by PCR, we are still subjected to the PCR biases (Wintzingerode *et al.* 1997), and to the possible errors that can be introduced during the PCR and sequencing process, such as chimeric sequences (Berney *et al.* 2004) that need to be identified and removed to not increase artificially the diversity. On the other hand, this huge amount of sequences gives us the possibility of detect and characterize the rare biosphere (Sogin *et al.* 2006; Pedrós-Alió 2012; Logares *et al.* 2014), usually described as phylotypes with an abundance <0.1% (Fig. 4).

The 'rare biosphere' was firstly named by Sogin (Sogin *et al*. 2006), and since then has become very popular in molecular studies. It is based in the fact that all communities are dominated by a few species that are very abundant, accompanied by a large number of species at very low abundance (Fig. 4). It has been proved that these rare communities could present seasonality (Alonso-Sáez *et al.* 2015), respond to environmental change (Campbell *et al.* 2011) and be metabolically active (Logares *et al.* 2015), which indicate that are active members of the

community, but is still unresolved why so many species are rare. However when thinking about the rare biosphere, we have to keep in mind that local rarity (being rare in a sample) does not imply global rarity (be rare always or everywhere) (Logares *et al.* 2015).

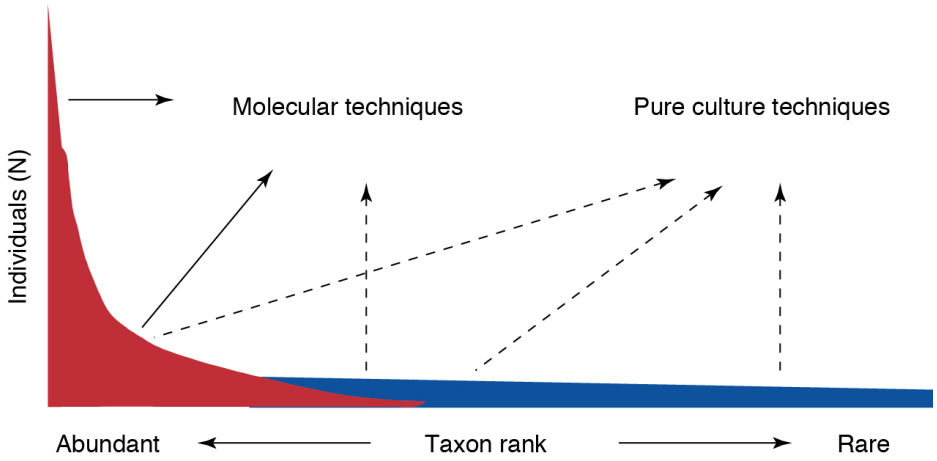


Fig.4. Plots of number of individuals of the different taxa ranked according to their respective abundance. The total curve represents biodiversity and it postulated to be composed of two sections. The red section represent the abundant taxa, an the blue section corresponds to rare taxa, which survive in the ecosystem at low abundance. From Pedrós-Alió 2006.

The molecular survey studies give us a bunch of sequences that increase our knowledge about diversity, but most of the organisms detected have never been seen before and we do not know anything about their morphology or ecology. To address these gaps, FISH (Fluorescent *in-situ* hybridization) allows targeting specific cells by using specific probes. FISH has been and is used as a conformational tool to identify morphologically similar protistan species (Scholin *et al.* 1996), and to provide morphotypes for sequences of uncultured taxa, known only from genetic surveys (Massana *et al.* 2002; Cuvelier *et al.* 2008; Rodríguez-Martínez *et al.* 2009).

Protist diversity in the eukaryotic tree of life

During the last years there have been several proposals for the organization of the eukaryotic tree of life, and many of them are still debated without a clear established consensus (Simpson & Roger 2004; Adl *et al.* 2005; Burki *et al.* 2007; Baldauf 2008). Nowadays the eukaryotic tree of life is divided into five main supergroups (Fig. 5), all of them frequently retrieved in molecular surveys: Opisthokonta, Amoebozoa, Excavata, Archaeplastida (Plantae) and SAR. The SAR is the most recently proposed supergroup and its existence is only supported by molecular data

(Burki 2014). It contains the Stramenopiles (or heterokonts), Alveolates and Rhizaria. The alveolates include two clades well represented in marine ecosystems: dinoflagellates and ciliates. Protists have representatives in all of the supergroups while multicellular plants, animals and fungi are present only in specific branches. Molecular approaches have been central in the rearrangements of the eukaryotic tree of life ocurring during the last 20 years (Burki 2014). In addition, molecular surveys of natural protist diversity have detected new undetected species, groups and entire clades of protists, which have also found a place in the eukaryotic tree of life. A particular case are the MALV (Marine Alveolates) detected in 2001 by López-García (López-García *et al.* 2001) and the MAST (Marine Stramenopiles) defined by Massana *et al.* in 2004. These two main ribogroups are very frequent in marine molecular surveys, being widespread around the world, but there is still a lack of knowledge about them. Those cells are mainly heterotrophic which could explain why they have not been detected before, due to a larger culturing bias among heterotrophs than phototrophs.

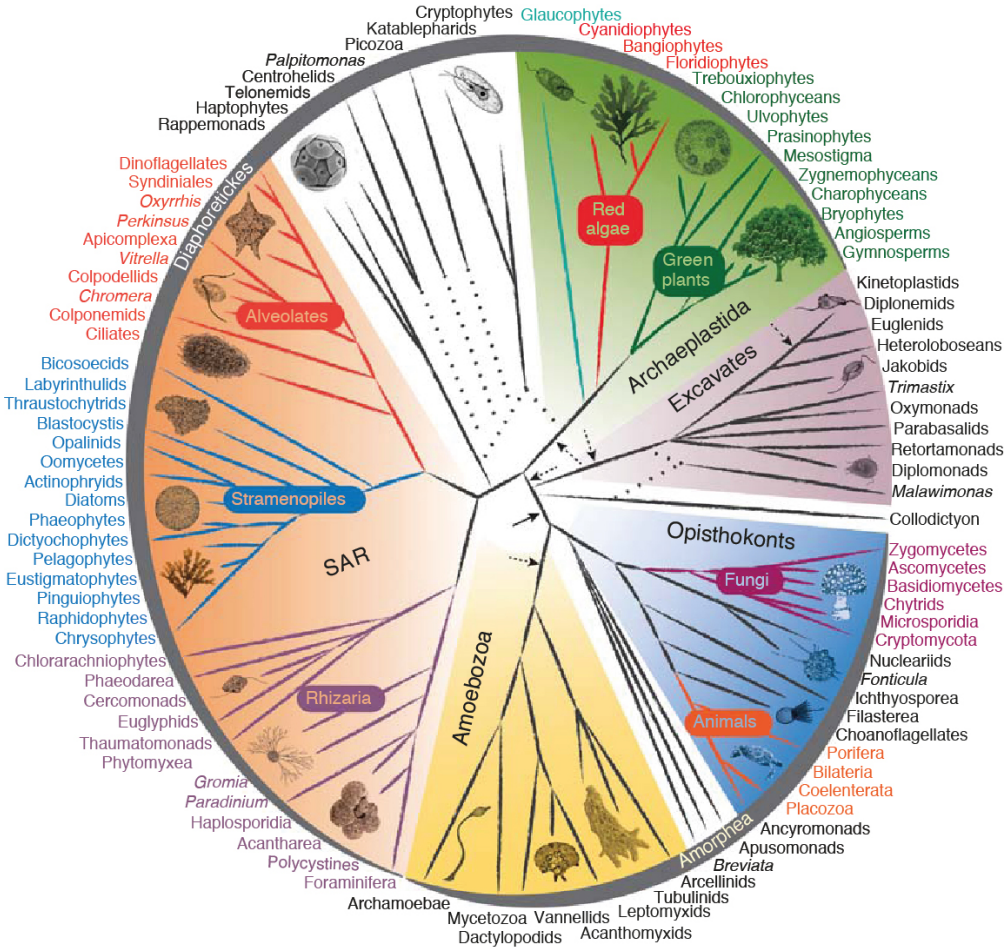


Fig. 5. Global tree of eukaryotes. Cartoons illustrate the diversity constituting the largest assemblages (colored boxes). Dotted lines denote uncertain relationships. From Burki 2014.

21st CENTURY: THE ERA OF LARGE DATASETS

Marine microbial communities are affected by physical, chemical and biological factors that could change in time and space. The differential adaptation to the variability of these factors will determine their dynamics and distribution of microbial communities. The dramatically decrease in the sequencing costs and the large number of sequences obtained has changed the way to study microbial ecology, unlocking the funding that was previously used for sequencing to other purposes. This has allowed a more extensive exploration of the marine environment. When thinking on large datasets we need to differentiate two perspectives or gradients: (1) the spatial variability and (2) the temporal dynamics.

Spatial variability

The physicochemical conditions of the ocean are not constant, being different in the horizontal scale, thus there are marked differences between coastal and offshore waters and between the tropics and poles. Also, there are contrasting patterns along the water column, from surface to the deep ocean. The distribution of organisms is mostly controlled by their dispersion and adaptation to the environment (local adaptation), so knowing these two factors is important to understand the biogeographical patterns of species. However, the majority of the spatial studies done until today only comprise a limited region of the global ocean, i.e. sampling points around a concrete area. Often, the difficulty to access to some areas is the reason behind their under-sampling.

The availability of large datasets covering a wide proportion of the ocean allows us to have a better understanding of the distribution and ecology of the different taxa, and to analyze their dispersion and response to different environments. This gives us a broad picture of the global ocean, though samples are geographically very separated one from the other. The diversity of protists has been studied in different areas such as the Artic (Lovejoy *et al.* 2006), Antarctic (Diez *et al.* 2001b) the Mediterranean Sea (Massana *et al.* 2004a), the coastal Pacific (Worden 2006), the Sargasso Sea (Not *et al.* 2007) and in the Indian Ocean (Not *et al.* 2008) during different years. However, the different sampling methodologies used in these studies difficults a deepful comparison.

Since few years ago, most of the studies published were based only in a few sampling points. However, in 2004 The *Sorcerer II* Global Ocean Sampling (GOS) by Craig Venter was the first

approach to collect samples from the entire world. Notwithstanding, in the recent years two circumnavigations expeditions have been done: The Malaspina expedition from 2010-2011 (Duarte 2015, Fig. 6) and the *Tara-Oceans* (Karsenti *et al.* 2011) from 2010-2011. These expeditions performed a huge sampling effort and achieved a more comprehensive knowledge of microbial diversity by unveiling global surface and deep ocean patterns (de Vargas *et al.* 2015; Sunagawa *et al.* 2015; Pernice *et al.* 2016; Salazar *et al.* 2016). These circumnavigations were focused in the open ocean, whereas also in 2010 the BioMarKs project collected coastal samples from all Europe giving new insights to the protist diversity and distribution in a coastal gradient (Massana *et al.* 2015).

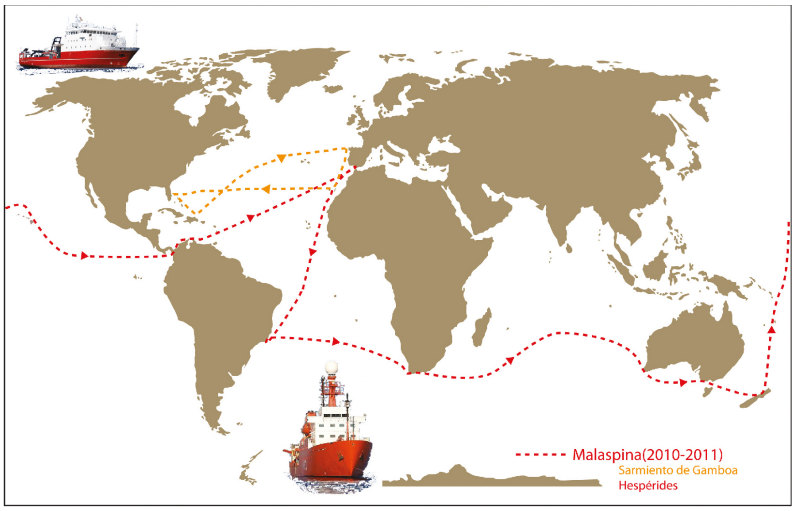


Fig. 6. Tracks of Spanish R/V Sarmiento de Gamboa (upper left corner, in orange), and R/V Hespérides (bottom center, in red) during the Malaspina 2010 Circumnavigation Expedition. From Duarte 2015.

Understanding how microbial communities vary at different spatial scales is important because it allows us to identify diversity hotspots, but also detect correlations with environmental drivers. Furthermore, it is important to have in mind that the spatial distribution gives a frozen picture of which microorganisms are present in a concrete moment under specific environmental conditions. However, we cannot be sure if the sample represents the system at that time-point, or if is an exception to the typical community at that specific moment due to a disturbance event or perturbation, which highlights the importance of the temporal data.

Temporal dynamics

We have to know the temporal dynamics of the system in order to understand it, and also to have the capacity to detect any perturbation. Furthermore, to address ecological trends that occur over periods of several years, such as changes and consequences produced by climate change, long time-series observations are required. Large datasets enable the examination of long-term relationships, which give insights into the principal factors that control microbial dynamics, and it allows to detect noise and exceptions from typical measurements.

It has been observed that microorganisms change over multiple timescales (hours, days, weeks) and in response to different environmental forces (biological and non-biological) that drive changes in microbial community composition (Fuhrman *et al.* 2015). Nowadays there are several long-term time series operating around the world, e.g. BATS (Bermuda Atlantic Time-series), HOT (Hawaii Ocean Time-series), SPOT (San Pedro Ocean Time-series), English Channel, BBMO (Blanes Bay Microbial Observatory) among others (see Fig. 7). Most of them use the monthly sampling interval, which is the most common in long-term time-series studies. When comparing the microbial dynamics observed in the different sites, consistent patterns emerge. For instance, seasonal variation is observed in all the time-series sites but with different strength, which is not surprising as seasonal changes in microbial community composition reflects seasonal changes in the environment, and these are different in the equatorial regions than in the tropics. Besides, more seasonal variation is generally observed in surface than in deep waters (Fuhrman *et al.* 2015).

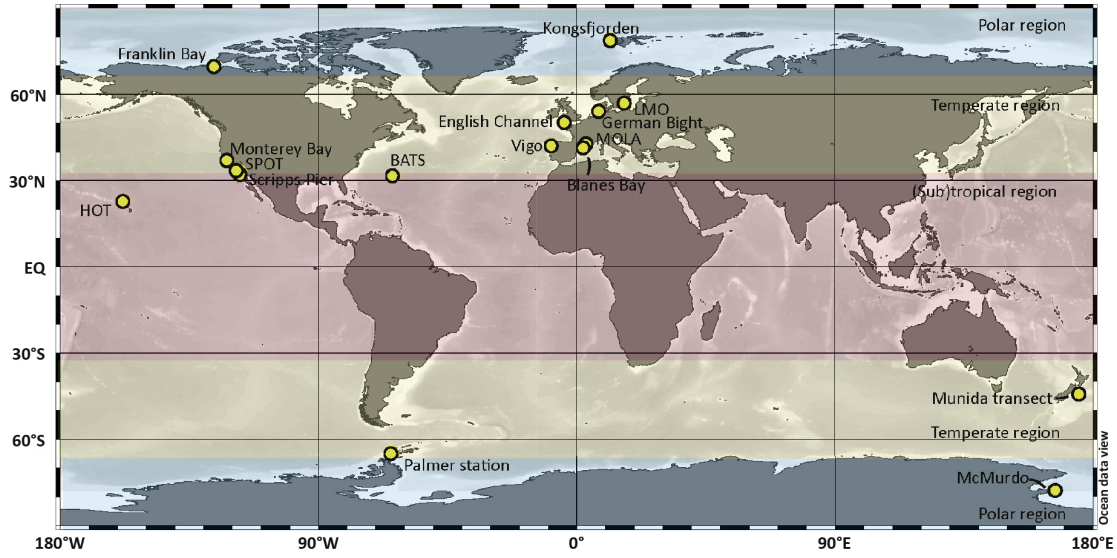
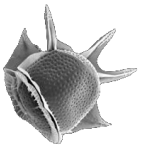


Fig. 7. World map representing long-term microbial sampling stations. Modified from Bunse & Pinhassi 2017.



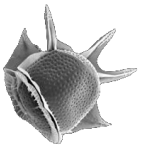
In order to show that there is an annual repeating pattern, data needs to be collected during several years to demonstrate that the pattern is predictably seasonal rather than just changing over months without a repeteability. Furthermore, the predictability of community composition from environmental parameters implies the presence of well-defined niches for the predictable organisms (Furhman 2009). In addition, the study of long time series will tell us if the community composition is stable over time or, on the contrary, it is changing due for instance to loss of species or successful invasions.

Nowadays microbial ecologists are working on the biggest large scale ever done, with good sequencing tools that allows quantifying the relative abundances of the organisms using a reliable methodology. Understanding the changes in the community structure in space and time is needed to figure out factors that control communities, and it is crucial to extrapolate from individual samples to the global scale. To assess the processes that shape microbial distribution is fundamental to study spatial and temporal variation.

FUTURE AND PRESENT

Sequencing technologies are evolving that fast that present and future are very close to each other. Every year new techniques are developed or improved, which makes cheaper the use of techniques previously unaffordable for many laboratories, like the analyses of new genomes, metagenomes or metatranscriptomes. Indeed, these -omics studies are becoming increasingly feasible. Metagenomes consist in the analysis of genomic DNA from a whole community that provides an inventory of the gene pool present in the community. In contrast, metatranscriptomics allows to identify which genes are being transcribed in a given moment, which enable to detect how assemblages respond to perturbations or under a specific environmental condition (Gilbert *et al.* 2008). Transcriptomic data is starting to be available for different protist taxa, and these studies are beginning to elucidate the specific physiological responses of specific protists to environmental cues (Caron *et al.* 2017). However, sometimes the findings and real information that we could obtain and understand from metagenomic or metatrascriptomic analyses do not describe the ecology of the microorganisms in their environment. This is due to the fact that most analyses depend on and exhaustive nd well-curated reference database to know what is the meaning of each environmental sequence (from which species or what is it for). It is mandatory to enrich the reference databases used in order to explore and extract all the information from the data obtained.

On the other hand, the tools that we have nowadays, allow us to move a step forward to unveil the interactions between all types of organisms (Krabberød *et al.* 2017), to study the ecological role of the rare biosphere, and to try to understand the different functional diversity present in a community. The recently applied SAGs (Single Amplified Genomes) are opening new avenues in microbial ecology, giving the possibility to study individual cells and obtain their genomes, which may help to understand ecological interactions (such as grazing or symbiosis and chemotactic response) between different organisms.



AIMS AND OBJECTIVES OF THE THESIS

The main goal of this thesis is to provide new insights to better understand the ecology, distribution and behavior (response) of marine protists. We aim to address this general topic by using data from natural communities deriving mostly from big datasets. To achieve the main goal, this dissertation contains four chapters that are structured in the following three main objectives:

1. Characterize changes in the microeukaryotic community through space and time. Chapters 1 and 2.

2. Assess changes in the chemotactic response of marine protists due to the detection of different chemical signals. Chapter 3.

3. Relate the information provided by molecular surveys with the information obtained by microscopy. Chapter 4.

Each chapter is structured as a scientific papers, some already published or submitted to the journal. For this reason, their reading can result in some reiteration specially in the methodology section. A brief introduction of each chapter and the main objectives assessed are presented below.

Chapter 1: *Seasonal and long-term community patterns in temperate marine planktonic protists*

Long temporal series give us the possibility of detecting recurrent patterns and differentiate them from specific disturbances that deviate from the natural yearly succession or multianual tendencies. Previous studies have shown that microorganisms in surface waters exhibit seasonal succession patterns, governed by physicochemical factors. However, no long-term studies have been carried in marine picoeukaryotes. In this chapter we focused in the long temporal community dynamics of marine pico- and nanoeukaryotes in an oligotrophic coastal site, the Blanes Bay Microbial Observatory, through 10 years. We aimed to:

- Establish how repeatable is the community composition through time.

- Develop an *Index* to easily quantify the degree of seasonality of the different taxonomic groups and OTUs (Operational Taxonomic Units).
- Identify seasonal patterns in the rare community.

Chapter 2: *Global changes in activity and community structure of marine picoeukaryotes through the water column*

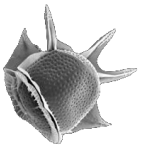
During the last years, global molecular surveys of the photic and the aphotic zone of the ocean have been published, increasing our knowledge on the distribution and diversity of marine protists. However, there is still a lack of information about changes through the water column. The Malaspina 2010 expedition generated vertical profiles in different regions of the global ocean that allowed to explore this. In this chapter, we analyzed the picoeukaryotic community through the vertical profile to assess the following objectives:

- Evaluate changes in community composition through the vertical gradient.
- Explore the vertical segregation of the different taxonomic groups.
- Assess changes in the activity of the different taxonomic groups with depth.

Chapter 3: *Chemotactic response of natural protists communities towards various stimuli*

The capacity of having a directional response towards a chemical gradient (chemotaxis) such as hotspots of organic matter, sulfur compounds or inorganic nutrients has been widely studied in bacteria. However, despite protist are known to be the major grazers in the ocean and to respond to some chemical cues, their chemotactic capacity has been poorly studied, with only few studies using cultured strains. In this chapter, we expand the knowledge of how protists present in natural communities respond to different chemical attractants by analyzing their preferences. We aimed to:

- Asses if the protists community has a marked preference for a specific chemoattractant.
- Evaluate wheather or not differently sized protists react equally.
- Compare the response of phototrophic and heterotrophic protists.



Chapter 4: *Environmental sequencing provides reasonable estimates of relative abundance of specific picoeukaryotes*

Nowadays most of the information used in microbial diversity studies is based only in the relative abundances obtained by high-throughput sequencing of the 18S rDNA. Nevertheless it is still unclear how the relative abundance correlates with the true cell abundance obtained by microscopy. Making use of the data obtained in the BioMarKs project we compared both approaches in six picoeukaryotic taxa. The objectives of this chapter were:

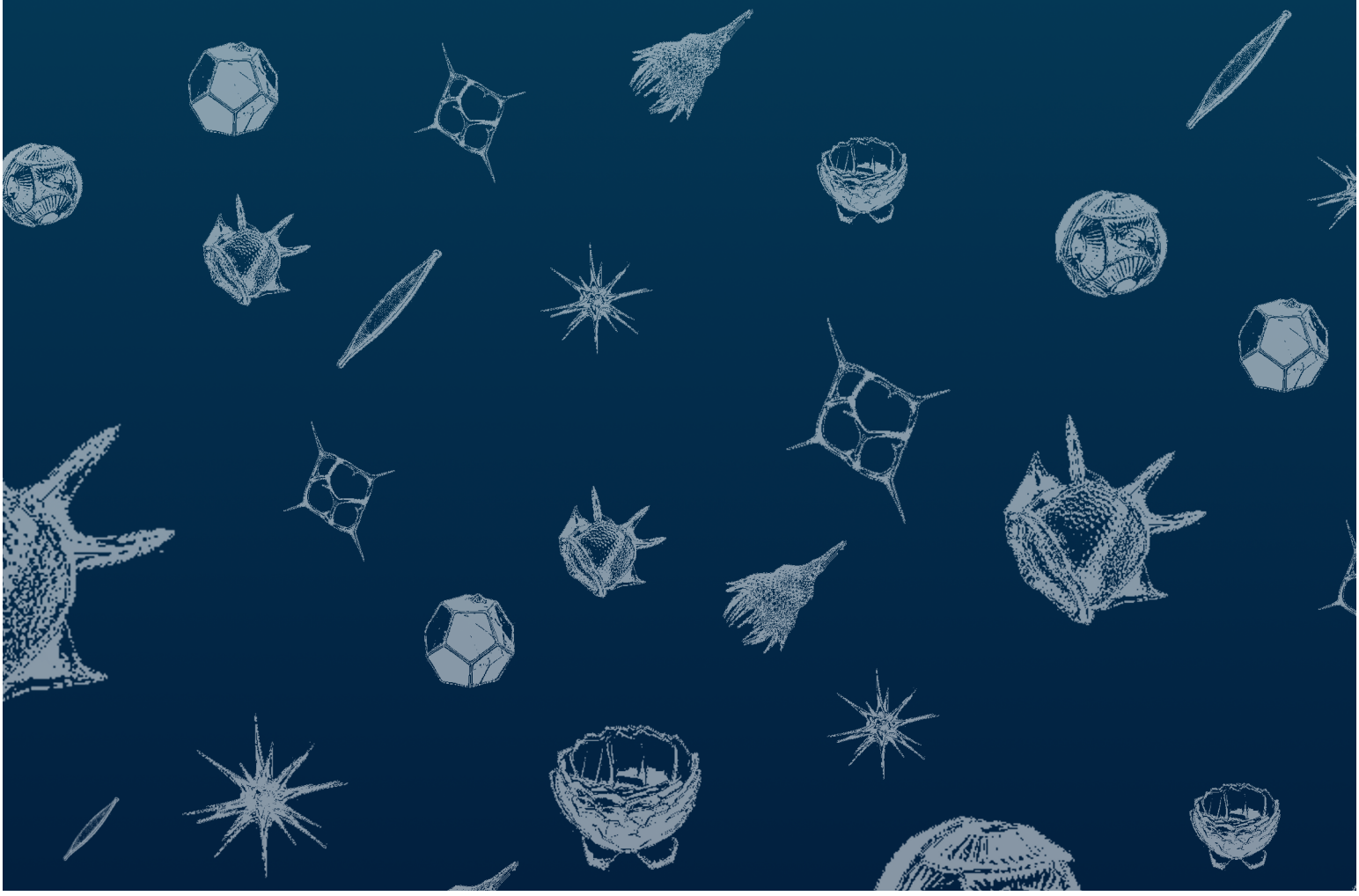
- Assess differences between the information obtained by sequencing different regions of the 18S rDNA (V4 and V9) and cell counts.
- Determine differences between sequencing environmental DNA or RNA, and relate it with the cell counts.
- Develop a new FISH probe to tag a picoeukaryotic group, the MAST-7.

The four studies presented focus in the environmental diversity of marine protists. Notwithstanding, they differ in the methodology and/or in the organismal size used. The following table summarizes the differences between each chapter.

	Protist size		Technique		Nucleic acid	
	Picoeukaryotes	Nanoeukaryotes	Sequencing	Microscopy	DNA	RNA
Chapter 1	×	×	×		×	
Chapter 2	×		×		×	×
Chapter 3	×	×		×		
Chapter 4	×		×	×	×	×

Chapter 1

Seasonal and long-term community patterns in temperate marine planktonic protists



Caterina R. Giner, Vanessa Balagué, Anders K. Krabberød, Isabel Ferrera, Albert Reñé, Esther Garcés, Josep M. Gasol, Ramiro Logares and Ramon Massana

ABSTRACT

Marine microbes are essential for the functioning of the biosphere, yet their long-term dynamics are poorly understood. In particular, we know little on how predictable is the assembly and disassembly of communities in temperate zones featuring annual climate cycles. Here we have analyzed the temporal dynamics of microeukaryotes inhabiting a temperate oligotrophic coastal site during 10 years using metabarcoding. The community showed two recurrent states, corresponding to summer and winter months. Using a 'Seasonality Index' developed *ad hoc*, we identified seasonal taxa featuring wave or pulse strategies. Opportunistic taxa with exceptional peaks of abundance were also detected. We also identified seasonality in a number of permanently rare taxa. Overall, the analysis of a 10-year time series allowed us to explore in depth the long-term patterns that emerge in dynamic microbial communities as well as to determine how much predictability vs. stochasticity is present in the seasonal re-assembly of microbial communities.

INTRODUCTION

A major challenge in microbial ecology is to unveil the mechanisms that determine community structure across space and time. Multiple studies have investigated community turnover (beta diversity) in a spatial context (Lindström & Langenheder 2012; Logares *et al*. 2014), which is assumed to be the product of contemporary and historical processes (Leibold *et al*. 2004; Ricklefs 2007). In a metacommunity framework (Leibold *et al*. 2004) important assembly mechanisms include species sorting (i.e. local environmental conditions determining community composition), mass effects, and neutral dynamics. In planktonic prokaryotes, species sorting seems to be the most important process for structuring communities in different environments (Lindström & Langenheder 2012). Similarly, temporal patterns of microbial community assembly could be explained by variation of the same structuring mechanism (e.g. changes in species sorting that select for different assemblages over time) or by predominance of different mechanisms at different times (e.g. species sorting vs. neutral dynamics). An important difference between the latter is that whereas in a spatial scenario species are mainly recruited via regional dispersal, in a temporal scenario taxa can also be recruited from the pool of low-abundant or dormant species.

Most planktonic ecosystems experience natural environmental fluctuations that take place at different timescales and influence the structure of microbial assemblages. In temperate zones, the annual cycle driven by the different meteorological seasons is the most prevalent. This results in a cyclical change in light, temperature, and nutrient availability that could change microplankton diversity and biomass. Long-term studies conducted over several years may allow characterizing the response of microbial communities to the annual cycle. Recurrent fluctuations of environmental conditions are expected to promote a cyclical annual community dis-assembly and re-assembly, which is called seasonality. In a seasonal community, not all of its members may show cyclic abundance patterns. Therefore, it is important to know which fraction of taxa displays predictable cyclic patterns and which fraction is stochastic. This is not only important to understand the dynamics of the system, but also to understand the long-term response of communities to disturbance or global change, as deviations from a cyclic pattern could point to ecosystem change.

To date, most of our knowledge on the temporal dynamics of microbial communities originate from short-time (i.e. <2 years) bacterioplankton studies, which typically have found community turnover related related with to different seasons (Andersson *et al.* 2010; Lindh *et al.* 2015). The few long-term studies (>5 years) conducted to date have confirmed this seasonal pattern

(Fuhrman *et al.* 2006; Gilbert *et al*. 2012; Cram *et al.* 2015), suggesting that bulk community turnover is mostly driven by cyclically fluctuating environmental conditions (i.e. dynamic species sorting). Despite the fundamental role played by protists in planktonic marine ecosystems, the few studies on their community temporal turnover have only analyzed short time-scales (Romari & Vaulot 2004; Countway *et al*. 2010; Kim *et al*. 2014; Genitsaris *et al*. 2015; Piredda *et al*. 2017) suggesting seasonal patterns similar to those observed in bacterioplankton.

Most microbial assemblages consist of a few abundant species, responsible for the majority of the processes that guarantee continuous ecosystem functioning, and many low-abundant taxa (Pedrós-Alió 2006; Logares *et al.* 2014). Yet, some low abundance taxa could have key roles for ecosystem function (Logares *et al.* 2015). In addition, those rare taxa that have the potential to become abundant may represent a source of diversity for the abundant community under fluctuating environmental conditions or disturbances. Rare taxa can be dormant or metabolically active (Logares *et al.* 2015), they can respond to environmental change (Campbell *et al.* 2011; Lindh *et al.* 2015) and present seasonality (Alonso-Sáez *et al.* 2015). Long-term temporal surveys are essential to understand the behavior and ecological role of low-abundance taxa, which can belong to one of these categories as per our definition: a) taxa systematically recruited to the abundant community during specific time periods (seasonal taxa), b) globally rare taxa that become abundant exceptionally and for a short time (opportunistic taxa; a.k.a conditionally rare taxa (Shade *et al.* 2014), and c) taxa that never become abundant (permanently rare taxa). We also propose that seasonal taxa may display different dynamics, being abundant for long (wave-behavior) or short (peak-behavior) periods, which suggest different ecological strategies. Finally, we indicate that the community could include taxa that are typically not rare (that is, that show moderate or high abundances most of the time), as for example microeukaryotic parasites that switch between seasonal hosts.

Here, we present the first long time-series exploration of protist communities inhabiting a model oligotrophic temperate coastal site in the Mediterranean Sea (Blanes Bay Microbial Observatory, Gasol *et al.* 2016). To explore long-term community dynamics and re-assembly patterns, picoeukaryotic and nanoeukaryotic communities were sampled every month for 10 years and their composition was analyzed by high-throughput sequencing of the V4 region of the 18S rDNA. We found that the system presents two main configurations, corresponding to summer and winter months. Using a seasonality index we developed ad hoc we determined that 13.2% of the OTUs and 22.4% of the groups are seasonal. In contrast, we identified OTUs and groups that presented no predictability in their dynamics. In particular, we identified taxa that

can be regarded as opportunistic. Finally, we found a number of permanently rare taxa that feature seasonality.

MATERIALS AND METHODS

Study site and sampling of planktonic protists

We carried out a monthly sampling during 10 years at the Blanes Bay Microbial Observatory (BBMO) located in the North Western Mediterranean Sea (41º 40' N, 2º 48' E). This is a well-studied temperate oligotrophic coastal site that has relatively little human or riverine influence (Schauer *et al.* 2003; Massana *et al*. 2004; Alonso-Sáez *et al.* 2007; Alonso-Sáez *et al.* 2008; Guadayol *et al.* 2009; Gasol *et al*. 2016). Surface water was sampled about 1 km offshore over a water column of 20 m depth, from January 2004 to December 2013. Water temperature and salinity were measured in situ with a CTD. Seawater was pre-filtered through a 200 µm nylon-mesh, transported to the laboratory under dim light in 25 L plastic carboys, and processed within 2 h.

Samples for determination of chlorophyll a concentration were filtered in GF/F filters, extracted by acetone and processed in a fluorometer (Yentsch & Menzel 1963). Inorganic nutrients (NO_3-, NO_2-, NH_4^+, PO_4^{3-}, SiO_2) were measured spectrophotometrically using an Alliance Evolution II autoanalyzer (Grasshof *et al.* 1983). Samples for determination of phototrophic and heterotrophic protists abundances (sizes from ~1 to 5 µm) were fixed with glutaraldehyde (1% final concentration), mounted in a slide with low-fluorescence oil, and counted using epifluorescence microscopy at 1000X (Porter & Feig 1980). For statistical analyses, these variables were standardized as z-scores, that is, deviations of the values from the global mean in standard deviation units.

About 6 liters of the 200 µm prefiltered seawater were sequentially filtered using a peristaltic pump through a 20 µm nylon mesh, a 3 µm pore-size polycarbonate filter of 47 mm diameter (nanoplankton fraction, 3-20 µm), and a 0.2 µm pore-size Sterivex unit (Millipore, Durapore) (picoplankton fraction, 0.2–3 µm). Sterivex units and the 3 µm filters were filled with 1.8 ml of lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) and stored at -80ºC until processed. DNA extractions were done at the end of the sampling period using the standard phenol-chloroform protocol (Schauer *et al*. 2003; Massana *et al*. 2004), with a final step of purification in

Amicon units (Millipore). Nucleic acid extracts were quantified in a NanoDrop 1000 spectrophotometer (Thermo Scientific) and stored at -80ºC until analysis.

Illumina sequencing and bioinformatics

The eukaryotic universal primers TAReuk454FWD1 and TAReukREV3 (Stoeck *et al*. 2010) were used to amplify the V4 region of the 18S rDNA (~380 bp). PCR amplifications and amplicon sequencing were carried out at the Research and Testing Laboratory (Lubbock, TX, USA; http://www.researchandtesting.com) using the Illumina MiSeq platform (2x250bp paired-end reads). Illumina reads were processed following an in-house pipeline (Logares 2017). Briefly, raw reads were corrected using BayesHammer (Nikolenko *et al.* 2013) as indicated by Schirmer (Schirmer *et al.* 2015). Corrected paired-end reads were subsequently merged with PEAR (Zhang *et al.* 2014) and sequences longer than 200 bp were quality-checked and dereplicated using USEARCH (Edgar 2010). OTU clustering at 99% similarity was done using UPARSE (Edgar 2013) as implemented in Usearch v8. Chimera check and removal was performed both de novo and using the SILVA reference database (Quast *et al*. 2013). Only OTUs present in at least 3 samples were retained. Taxonomic assignment was done by BLASTing OTU representative sequences against three reference databases: PR2 (Guillou *et al.* 2013) and two in-house marine protist databases (available at https://github.com/ramalok) based in a collection of Sanger sequences from molecular surveys (Pernice et al. 2013) or 454 reads from the BioMarKs project (Massana *et al.* 2015). Metazoan, Charophyta and nucleomorphs were removed after taxonomic assignment. In downstream analyses, protists diversity was assessed in two size fractions, the picoeukaryotes (0.2-3 μm) and the nanoeukaryotes (3-20 μm). Nanoplankton samples from May-2010 to July-2012 and from 4 additional dates were discarded due to suboptimal amplification or sequencing. The combined final OTU table had 209 samples (120 picoeukaryotes and 89 nanoeukaryotes) and 17,518 OTUs. The picoeukaryotic-OTU table had 120 samples and 16,989 OTUs. To enable sample comparisons, both tables were randomly subsampled to the lowest number of reads per sample using the *rrarefy* function in *vegan* (Oksanen *et al.* 2008). The combined subsampled table had 14,771 OTUs and 5,898 reads per sample, while the picoeukaryotic-table had 13,040 OTUs and 7,553 reads per sample.

Analyses of rare or opportunistic taxa

Only the picoplankton (picoeukaryotic-OTU table) was used in rarity analyses, as this was the most complete dataset. OTUs with abundances per sample that were always <0.1% were considered rare (Logares *et al*. 2014). To exclude the possibility that rare OTUs were aberrant variants of abundant ones, we only analyzed rare OTUs that had a similarity <97% with any abundant counterpart. We considered as temporally abundant those OTUs with a mean abundance >0.1% along 10 years. Conditional Rare Taxa (opportunistic) were detected following the protocol described in Shade *et al.* (Shade *et al*. 2014).

Alpha and beta diversity analyses

Alpha-diversity was estimated using richness (i.e. number of OTUs) and the Shannon index (H'), while Beta-diversity was estimated using Bray-Curtis dissimilarities between pairs of samples. These indices were calculated from combined subsampled OTU tables. Non-metric multidimensional scaling (NMDS) was performed based on Bray-Curtis dissimilarity matrices. In NMDS, differences between predefined groups were tested with ANOSIM (analysis of similarity, Clarke 1993) performing 1,000 permutations. To identify the OTUs contributing to group dissimilarities we used SIMPER (SIMilarity PERcentage analysis). To determine the proportion of variation in community composition explained by the measured environmental variables we used PERMANOVA. In addition, we analyzed the correlation between environmental variables and community differentiation using Partial Mantel tests (Legendre & Legendre 1998). Further relationships between community and environmental variation were analyzed by fitting environmental variables onto the ordination space of the NMDS (*envfit* function in *vegan*). Finally, we performed an IndVal analysis (INDicator VALues, Dufrene & Legendre 1997) to identify OTUs associated to a specific season. OTUs with statistically significant ($p<0.05$) IndVal values >0.3 were considered, following Logares et al. (2013). All the mentioned analyses were performed using functions implemented in the packages *vegan* (Oksanen *et al.* 2008), *pvclust* (Suzuki & Shimodaira 2006) and *labdsv* (Roberts 2016) of the R Statistical environment (R Developement Core Team, 2015).

Local Similarity Analysis

Pairwise co-occurrence and correlation matrices were generated using the extended local similarity analysis (eLSA) (Ruan *et al*. 2006; Xia *et al*. 2011). The analysis was performed on the

subsampled OTU-table together with the environmental variables. OTUs that were not present in at least 10 of the 120 months were excluded from the analysis, resulting in a dataset with 1,065 OTUs and 14 selected environmental variables. ELSA was run with default normalization (a Z-score transformation using the median instead of mean and median absolute deviation instead of standard deviation) and p-value estimations under a mixed model that performs a random permutation test of a co-occurrence only if the theoretical p values for the comparison is <0.05. Missing data was interpolated linearly from adjacent months, and we did not allow any time delay.

Seasonality analyses

To investigate picoplankton seasonality, we computed the mean Bray-Curtis dissimilarity for all pairs of samples taken n months apart (n ranging from 1 to 119 months). Furthermore, we developed a Seasonality Index (SI) aiming to identify which taxa present seasonality or not. To calculate the SI, we computed the ACF (Auto Correlation Function) comparing taxa (OTUs or taxonomic Classes) relative abundances at different time lags. Then, we sum the absolute ACF values for each taxa in the complete temporal series (SF). Afterwards, we repeated the latter process 1,000 times with randomized abundances, and calculated the mean (SFrandom) plus its 97% confidence intervals. The SI was calculated as: SI=SF/SFrandom. Based on empirical tests, a given taxa was considered seasonal if its SI was above 1.2, and to avoid seasonality by randomness, its SF was significantly higher than SFrandom (that is, outside the upper 97% confidence interval). The SI was applied only to picoeukaryotes, which encompassed the complete time series of 120 points. The SI is implemented and publicly available in EcolUtils (Salazar, 2015; https://github.com/GuillemSalazar/EcolUtils).

To analyze the ‘pulse-‘ or ‘wave-strategy’ of the seasonal taxa, we computed the mean abundance of each taxon. Wave OTUs had an abundance higher than their mean for more than 30 months, while the ‘pulse-strategy' was shorter.

RESULTS

Communities show two main states along the 10 years

The samples were clearly differentiated by cell size (picoeukaryotes vs. nanoeukaryotes) in the NMDS (Fig. 1a). In addition, both communities showed the same temporal pattern with samples generally grouped by season, but forming two clearly differentiated groups corresponding to Winter and Summer (ANOSIM test: R_{pico}=0.717; R_{nano}=0.713, p<0.001) and with spring and autumn communities appearing between these two seasons and not forming clear groups (Fig. 1a), if anything slightly more similar to the summer state (Table S1). Furthermore, winter communities were more similar between themselves (i.e. formed a tighter cluster) than the communities from other seasons, as also shown by their smaller average Bray-Curtis dissimilarity (Fig. 2). We further explored which OTUs contributed to explain the differences between groups using SIMPER tests. Within picoeukaryotes, 71 OTUs explained 52% of the difference between winter and summer assemblages, whereas 240 OTUs had to be considered to explain the same percentage difference between spring and summer groups. Finally, the IndVal analysis detected 173 season-specific OTUs (IndVal >0.3, p<0.05), most of them associated to winter and summer states (56 and 59 OTUs respectively, Table S2).

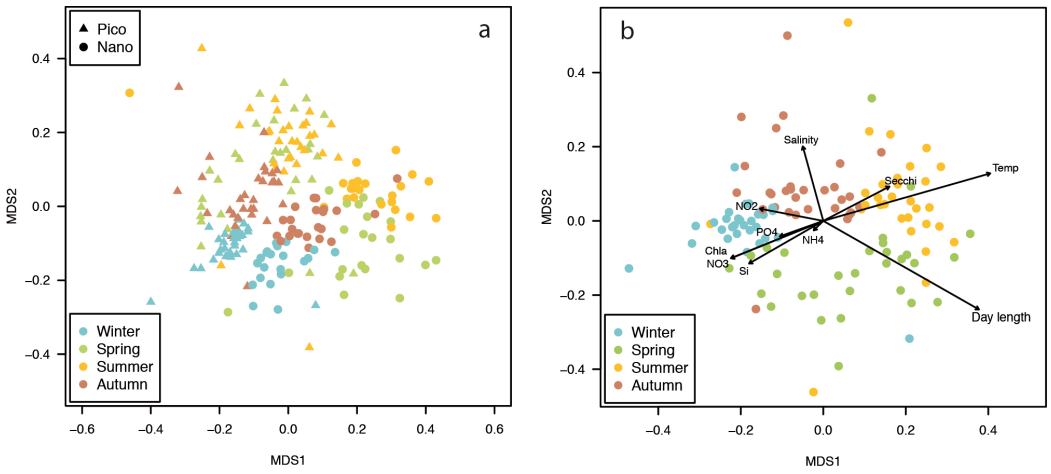


Fig. 1. Community comparison of protists assemblages in monthly samples taken during 10 years in Blanes Bay. (a) Non-metric multidimensional analysis (NMDS) based on Bray-Curtis dissimilarities among 209 samples (120 of picoeukaryotes and 89 of nanoeukaryotes) taken at different seasons. (b) NMDS analysis of the picoeukaryotic communities showing the environmental vectors that better fit the plot after an *Envfit* test.

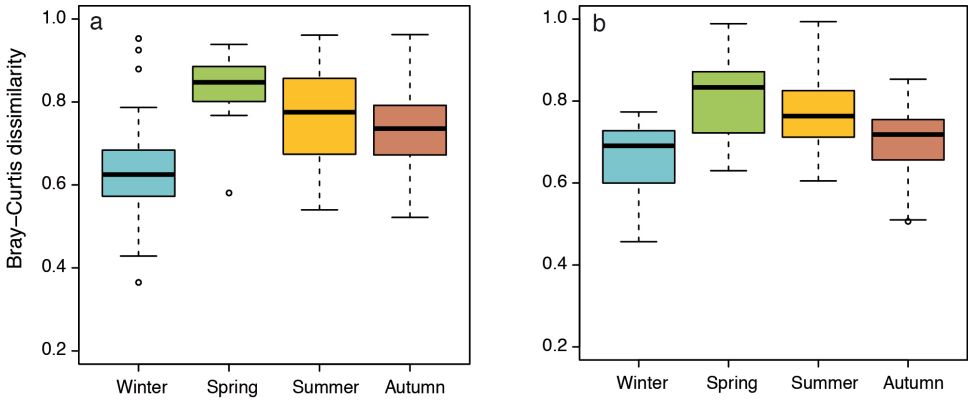


Fig. 2. Differences within picoeukaryotic (a) and nanoeukaryotic (b) community composition in each season, shown as boxplots of Bray-Curtis dissimilarities among all pairs of samples taken at the same season during 10 years.

Limited explanation of community turnover by the measured environmental factors

The Blanes Bay site features annual cyclic fluctuations of environmental conditions that could drive dynamic species sorting (that is, different and repeatable environmental selection at different times). At this site, the day was longer in early summer, water temperature was maximal two months later, and inorganic nutrients, particularly nitrate, nitrite and silicate, peaked in winter (Fig. S1). This promotes algal blooms in late winter or early spring as seen by chlorophyll a concentration. Cell abundances of phototrophic and heterotrophic protists (1-5 µm in size) also followed temporal trends, with maximal abundance observed in winter and summer respectively (Fig. S1).

Selected environmental variables were fitted to the NMDS separately for picoeukaryotes (Fig. 1b) and nanoeukaryotes (Fig. S2). In both cases day length and temperature were the variables most correlated with community variation (envfit day length r^2=0.62, temp r^2=0.56, p<0.001; Table S3). When controlled by each other in partial mantel tests both variables still presented a moderate correlation with community composition (r=0.44 for temperature, r=0.40 for day length; p=0.001). The remaining environmental variables presented weaker or non-significant correlations with community composition (Table S3). Additional analyses indicated that a large part of community variance (76.8% in PERMANOVA) was not explained by any of the measured

environmental variables. Day length and temperature together, only explained 16% of community variance ($p<0.001$) in PERMANOVA analysis, a value that increased to 26% when running the analysis only with the OTUs that have seasonality (defined below).

Detection of OTUs with differential response to environmental variables

Even though there was a weak correlation between community composition and environmental variability, multiple taxa could still present strong differential responses to environmental heterogeneity (e.g. two OTUs could be highly correlated to environmental variables A and B respectively, which present different patterns, while the remaining are not). To disentangle which OTUs react differentially to the analyzed environmental variables, we investigated the correlation of individual OTUs with each variable using eLSA analyses. We detected 2,375 OTUs that were positively or negatively correlated with most of the analyzed environmental variables (Table S4). Although these OTUs were always few in numbers, they tended to be abundant. For example, about 3-4% of the OTUs correlated positively or negatively with temperature and day length, representing ~47% of the total abundance. Other variables that presented differential correlations were chlorophyll *a*, nitrate, silicate and salinity (Table S4).

Diversity patterns

Most individual samples (~80%) were close to richness (as number of OTUs) saturation. In addition, we found richness saturation when constructing rarefaction curves based on the complete dataset (total number of reads) of pico- and nanoeukaryotes (Fig. S3a), indicating that we recovered most of their diversity present in Blanes Bay throughout the 10 years. Richness increased rapidly until approximately the 60th sample (60th month of sampling), and after that, subsequent samples contributed with very few new OTUs (Fig. S3b).

Alpha diversity presented clear temporal trends. For the pico- and nanoeukaryotes, averaged richness and Shannon index (H) were highest during the autumn and winter months and significantly lower during spring (Fig. 3, $p<0.05$ Wilcoxon test). No statistical differences were found between pico- and nanoplankton when comparing all samples together, nor when we compared each of the seasons separately (Wilcoxon test $p>0.05$).

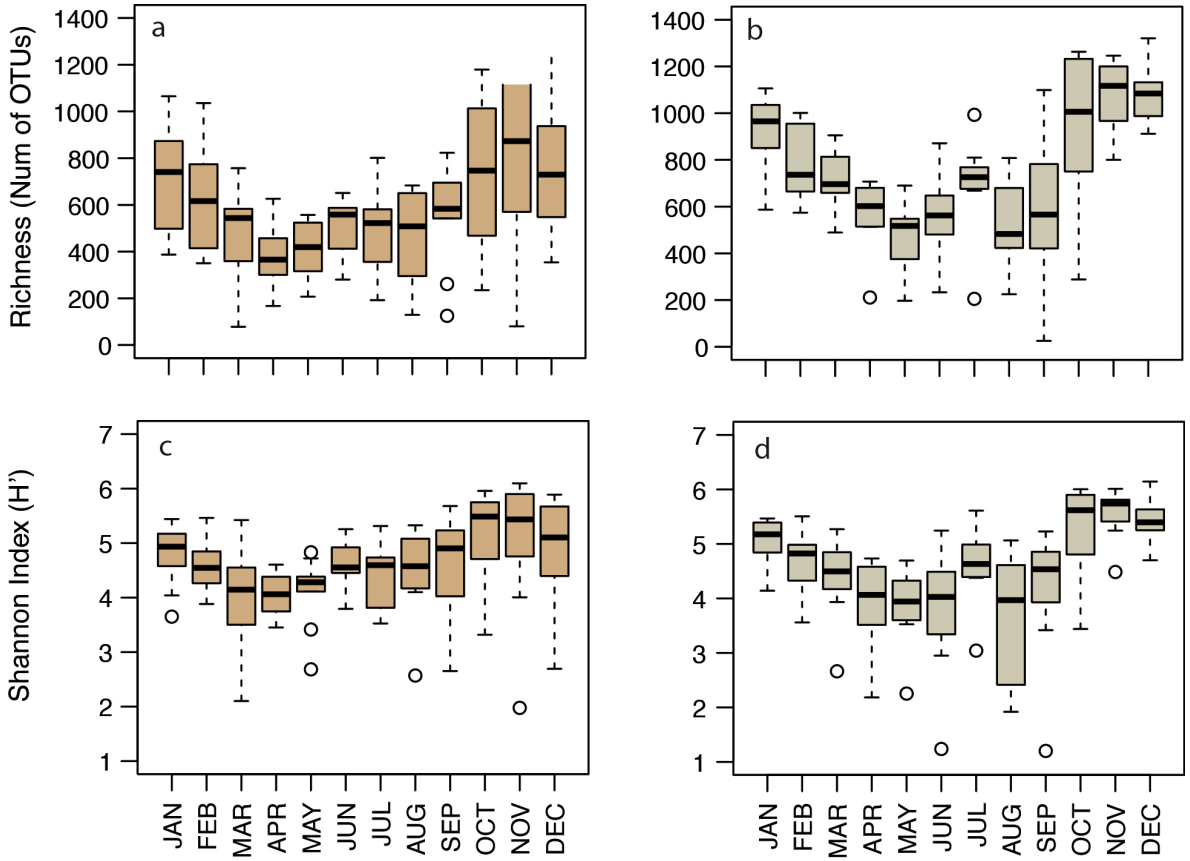


Fig. 3. Average monthly variation of alfa diversity in Blanes Bay protist assemblages. Boxplots display the monthly variability of the richness (a, b) and Shannon Indices (c, d) for the picoplankton (a, c) and nanoplankton (b, d) during the 10 years.

Community seasonality

In order to assess recurrent patterns of community composition, we calculated pairwise Bray-Curtis dissimilarities between communities separated by different time lags. Communities separated 12 months and their multiples (24, 36 and so on) showed the highest similarity, while those separated 6 months and their multiples showed the highest dissimilarity for both pico- and nanoeukaryotes (Fig. 4). Despite this dissimilarity cycling, the community composition was not exactly the same in successive years, as the averaged Bray-Curtis values remained always rather high, from 0.7 to 0.9. Yet, the differentiation among communities did not increase with time; Bray-Curtis distances among samples separated by 12 months were very similar to those from samples separated e.g. 60 or 84 months.

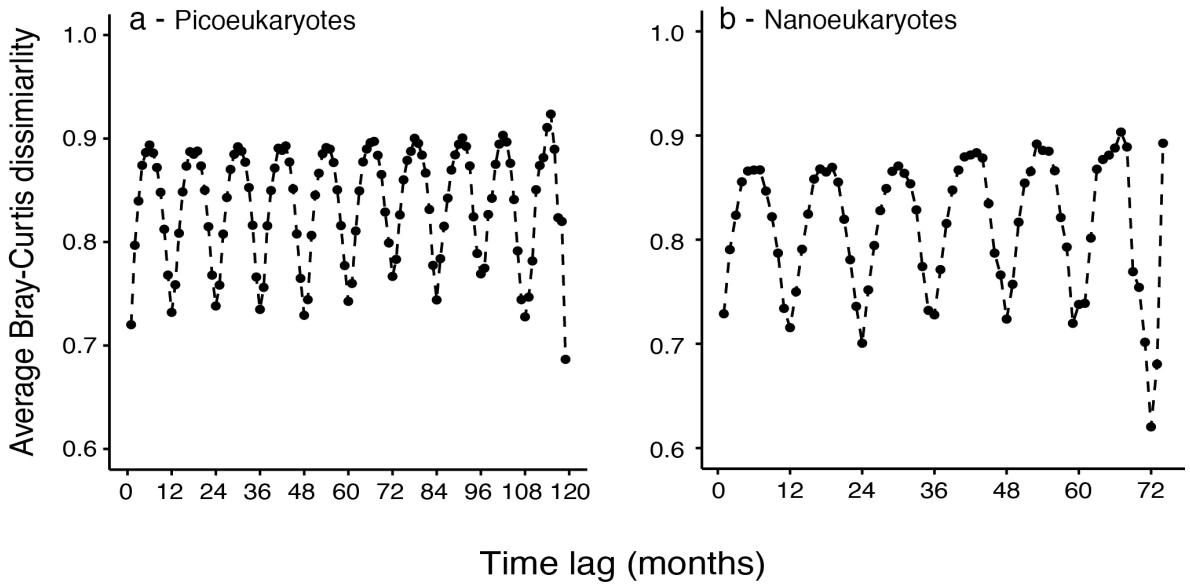


Fig. 4. Interannual reccurrence of communities of picoeukaryotes (a) and nanoeukaryotes (b), shown by the average of Bray-Curtis dissimilarities of all pairs of communities separated by a given number of months (from 1 to 119 in 'a' from 1 to 74 in 'b').

Community patterns in the rare sub-community

The seasonal patterns at the whole community level shown above are driven by the most abundant OTUs, which have a stronger weight in Bray-Curtis dissimilarities. Therefore, we investigated whether the rare biosphere exhibited similar seasonality. Within picoeukaryotes, 3,095 OTUs were considered permanently rare. Similar to what we found for the entire community, we observed two main rare sub-community states associated to winter and summer months (Fig. 5a) with transitions in spring and autumn. Furthermore, we also found that the averaged Bray-Curtis values were most similar between rare communities separated by 1 year (and their multiples), and most different when separated by half a year (Fig. 5b). However, the Bray-Curtis values were higher than the ones found for the entire community (from 0.9 to almost 1) indicating that even though there was evidence of seasonality for this sub-community, the rare assemblage was very different from year to year.

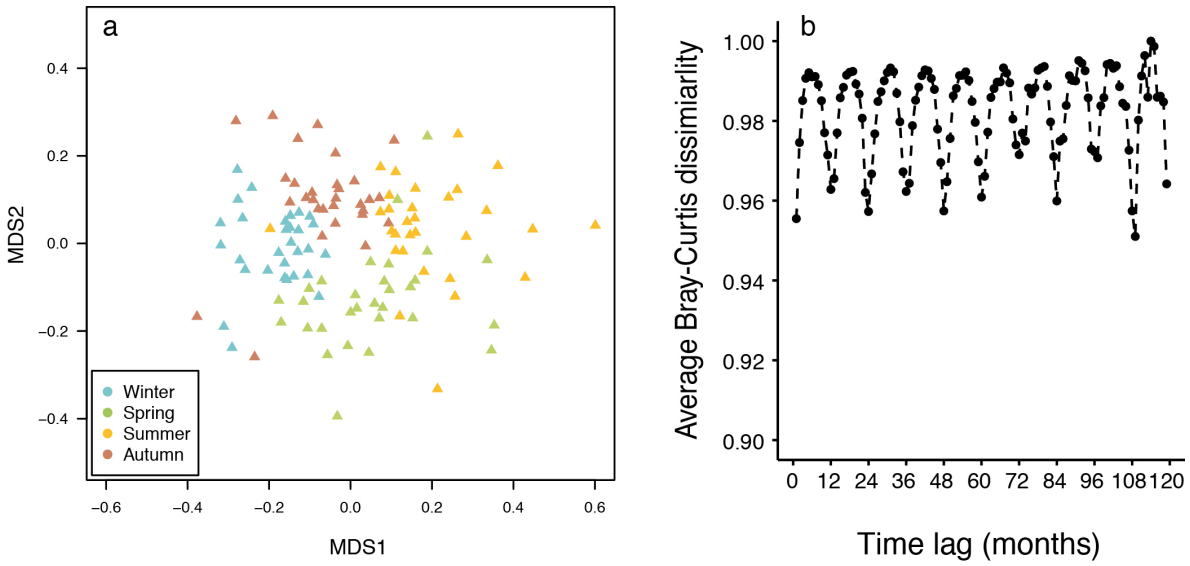


Fig. 5. Seasonality and interannual reccurrence of rare OTUs within picoeukaryotes. (a) NMDS plot based on Bray-Curtis dissimilarities. (b) Average of Bray-Curtis dissimilarities of all pairs of communities separated by a given number of months (from 1 to 119).

Contrasting seasonal patterns in different taxa

The protist community present in the Blanes Bay was very diverse, including more than 63 taxonomic groups at the class level (Table S5). Most of these groups were more abundant in the picoplankton than in the nanoplankton (Table S5). For an exploration of temporal patterns we focused on the picoeukaryotic assemblage, which included 10 years of uninterrupted monthly data. Picoeukaryotes were mainly dominated by different alveolates (MALV-I, Dinoflagellata, MALV-II) and Mamiellophyceae (Fig. 6a). The relative abundance of these groups changed along the year, indicating a recurrent behavior for some of the taxa (Fig. S4). In particular, there was a clear peak of Mamiellophyceae each winter. The seasonality index (SI) developed here allowed us to identify 'Seasonal' behavior divided in: 'Strongly-Seasonal' (SI >2), 'Moderately-Seasonal' (SI between 2 and 1.2), and 'Non-Seasonal' (SI <1.2) behaviors (examples in Fig. 6b). We found that 13 groups within the picoeukaryotes (35.2% of the reads) were 'Seasonal' (Fig. 6a, Table S6), yet only two of them, MALV-III and Mamiellophyceae, were 'Strongly-Seasonal'. The remaining 11 groups were 'Moderately-Seasonal' and included Dinoflagellata and several environmental clades. The remaining groups were 'Non-Seasonal'.

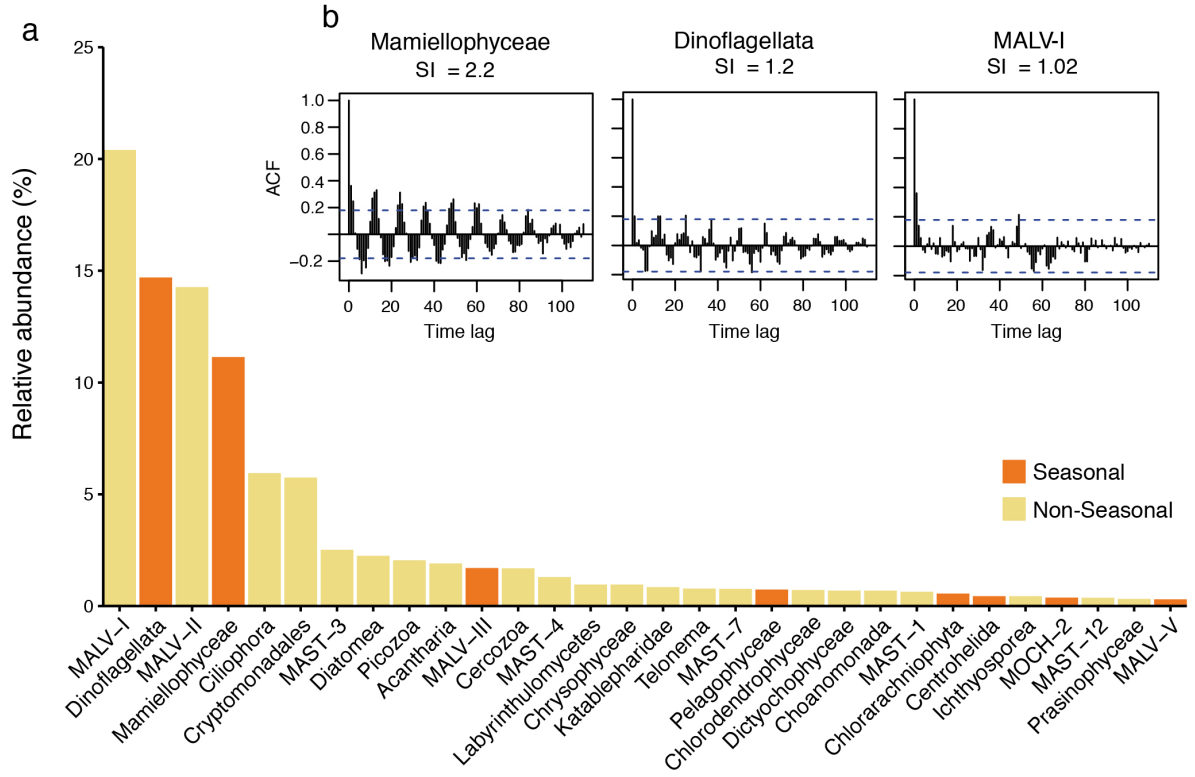


Fig. 6. Taxonomic groups forming the community of picoeukaryotes in Blanes Bay and indications of their seasonality. (a) Average relative abundances of groups accounting for more than 0.3% of reads. The bars are colored according to whether the group as a whole exhibits seasonality. (b) Autocorrelation function (ACF) plots of examples being 'Strong-seasonal' (Mamiellophyceae), 'Moderately-Seasonal' (Dinoflagellata) and 'Non-Seasonal' (MALV-I), together with the SI value for each case.

Within each taxonomic group, the composing OTUs could have different seasonal behaviors. That is, if a taxonomic group is seasonal, that does not necessarily imply that all composing OTUs are also seasonal and vice versa. Therefore, we explored the seasonality of those OTUs present in at least 10 samples (1,898 OTUs accounting for ~90% of reads). Applying the Seasonality Index (SI) only 251 OTUs (representing 39.3% of reads) were seasonal (SI>1.2). As expected, seasonal groups generally contained most reads (abundance) within seasonal OTUs (Fig. 7). Exceptions were low abundance groups (e.g. MALV-V, RAD-B) and the Dinoflagellata, which had a seasonal index just above the cut-off (SI=1.23). We also identified seasonal OTUs in groups that did not show seasonality as a group, e.g. Acantharia, Bolidomonas, Cryptomonadales, Dictyochophyceae, MAST-1, MAST-10 had more reads belonging to seasonal OTUs than to non-seasonal.

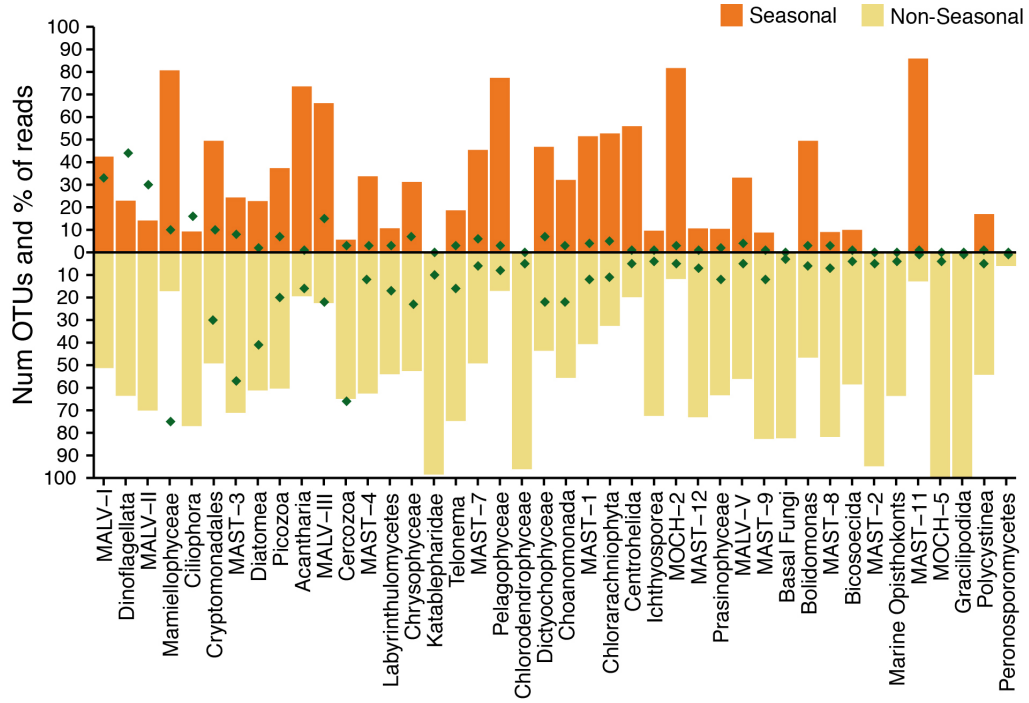


Fig. 7. Seasonal and non-seasonal signal for the main picoeukaryotic taxonomic groups in Blanes Bay. The figure shows the number of OTUs (dots) and their percentage contribution to all reads (bars) for each taxonomic group.

Wave versus pulse strategists

Seasonal taxa exhibited different strategies based on how fast was the increase of their abundance and how long was their persistence as abundant members of the assemblage. Taxa with a "Wave-strategy" increase and decrease their abundance at a slow pace, whereas taxa with a "Pulse-strategy" increase and decrease their abundance very fast, in 1-3 months. We searched for these strategies within the 13 groups identified as seasonal (examples in Fig. S5), and found that nine of them (99.5% of seasonal reads) had a wave-strategy, while only 4 groups (0.5% reads), generally the less abundant, presented a pulse-strategy (Table S6). Wave and pulse strategies were also analyzed for individual OTUs. Out of 251 seasonal OTUs, 31.5% presented a wave-strategy. These belonged to groups that also presented this strategy as a group, and included OTUs with high relative abundance. The remaining 68.5% of seasonal OTUs had a pulse-strategy. Furthermore, among the rare OTUs that appeared in at least 10 samples (89 OTUs), application of the SI allowed detection of nine OTUs that were 'moderately-seasonal' and followed the pulse-strategy.

In summary, we identified several remarkable temporal strategies at the level of individual OTUs. The most evident was the seasonal pattern shown by 251 OTUs that accounted for 39.3% of the reads. Second, we found 3095 permanently rare OTUs (24% of OTUs representing 1.8% of reads), with a few of them being seasonal. Finally, we identified 212 OTUs (1.6%) representing 11.5% of the reads that were conditionally rare taxa or opportunistic taxa, with abundances changing abruptly and unpredictably from being rare to abundant.

DISCUSSION

Increasing our knowledge on the seasonal re-assembly of microbial communities and connecting this process to ecological paradigms like those derived from the metacommunity framework is needed to continue incorporating microbes into ecological theories. This exploration will reveal how much predictability vs. stochasticity is involved in the seasonal community re-assembly as well as the role of low abundance taxa in the architecture of communities. Seasonal analyses can also help us understand differential responses of individual taxa to environmental and biological variability. Here we explored multiple questions related to these issues using data from one of the longest microbial, particularly protistan, time-series analyzed to date (BBMO; Gasol *et al*., 2016). We put particular emphasis in disentangling different behaviors in the community, determining the seasonality of different taxa using our "Seasonality Index" and analyzing differential responses of the various OTUs to environmental variables. Overall, our results are likely representative of other temperate marine-surface environments around the world.

Two main community states and differential OTU response to environmental fluctuations

There was a clear structure in the protist communities through the analyzed 10 years, with a recurrent configuration of pico- and nanoeukaryotes into two main states consisting of winter and summer communities. Spring and autumn communities appear as transitional states between winter and summer, but are generally more similar to summer communities. Other studies of marine protists (Genitsaris *et al*. 2015) and prokaryotes (Andersson *et al*. 2010) have revealed cyclical changes in community composition, but few have analyzed in depth the different states through which communities oscillate. In a 2.5 years time-series three seasonal clusters were observed in the protists community of the English Channel (Genitsaris *et al*. 2015).

The fact that winter communities were the most similar along the 10 years points to assembly mechanisms likely linked to low temperatures. In contrast, summer communities presented a relatively higher variation in community composition over the years, suggesting weaker species sorting. In between these two main community configurations are the compositionally variable transitional states, which could be linked to the disassembly of one community state and the re-assembly of the next. Consequently, species sorting seems to be operating, to certain extent, in winter and summer, but not in the transitional states in-between. An extra aspect pointing to species sorting was the larger number of OTUs exclusively associated to summer and winter (59 and 56 OTUs respectively) as compared to those associated to autumn and spring (36 and 23 OTUs respectively). These OTUs can be considered as season-specialists, and a higher number of them are expected under stronger environmental selection.

Since environmental variables showed seasonality, we naturally hypothesized that dynamic species sorting could be driving the turnover of the community. However, our results indicate that the measured environmental variables explain a minor fraction of overall community variability along the 10 years. One possible reason is that environmental selection has different intensities throughout the year, i.e. stronger in summer and winter and weaker in spring and autumn, giving space for multiple community compositions in spring and autumn. The rationale is that species sorting is expected to be a relevant structuring mechanism under moderate or high local environmental selection (Lindström & Langenheder 2012), and under these circumstances habitats with similar environmental selection are expected to contain comparable microbial assemblages.

Even though the measured environmental variables could explain a minor fraction of the overall community turnover, we hypothesized that individual OTUs could be differentially correlated to environmental variation. We found support for the latter hypothesis. For example, we found that a small percentage of OTUs (~4%), yet representing ~47% of the total abundance, correlated positively or negatively with temperature and day length. Thus, it appears that environmental heterogeneity generates multiple taxa specific responses, which are poorly detected with multivariate analyses encompassing the whole community.

Predictability and stochasticity in community re-assembly

The investigated pico- and nanoeukaryotic communities displayed a dynamic composition during the 10 years, with increased similarity when separated every 12 months. This points to a

certain degree of resilience and predictability in species composition (Fuhrman *et al*. 2015), in contrast to a scenario where different ecologically or functionally redundant species predominate in community re-assembly. In this context, there is evidence of high resilience in planktonic protists after facing droughts in ponds (Simon *et al.* 2016). Nevertheless, communities separated by a year were far from identical (averaged BC dissimilarity ~0.8), indicating stochasticity in the re-assembly process as well. However, the differentiation between communities did not increase with time along the 10 years, indicating no temporal trends in community dissimilarity. This indicates that at least some abundant taxa should present annual seasonality and that large immigration events probably do not have relevant effects among the abundant taxa. Overall, we did not find evidence of large-scale immigration to the system. We have observed that protist diversity was saturated after 5 years of sampling.

The permanently rare sub-community mirrored the recurrent annual pattern found in the whole community. Yet, the overall dissimilarity values were higher even for samples separated by 12 months (BC dissimilarity 0.9-1). This points to a limited compositional repeatability among rare taxa, implying a larger stochasticity when compared to the whole community. In any case, we should also consider that some members of the rare sub-community are not identified due to the detection limits of our sequencing approach, thus inflating the dissimilarity between rare sub-communities. Accordingly, other studies have found that rare OTUs were randomly sampled during DNA sequencing (Leray & Knowlton 2017), inflating beta diversity estimates between replicates.

Quantifying community seasonality and predictability

One of the main issues with temporal studies is the lack of mechanisms to quantify the seasonality of taxa. Here, we developed a seasonality index (SI) that allowed us quantifying seasonality for each taxonomic Class or OTU. We identified two groups, Mamiellophyceae and MALV-III, featuring strong seasonality. As expected, most OTUs within those groups were also seasonal, indicating that this was a conserved trait in most species of the group. The strong seasonality within MALV-III was remarkable, given that virtually nothing is known about this group, whereas Mamiellophyceae were already known to have a preference for low temperatures (Foulon *et al.* 2008). The opposite scenario, seasonal OTUs within non-seasonal groups, was also found and was explained by contrasting seasonality in different OTUs within a group.

We also used the seasonality index to quantify the proportion of the community that presented repeatable patterns through 10 years. We found seasonality in only 1.9% of the OTUs (yet accounting for 35.2% of the total abundance [as measured by number of reads]). Thus, the majority of taxa in the community did not present clear seasonal patterns, but the few that did were particularly abundant. This suggests a lack of seasonality among rare taxa or ecological or functional redundancy (Allison & Martiny 2008) by which different ecologically redundant OTUs become dominant during different years. So far, there is limited evidence supporting ecological redundancy in microbial communities, as different studies evidence that compositional changes are followed by ecosystem processes (Allison & Martiny 2008). Yet, more detailed studies are needed to determine the long-term role of ecological redundancy in the microbial plankton.

Seasonal, conditionally rare and permanently rare taxa

Among the investigated taxa, we identified at least three behaviors or strategies: OTUs that were seasonal, conditionally rare (opportunistic) or permanently rare. Seasonal taxa presented either pulse or wave behaviors. The pulse behavior reflects a fast growth on specific resources or a high predation and competitive pressure. On the other hand, the wave behavior could reflect relatively slower growth (and slower use of resources) accompanied with relatively lower predation or competition pressures, thus maintaining the taxa in the system for relatively longer periods. Wave strategists may also have their growth rate tightly associated to some environmental variables (e.g. temperature), with their abundance reflecting the environmental variability.

We have also found that 1.6% of the OTUs were Conditionally Rare Taxa (CRT), a value coinciding with that observed by Shade and Gilbert (2015) for prokaryotes. We considered these OTUs opportunistic, with an increase in abundance triggered by environmental cues. CRTs may contribute to the active component of protist communities and be responsible, in some cases, of increasing community stochasticity. Finally, some OTUs (23.7%) were permanently rare, and only a few of them showed seasonality, similarly to what was observed in bacterioplankton (Alonso-Sáez *et al*. 2015). Altogether, the latter suggests that some rare OTUs are adapted to live a low-abundance lifestyle, that is, they will never be abundant even under the most favorable environmental conditions (Logares *et al*. 2015).

Concluding Remarks

In summary, we have shown that the protist community has a recurrent annual pattern with two main states, summer and winter. Despite this clear separation, environmental variation explains a small proportion of the community variability, which is probably a result of different species reacting to different variables at different times. Additionally, we developed a 'Seasonality Index' to quantify the seasonality of the different taxa and revealed that only a small fraction of the OTUs (1.7%-check), albeit they contributed a large number of reads, was cyclic through the 10 years. We also characterized the different behaviors of the OTUs as seasonal, opportunistic or permanently rare and we detected seasonality in the rare sub-community, which followed the same trends as the whole community.

ACKNOWLEDGEMENTS

We thank all members of the Blanes Bay Microbial Observatory sampling team and the multiple projects funding this collaborative effort over the years. The data and analysis done here have been funded by the Spanish projects FLAME (CGL2010-16304, MICINN) and ALLFLAGS (CTM2016-75083-R, MINECO) to RM and INTERACTOMICS (CTM2015-69936-P, MINECO/FEDER, EU) to RL, and DEVOTES (funded by the European Union, grant agreement no. 308392). CRG was supported by a FPI fellowship. RL was supported by a Ramón y Cajal fellowship (RYC-2013-12554, MINECO, Spain). We thank Guillem Salazar for his help in R analyses.

SUPPLEMENTARY MATERIAL

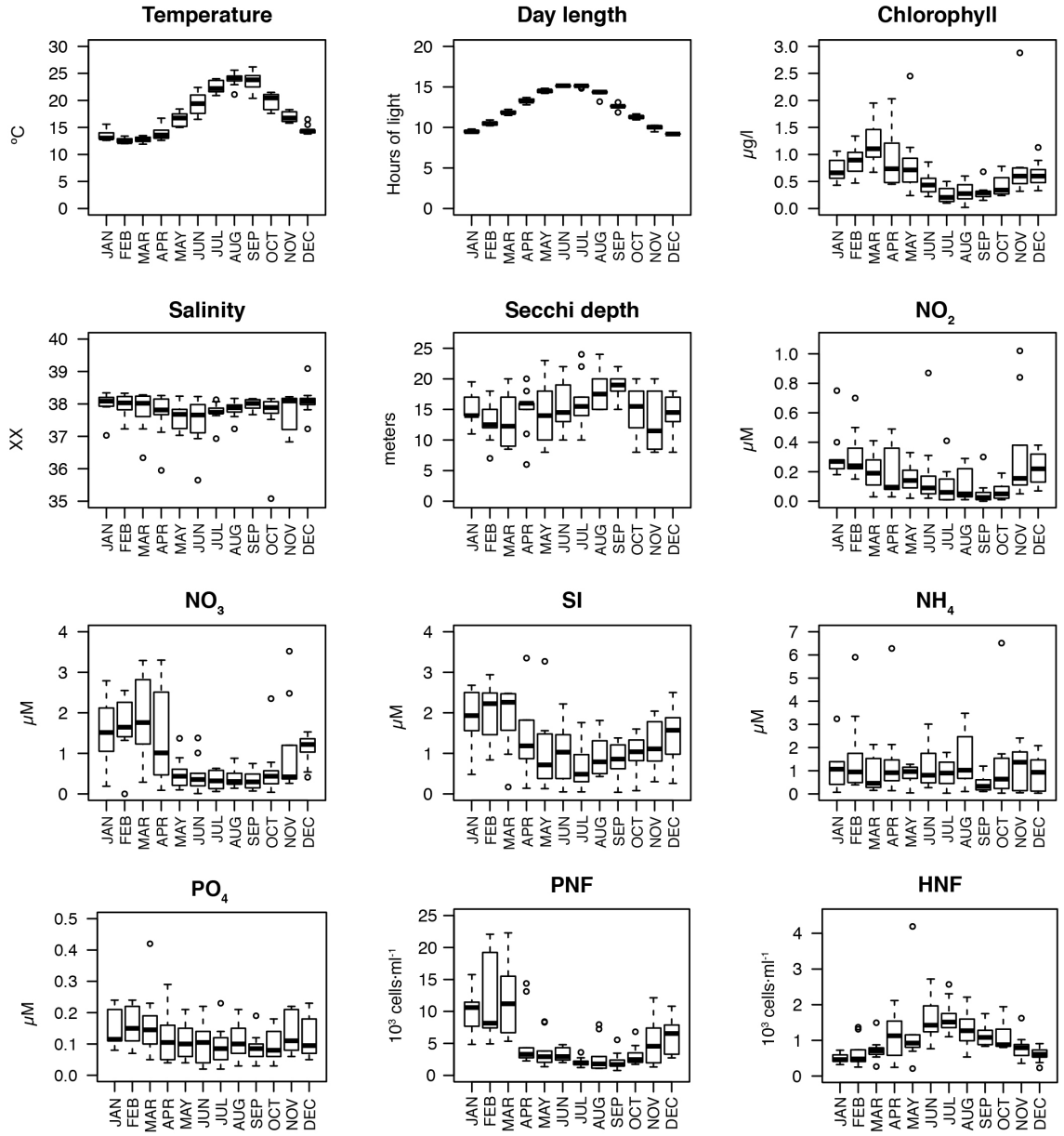


Fig. S1. Seasonal variation of environmental parameters in the Blanes Bay shown as box plots displaying the variability in each month across the 10 years.

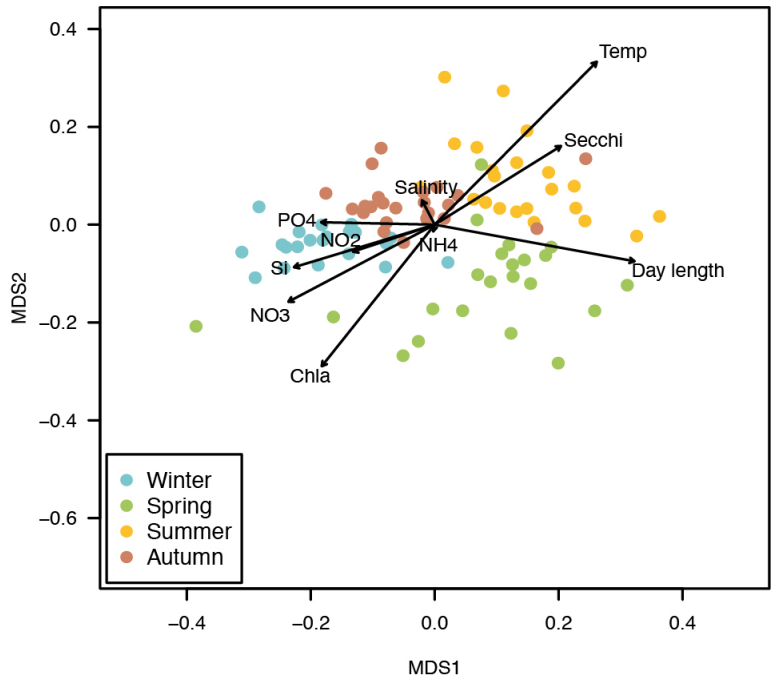


Fig. S2. Non-metric multidimensional analysis (NMDS) based on Bray-Curtis dissimilarities among nanoeukaryotic communities taken at different seasons, showing the environmental vectors that better correlates after an Envfit test.

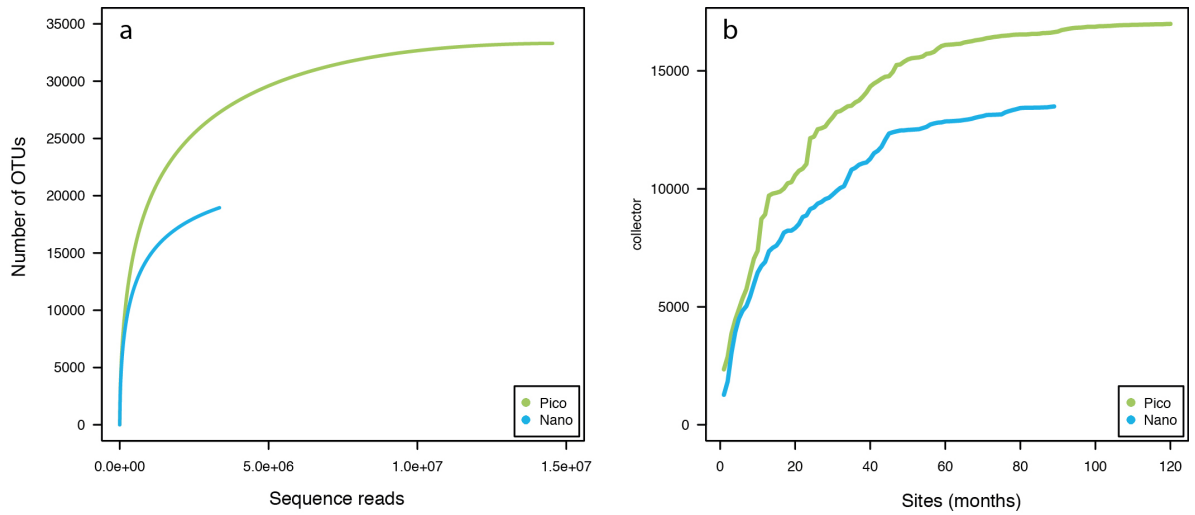


Fig. S3. Rarefaction curves of individual samples of the picoplankton and the nanoplankton fraction together (a). The second panel (b) shows the accumulation curves in both size fractions, i.e. the number of OTUs detected as a function of the number of samples.

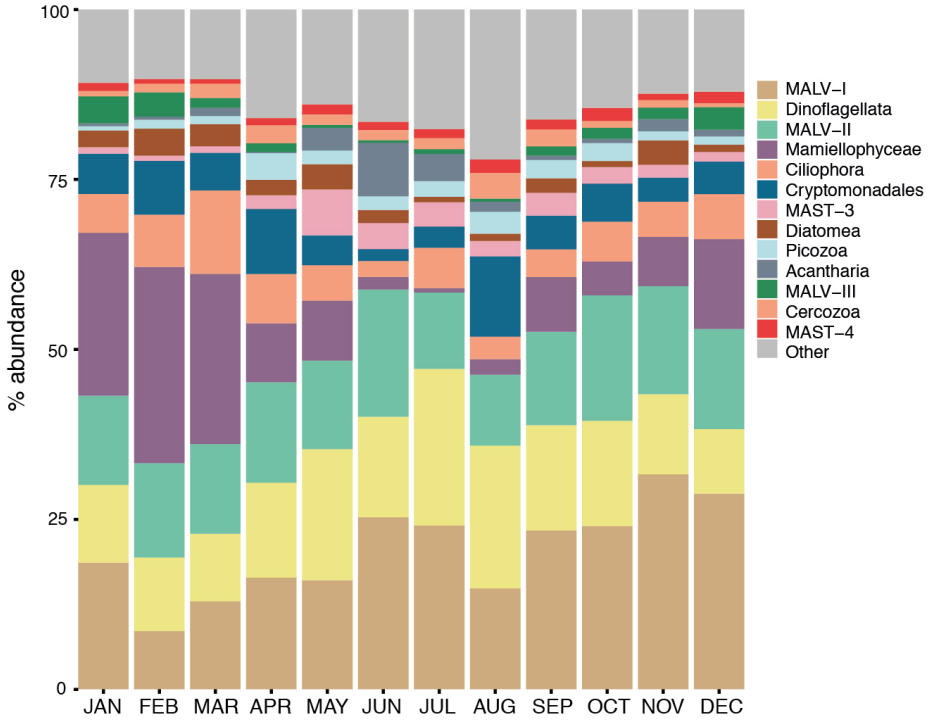


Fig. S4. Monthly changes (averaged for the 10 years) in the relative abundance of the main picoeukaryotic taxonomic groups (only shown groups with relative abundance >1%).

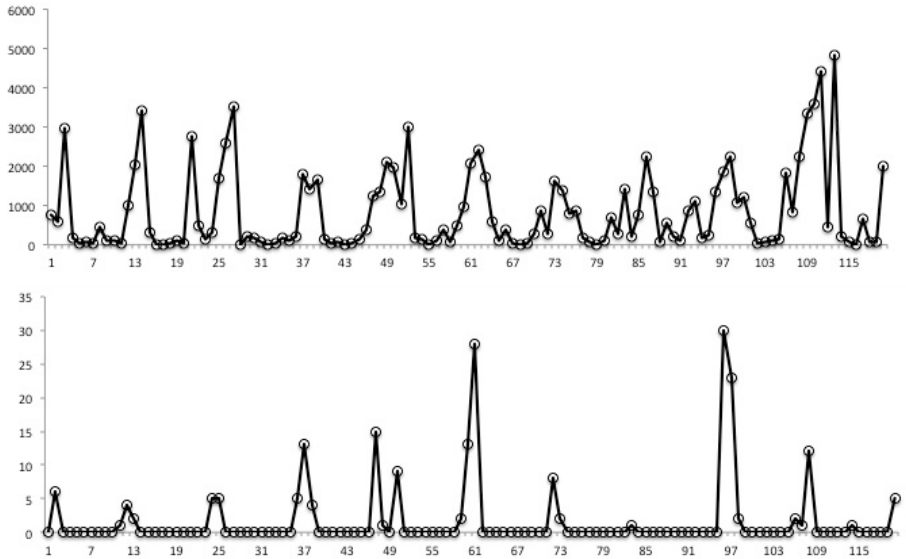


Fig. S5. Examples of different seasonality strategies, the wave strategy of Mamiellophyceae (a), and the pulse strategy of RAD-B (b).

Table S1. Statistics of the ANOSIM test among paired groups of samples taken at each season for picoeukaryotes and nanoeukaryotes. All tests were highly significative with p values below 0.001.

	R picoeukaryotes	R nanoeukaryotes
Winter-Spring	0.525	0.713
Winter-Summer	**0.717**	0.713
Winter-Autumn	0.405	0.384
Spring-Summer	0.232	0.279
Spring-Autumn	0.374	0.531
Summer-Autumn	0.378	0.426

Table S2. Indicator OTUs within picoeukaryotes that are specific for a given season identified using IndVal. The list shows OTUs with an IndVal value above 0.3, highly significant ($p<0.001$) and with a relative abundance >0.05%.

OTUId	Group	Indval	Occurrence	Relative abundance	Taxonomic affiliation
OTU_155	Winter	0.84	33	0.11	MALV-II
OTU_17	Winter	0.81	74	0.84	Ciliophora
OTU_46	Winter	0.76	53	0.37	MALV-II
OTU_216	Winter	0.74	45	0.07	MALV-II
OTU_1	Winter	0.74	98	4.30	Mamiellophyceae
OTU_63	Winter	0.74	57	0.24	Cryptomonadales
OTU_15	Winter	0.71	53	0.51	Diatomea
OTU_104	Winter	0.67	71	0.19	Dinoflagellata
OTU_24284	Winter	0.64	67	0.42	Dinoflagellata
OTU_48	Winter	0.62	55	0.32	Cryptomonadales
OTU_22	Winter	0.60	62	0.62	Mamiellophyceae
OTU_269	Winter	0.59	42	0.06	Pelagophyceae
OTU_2	Winter	0.58	111	3.56	Mamiellophyceae
OTU_7915	Winter	0.58	32	0.05	Telonema
OTU_126	Winter	0.58	44	0.13	MALV-II
OTU_562	Winter	0.57	46	0.07	Pelagophyceae
OTU_31	Winter	0.57	92	0.64	MALV-III
OTU_1395	Winter	0.57	35	0.05	Cryptomonadales
OTU_11782	Winter	0.57	49	0.13	MALV-II
OTU_202	Winter	0.56	66	0.14	MAST-7
OTU_299	Winter	0.56	55	0.06	Dictyochophyceae
OTU_33	Winter	0.55	89	0.44	Mamiellophyceae
OTU_267	Winter	0.55	44	0.07	MALV-II
OTU_242	Winter	0.54	46	0.07	Picozoa
OTU_175	Winter	0.54	49	0.10	Dinoflagellata
OTU_237	Winter	0.54	31	0.06	Cercozoa
OTU_308	Winter	0.52	52	0.05	Dinoflagellata
OTU_324	Winter	0.51	56	0.05	MAST-10
OTU_284	Winter	0.50	47	0.05	Dictyochophyceae
OTU_290	Winter	0.49	22	0.07	MALV-II
OTU_278	Winter	0.47	33	0.07	Ciliophora
OTU_281	Winter	0.47	41	0.06	Dinoflagellata
OTU_174	Winter	0.47	42	0.12	Ciliophora
OTU_326	Winter	0.46	43	0.06	Ciliophora
OTU_65	Winter	0.46	67	0.26	Dinoflagellata

OTU_122	Winter	0.46	61	0.14	MAST-12
OTU_53	Winter	0.45	85	0.32	Ciliophora
OTU_287	Winter	0.45	41	0.07	MALV-II
OTU_32	Winter	0.44	105	0.56	Dinoflagellata
OTU_130	Winter	0.44	40	0.16	Ciliophora
OTU_184	Winter	0.43	37	0.10	Ciliophora
OTU_12156	Winter	0.42	64	0.11	MALV-III
OTU_99	Winter	0.42	54	0.18	MALV-II
OTU_296	Winter	0.41	64	0.05	Bolidomonas
OTU_227	Winter	0.41	53	0.07	MAST-1
OTU_34	Winter	0.40	84	0.44	Pelagophyceae
OTU_55	Winter	0.39	71	0.31	Cryptomonadales
OTU_258	Winter	0.39	69	0.05	MOCH-2
OTU_159	Winter	0.39	50	0.11	MALV-II
OTU_20193	Winter	0.38	60	0.07	Ciliophora
OTU_111	Winter	0.38	83	0.23	MALV-II
OTU_261	Winter	0.36	55	0.06	Telonema
OTU_266	Winter	0.36	68	0.08	MALV-III
OTU_214	Winter	0.34	26	0.09	Mamiellophyceae
OTU_3341	Winter	0.34	55	0.05	Dinoflagellata
OTU_136	Winter	0.32	21	0.12	MALV-II
OTU_24	Spring	0.61	97	0.66	Picozoa
OTU_40	Spring	0.55	63	0.43	MALV-II
OTU_2310	Spring	0.52	32	0.07	Katablepharidae
OTU_18	Spring	0.49	90	0.65	MALV-II
OTU_182	Spring	0.49	25	0.08	Diatomea
OTU_289	Spring	0.48	43	0.05	MAST-3
OTU_98	Spring	0.47	66	0.12	Labyrinthulomycetes
OTU_67	Spring	0.47	100	0.29	Katablepharidae
OTU_75	Spring	0.45	68	0.13	MAST-3
OTU_110	Spring	0.44	77	0.17	Dinoflagellata
OTU_165	Spring	0.43	55	0.10	Chlorarachniophyta
OTU_7	Spring	0.42	71	1.40	Acantharia
OTU_41	Spring	0.42	44	0.42	Ciliophora
OTU_11180	Spring	0.40	25	0.06	Chlorodendrophyceae
OTU_115	Spring	0.40	58	0.12	MAST-3
OTU_8	Spring	0.38	79	1.19	MALV-I
OTU_80	Spring	0.36	16	0.16	Ichthyosporea
OTU_47	Spring	0.36	100	0.38	Dinoflagellata
OTU_282	Spring	0.35	62	0.06	MAST-3
OTU_232	Spring	0.35	42	0.07	MALV-II
OTU_250	Spring	0.34	56	0.09	Dinoflagellata
OTU_274	Spring	0.34	48	0.08	MALV-II
OTU_226	Spring	0.33	31	0.06	MALV-II
OTU_134	Summer	0.62	70	0.16	Ciliophora
OTU_71	Summer	0.60	29	0.32	Chlorodendrophyceae
OTU_197	Summer	0.59	29	0.08	Chlorarachniophyta
OTU_168	Summer	0.59	32	0.17	Chlorodendrophyceae
OTU_167	Summer	0.59	40	0.16	Dinoflagellata
OTU_264	Summer	0.58	43	0.15	Cercozoa
OTU_443	Summer	0.57	39	0.05	MAST-7
OTU_246	Summer	0.54	52	0.09	MAST-9
OTU_57	Summer	0.54	36	0.15	Chlorodendrophyceae
OTU_203	Summer	0.53	38	0.11	Dinoflagellata
OTU_28	Summer	0.53	102	0.57	Dinoflagellata
OTU_4	Summer	0.51	114	2.45	Cryptomonadales
OTU_20	Summer	0.51	87	1.26	MALV-I
OTU_86	Summer	0.50	33	0.19	Unknown Stramenopiles
OTU_43	Summer	0.48	93	0.37	Picozoa

OTU_83	Summer	0.48	52	0.24	Dinoflagellata
OTU_219	Summer	0.47	67	0.08	Dinoflagellata
OTU_3	Summer	0.45	96	2.12	MALV-I
OTU_145	Summer	0.45	63	0.12	Dinoflagellata
OTU_191	Summer	0.45	55	0.09	Bolidomonas
OTU_1495	Summer	0.45	38	0.18	Dinoflagellata
OTU_285	Summer	0.45	61	0.08	Picozoa
OTU_118	Summer	0.44	51	0.14	MAST-11
OTU_340	Summer	0.44	41	0.06	Dinoflagellata
OTU_54	Summer	0.41	90	0.43	Dinoflagellata
OTU_309	Summer	0.41	62	0.12	Dinoflagellata
OTU_430	Summer	0.41	26	0.05	Labyrinthulomycetes
OTU_23	Summer	0.40	96	0.63	Dinoflagellata
OTU_170	Summer	0.40	44	0.08	Chrysophyceae
OTU_229	Summer	0.40	42	0.06	Chrysophyceae
OTU_303	Summer	0.39	57	0.10	Unknown Archaeplastida
OTU_193	Summer	0.39	42	0.08	Cercozoa
OTU_94	Summer	0.39	55	0.15	Chlorarachniophyta
OTU_82	Summer	0.39	92	0.21	MAST-1
OTU_114	Summer	0.39	49	0.17	Choanomonada
OTU_199	Summer	0.39	49	0.10	Dinoflagellata
OTU_206	Summer	0.38	63	0.08	MAST-1
OTU_176	Summer	0.37	70	0.14	Dinoflagellata
OTU_238	Summer	0.37	17	0.10	Dinoflagellata
OTU_2805	Summer	0.37	79	0.07	Dinoflagellata
OTU_30	Summer	0.37	33	0.43	MALV-I
OTU_189	Summer	0.36	49	0.08	Chrysophyceae
OTU_76	Summer	0.36	71	0.25	Centrohelida
OTU_143	Summer	0.36	76	0.13	Telonema
OTU_316	Summer	0.35	19	0.05	Pelagophyceae
OTU_241	Summer	0.35	19	0.08	Cercozoa
OTU_823	Summer	0.35	43	0.10	MALV-I
OTU_97	Summer	0.35	86	0.17	MAST-3
OTU_137	Summer	0.34	44	0.12	MAST-4
OTU_208	Summer	0.34	50	0.07	MAST-7
OTU_100	Summer	0.33	58	0.16	Dinoflagellata
OTU_400	Summer	0.33	12	0.05	MALV-II
OTU_322	Summer	0.33	49	0.06	MAST-3
OTU_333	Summer	0.32	52	0.05	Picozoa
OTU_190	Summer	0.31	36	0.10	MALV-I
OTU_235	Summer	0.31	37	0.07	MALV-II
OTU_14557	Summer	0.31	58	0.05	Dinoflagellata
OTU_109	Summer	0.30	58	0.15	MALV-II
OTU_140	Summer	0.30	35	0.15	Dinoflagellata
OTU_5	Autumn	0.62	98	3.53	MALV-I
OTU_353	Autumn	0.62	31	0.05	MALV-V
OTU_207	Autumn	0.62	28	0.07	MALV-II
OTU_117	Autumn	0.60	55	0.13	MALV-I
OTU_200	Autumn	0.56	28	0.06	MALV-II
OTU_116	Autumn	0.56	60	0.14	Cryptomonadales
OTU_254	Autumn	0.55	45	0.06	MALV-II
OTU_29	Autumn	0.53	66	0.53	MALV-I
OTU_131	Autumn	0.52	58	0.13	MALV-I
OTU_135	Autumn	0.50	69	0.13	Ciliophora
OTU_1862	Autumn	0.47	71	0.14	Mamiellophyceae
OTU_142	Autumn	0.45	52	0.10	Picozoa
OTU_1161	Autumn	0.43	29	0.05	MALV-II
OTU_103	Autumn	0.43	87	0.17	Dinoflagellata
OTU_178	Autumn	0.42	43	0.12	MALV-II

OTU_91	Autumn	0.39	85	0.28	MAST-4
OTU_270	Autumn	0.39	71	0.08	Dinoflagellata
OTU_160	Autumn	0.38	67	0.11	MALV-III
OTU_107	Autumn	0.38	94	0.24	Ciliophora
OTU_60	Autumn	0.37	91	0.23	MOCH-2
OTU_38	Autumn	0.37	29	0.76	MALV-I
OTU_11374	Autumn	0.36	52	0.19	MALV-I
OTU_144	Autumn	0.35	57	0.14	MALV-II
OTU_42	Autumn	0.35	62	0.43	Mamiellophyceae
OTU_128	Autumn	0.35	45	0.12	MALV-V
OTU_222	Autumn	0.35	48	0.07	Labyrinthulomycetes
OTU_121	Autumn	0.34	100	0.32	Dinoflagellata
OTU_217	Autumn	0.34	36	0.05	Acantharia
OTU_260	Autumn	0.34	22	0.07	MALV-II
OTU_108	Autumn	0.33	55	0.15	MALV-I
OTU_16	Autumn	0.33	46	0.76	MALV-II
OTU_265	Autumn	0.33	37	0.06	MALV-II
OTU_223	Autumn	0.33	44	0.09	MALV-II
OTU_105	Autumn	0.33	54	0.19	MALV-II
OTU_150	Autumn	0.32	37	0.14	MALV-I
OTU_305	Autumn	0.31	19	0.06	MALV-II

Table S3. Statistics of the Envfit test evaluating the impact of environmental variables on the NMDS plots shown for picoeukaryotes (Fig. 1b) and nanoeukaryotes (Fig. S2).

Picoeukaryotes

Variable	NMDS1	NMDS2	r2	p-value
Day length	0.845	-0.535	0.623	0.001
Temperature	0.953	0.302	0.563	0.001
Secchi depth	0.866	0.500	0.107	0.002
Salinity	-0.240	0.971	0.134	0.001
Chlorophyll a	-0.912	-0.410	0.188	0.001
PO_4	-0.929	-0.369	0.040	0.086
NH_4	-0.674	-0.739	0.004	0.790
NO_2	-0.978	0.210	0.077	0.010
NO_3	-0.912	-0.411	0.188	0.001
SiO_2	-0.843	-0.538	0.142	0.001

Nanoeukaryotes

Variable	NMDS1	NMDS2	r2	p-value
Day length	0.937	-0.350	0.445	0.001
Temperature	0.881	0.473	0.515	0.001
Secchi depth	0.975	0.222	0.228	0.001
Salinity	-0.409	0.913	0.006	0.789
Chlorophyll a	-0.713	-0.702	0.367	0.001
PO_4	-0.937	-0.350	0.136	0.002
NH_4	-0.124	-0.992	0.002	0.910
NO_2	-0.986	0.166	0.082	0.030
NO_3	-0.953	-0.302	0.296	0.001
SiO_2	-0.972	-0.237	0.235	0.001

Table S4. eLSA analyses showing the percentage of OTUs that have a significant correlation with the different environmental variables.

Variable	Correlation	% total OTUs	% total reads
Day length	Negative	2.88	33.2
	Positive	1.13	15.6
Temperature	Negative	1.77	27.0
	Positive	1.49	20.0
Salinity	Negative	0.15	3.6
	Positive	1.76	24.9
Chlorophyll a	Negative	0.78	12.8
	Positive	1.23	19.7
PO_4	Negative	0.03	0.5
	Positive	0.30	3.4
NH_4	Negative	0.49	4.1
	Positive	0.88	6.4
NO_2	Negative	0.01	0.4
	Positive	1.58	18.4
NO_3	Negative	0.17	9.0
	Positive	1.34	18.8
SiO_2	Negative	1.33	17.2
	Positive	0.89	16.6

Table S5. Relative abundance of the taxonomic groups present in the communities of picoeukaryotes (pico) and nanoeukaryotes (nano) in Blanes Bay. Data derives from the OTU table with both fractions and subsampled at 5,898 reads. Only groups with abundances in the whole dataset above 0.02% are shown.

	Relative abundance (%)				
Group	**Whole dataset**	**Pico**	**Nano**	**Pico/Nano**	**Dominance in nano**
Dinoflagellata	28.21	14.71	46.42	0.32	Nano
MALV-I	18.23	20.39	15.33	1.33	
MALV-II	9.93	14.25	4.12	3.46	
Diatomea	8.07	2.24	15.93	0.14	Nano
Mamiellophyceae	7.68	11.21	2.93	3.83	
Cryptomonadales	4.45	5.75	2.71	2.12	
Ciliophora	4.19	6.00	1.75	3.43	
MAST-3	1.58	2.55	0.28	9.14	
MALV-III	1.56	1.70	1.38	1.23	
Acantharia	1.54	1.91	1.05	1.82	
Cercozoa	1.41	1.71	1.01	1.69	
Picozoa	1.34	2.06	0.37	5.52	
Chlorodendrophyceae	1.12	0.72	1.67	0.43	Nano
MAST-4	0.77	1.27	0.09	13.34	
Telonema	0.65	0.77	0.49	1.57	
Labyrinthulomycetes	0.65	0.96	0.22	4.27	
Chrysophyceae	0.61	0.94	0.16	5.77	
Katablepharidae	0.53	0.84	0.12	7.12	
MAST-7	0.51	0.79	0.13	6.06	
MAST-1	0.50	0.64	0.32	2.03	
Dictyochophyceae	0.49	0.70	0.21	3.38	
Choanomonada	0.48	0.69	0.21	3.33	
Pelagophyceae	0.48	0.72	0.16	4.58	
Chlorarachniophyta	0.35	0.57	0.06	10.34	

Ichthyosporea	0.31	0.43	0.15	2.80	
Centrohelida	0.31	0.44	0.12	3.55	
Prasinophyceae	0.27	0.31	0.20	1.57	
MOCH-2	0.26	0.38	0.08	4.60	
MAST-12	0.25	0.37	0.08	4.43	
MALV-V	0.23	0.30	0.15	1.98	
Unknown Stramenopiles	0.22	0.34	0.06	5.94	
Basal Fungi	0.21	0.25	0.15	1.72	
Ulvophyceae	0.20	0.07	0.38	0.18	Nano
MAST-9	0.19	0.28	0.08	3.72	
Bolidomonas	0.17	0.26	0.05	5.10	
Unknown Archaeplastida	0.16	0.18	0.12	1.49	
MAST-8	0.15	0.23	0.03	7.69	
MAST-2	0.14	0.20	0.06	3.42	
Bicosoecida	0.13	0.21	0.02	10.05	
Rhodophyta	0.13	0.07	0.21	0.34	Nano
Unknown Eukaryota	0.12	0.15	0.09	1.76	
Marine Opisthokonts	0.11	0.17	0.01	12.52	
Polycystinea	0.10	0.11	0.09	1.23	
MAST-11	0.10	0.16	0.01	14.37	
Peronosporomycetes	0.08	0.10	0.04	2.72	
MOCH-5	0.07	0.12	0.01	8.21	
Chlorophyceae	0.07	0.06	0.09	0.69	
Gracilipodida	0.07	0.12	0.01	13.87	
Unknown Alveolata	0.06	0.02	0.11	0.21	Nano
Apicomplexa	0.05	0.01	0.11	0.09	Nano
Unknown Opisthokonta	0.05	0.06	0.03	2.09	
MAST-10	0.04	0.07	0.01	5.39	
MALV-IV	0.04	0.03	0.06	0.56	
Ascomycota	0.04	0.04	0.03	1.29	
MOCH-1	0.04	0.05	0.02	2.38	
Pirsonia	0.04	0.02	0.06	0.36	Nano
RAD-A	0.03	0.03	0.04	0.82	
Unknown Rhizaria	0.03	0.03	0.03	0.79	
Perkinsidae	0.03	0.04	0.01	3.76	
Basidiomycota	0.02	0.03	0.02	1.30	
MOCH-3	0.02	0.02	0.01	2.46	
MAST-6	0.02	0.02	0.01	2.95	
RAD-B	0.02	0.02	0.01	3.24	

Table S6. Seasonality Index (SI) of the main taxonomic picoeukaryotic groups (those with abundance >0.01%). For each group. we removed the OTUs that appeared in less than 10 samples (Uncertain signal) and identified OTUs with SI values above or below 1.2 (seasonal and non-seasonal. respectively). The number of reads accounted for these OTUs are also shown.

				Seasonal Signal		Non-Seasonal Signal		Uncertain Signal	
Group	SI	Relative abundance (%)	Seasonal strategy	OTUs	% reads	OTUs	% reads	OTUs	% reads
MALV-III	2.54	1.70	Wave	15	66.1	22	22.4	343	11.5
Mamiellophyceae	2.16	11.14	Wave	10	80.7	75	17.2	527	2.1
MALV-V	1.70	0.30	Wave	4	33.1	5	56.1	76	10.8
Pelagophyceae	1.65	0.74	Wave	3	77.4	8	17.1	87	5.6
MAST-11	1.55	0.16	Wave	1	85.9	1	12.8	2	1.3
RAD-B	1.44	0.02	Pulse	0	0.0	3	61.2	8	38.8
Chlorarachniophyta	1.42	0.56	Wave	5	52.7	11	32.6	68	14.7
RAD-A	1.40	0.03	Pulse	1	32.5	2	46.7	12	20.8
Polycystinea	1.32	0.11	Pulse	1	17.0	5	54.2	34	28.8
MAST-25	1.30	0.01	Pulse	0	0.0	1	75.9	2	24.1
MOCH-2	1.25	0.38	Wave	3	81.7	5	11.8	30	6.6
Dinoflagellata	1.24	14.70	Wave	44	22.9	328	63.5	2994	13.6
Centrohelida	1.22	0.44	Wave	1	55.9	5	19.9	27	24.2
Ichthyosporea	1.15	0.44	-	1	9.6	4	72.4	38	18.0
MAST-3	1.14	2.52	-	8	24.3	57	71.1	181	4.6
Chrysophyceae	1.13	0.96	-	7	31.2	23	52.6	82	16.3
Acantharia	1.13	1.91	-	1	73.5	16	19.4	147	7.0
Perkinsidae	1.13	0.04	-	0	0.0	1	27.1	27	72.9
Chlorodendrophyceae	1.12	0.72	-	0	0.0	5	96.1	71	3.9
MAST-9	1.10	0.29	-	1	8.7	12	82.7	29	8.6
MAST-1	1.10	0.64	-	4	51.4	12	40.6	61	7.9
MALV-IV	1.09	0.03	-	0	0.0	1	93.8	4	6.2
Katablepharidae	1.09	0.85	-	0	0.0	10	98.5	34	1.5
MOCH-3	1.09	0.02	-	0	0.0	1	90.3	2	9.7
Labyrinthulomycetes	1.08	0.96	-	3	10.7	17	54.0	133	35.3
MAST-8	1.08	0.24	-	3	9.0	7	81.8	33	9.2
MAST-6	1.07	0.02	-	0	0.0	2	67.3	5	32.7
Ciliophora	1.07	5.95	-	16	9.2	123	77.0	584	13.8
Basidiomycota	1.07	0.02	-	0	0.0	1	26.9	8	73.1
Prasinophyceae	1.06	0.32	-	2	10.4	12	63.3	59	26.3
Bicosoecida	1.05	0.22	-	1	9.9	4	58.5	21	31.5
Pirsonia	1.05	0.02	-	0	0.0	1	26.8	11	73.2
MOCH-1	1.04	0.06	-	0	0.0	6	87.3	13	12.7
MOCH-5	1.04	0.12	-	0	0.0	4	100.0		0.0
Choanomonada	1.04	0.69	-	3	32.1	22	55.6	69	12.3
Cercozoa	1.03	1.69	-	3	5.6	66	64.9	336	29.5
Rhodophyta	1.03	0.07	-	0	0.0	2	26.8	44	73.2
MALV-I	1.02	20.40	-	33	42.4	267	51.3	1674	6.3
Nephroselmis	1.02	0.01	-	0	0.0	1	92.6	2	7.4
MAST-10	1.02	0.06	-	1	71.8	1	28.1	1	0.2
Chlorophyceae	1.01	0.06	-	0	0.0		0.0	6	100.0
Ulvophyceae	1.00	0.07	-	0	0.0	1	60.0	12	40.0
Gracilipodida	1.00	0.11	-	0	0.0	1	99.7	1	0.3
Marine Opisthokonts	1.00	0.18	-	0	0.0	4	63.6	42	36.4
Peronosporomycetes	1.00	0.11	-	0	0.0	1	6.0	39	94.0
Basal Fungi	0.97	0.27	-	0	0.0	3	82.3	17	17.7
Cryptomonadales	0.96	5.75	-	10	49.5	30	49.2	216	1.3
Picozoa	0.96	2.05	-	7	37.4	20	60.3	113	2.3

Ascomycota	0.96	0.04	-	0	0.0	2	37.5	11	62.5
Diatomea	0.95	2.25	-	2	22.7	41	61.2	325	16.1
MAST-4	0.95	1.30	-	3	33.7	12	62.5	147	3.8
MALV-II	0.95	14.27	-	30	14.1	308	70.1	1710	15.8
MAST-7	0.94	0.77	-	6	45.4	6	49.2	63	5.4
MAST-2	0.93	0.21	-	0	0.0	5	94.8	12	5.2
Telonema	0.93	0.78	-	3	18.6	16	74.7	83	6.7
Bolidomonas	0.93	0.27	-	3	49.5	6	46.7	13	3.9
MAST-12	0.92	0.37	-	1	10.6	7	73.1	49	16.3
Dictyochophyceae	0.92	0.69	-	7	46.8	22	43.6	77	9.6

REFERENCES

Allison, S.D. & Martiny, J.B. (2008). Resistance, resilience, and redundancy in microbial communities. Proc Natl Acad Sci U S A, 105 Suppl 1, 11512-11519.

Alonso-Sá´ez, L., Balague, V., Sa, E.L., Sanchez, O., Gonzalez, J.M., Pinhassi, J. et al. (2007). Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. FEMS microbiology ecology, 60, 98-112.

Alonso-Sáez, L., Diaz-Perez, L. & Moran, X.A. (2015). The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. Environ Microbiol, 17, 3766-3780.

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

Andersson, A.F., Riemann, L. & Bertilsson, S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. ISME J, 4, 171-181.

Campbell, B.J., Yu, L., Heidelberg, J.F. & Kirchman, D.L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 12776-12781.

Clarke, K.R. (1993). Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18, 117-143.

Countway, P.D., Vigil, P.D., Schnetzer, A., Moorthi, S.D. & Caron, D.A. (2010). Seasonal analysis of protistan community structure and diversity at the USC Microbial Observatory (San Pedro Channel, North Pacific Ocean). Limnology and Oceanography, 55, 2381-2396.

Cram, J.A., Chow, C.E., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A. et al. (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J, 9, 563-580.

Dufrene, M. & Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs*, 67, 345-366.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26, 2460-2461.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods, 10, 996-998.

Foulon, E., Not, F., Jalabert, F., Cariou, T., Massana, R. & Simon, N. (2008). Ecological niche partitioning in the picoplanktonic green alga Micromonas pusilla: evidence from environmental surveys using phylogenetic probes. *Environmental microbiology*, 10, 2433-2443.

Fuhrman, J.A., Cram, J.A. & Needham, D.M. (2015). Marine microbial community dynamics and their ecological interpretation. Nature reviews. Microbiology, 13, 133-146.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V. & Naeem, S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences of the United States of America, 103, 13104-13109.

Gasol, J.M., Cardelús, C., G. Morán, X.A., Balagué, V., Forn, I., Marrasé, C. et al. (2016). Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. Scientia Marina, 80, 63-77.

Genitsaris, S., Monchy, S., Viscogliosi, E., Sime-Ngando, T., Ferreira, S. & Christaki, U. (2015). Seasonal variations of marine protist community structure based on taxon-specific traits using the eastern English Channel as a model coastal system. FEMS Microbiol Ecol, 91.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbruck, L., Reeder, J., Temperton, B. et al. (2012). Defining seasonal marine microbial community dynamics. ISME J, 6, 298-308.

Grasshoff, K., Ehrhardt, M. & Kremling, K. (1983). *Methods on seawater analysis*. 2n ed. Winheim: Verlag Chemie.

Guadayol, Ò., Peters, F., Marrasé, C., Gasol, J.M., Roldán, C., Berdalet, E. et al. (2009). Episodic meteorological and nutrient-load events as drivers of coastal planktonic ecosystem dynamics: a time-series analysis. Marine Ecology Progress Series, 381, 139-155.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L. et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res, 41, D597-604.

Kim, D.Y., Countway, P.D., Jones, A.C., Schnetzer, A., Yamashita, W., Tung, C. et al. (2014). Monthly to interannual variability of microbial eukaryote assemblages at four depths in the eastern North Pacific. ISME J, 8, 515-530.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology*. 2n edn. Elsevier Science BV, Amsterdam.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. et al. (2004). The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters, 7, 601-613.

Leray, M. & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ, 5, e3006.

Lindh, M.V., Sjostedt, J., Andersson, A.F., Baltar, F., Hugerth, L.W., Lundin, D. et al. (2015). Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. Environmental microbiology, 17, 2459-2476.

Lindström, E.S. & Langenheder, S. (2012). Local and regional factors influencing bacterial community assembly. Environ Microbiol Rep, 4, 1-9.

Logares, R. (2017). Workflow for Analysing MiSeq Amplicons based on Uparse v1.5. https://doi.org/10.5281/zenodo.259579.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R. et al. (2014). Patterns of rare and abundant marine microbial eukaryotes. Current biology : CB, 24, 813-821.

Logares, R., Lindstrom, E.S., Langenheder, S., Logue, J.B., Paterson, H., Laybourn-Parry, J. *et al.* (2013). Biogeography of bacterial communities exposed to progressive long-term environmental change. *The ISME journal*, 7, 937-948.

Logares, R., Mangot, J.F. & Massana, R. (2015). Rarity in aquatic microbes: placing protists on the map. Research in microbiology, 166, 831-841.

Massana, R., Balague, V., Guillou, L. & Pedros-Alio, C. (2004). Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. FEMS microbiology ecology, 50, 231-243.

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C. et al. (2015). Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ Microbiol, 17, 4035-4049.

Nikolenko, S.I., Korobeynikov, A.I. & Alekseyev, M.A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics, 14 Suppl 1, S7.

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

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P. *et al.* (2008). vegan: Community Ecology Package. R package version 1.15-0.

Pernice, M.C., Logares, R., Guillou, L. & Massana, R. (2013). General patterns of diversity in major marine microeukaryote lineages. PLoS One, 8, e57170.

Piredda, R., Tomasino, M.P., D'Erchia, A.M., Manzari, C., Pesole, G., Montresor, M. et al. (2017). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. FEMS microbiology ecology, 93.

Porter & Feig (1980). The use of DAPI for identifying and counting aquatic microflora. Limnology and Oceanography.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res, 41, D590-596.

Ricklefs, R.E. (2007). History and diversity: Explorations at the intersection of ecology and evolution. American Naturalist, 170, S56-S70.

Roberts, D.W. (2016). labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.8-0.

Romari, K. & Vaulot, D. (2004). Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. Limnology and Oceanography.

Ruan, Q., Dutta, D., Schwalbach, M.S., Steele, J.A., Fuhrman, J.A. & Sun, F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics*, 22, 2532-2538.

Schauer, M., Balague, V., Pedros-Alio, C. & Massana, R. (2003). Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. Aquatic Microbial Ecology, 21.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res, 43, e37.

Shade, A. & Gilbert, J.A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. Trends Microbiol, 23, 335-340.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N. et al. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio, 5, e01371-01314.

Simon, M., Lopez-Garcia, P., Deschamps, P., Restoux, G., Bertolino, P., Moreira, D. et al. (2016). Resilience of Freshwater Communities of Small Microbial Eukaryotes Undergoing Severe Drought Events. Front Microbiol, 7, 812.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W. *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular ecology*, 19 Suppl 1, 21-31.

Suzuki, R. & Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics, 22, 1540-1542.

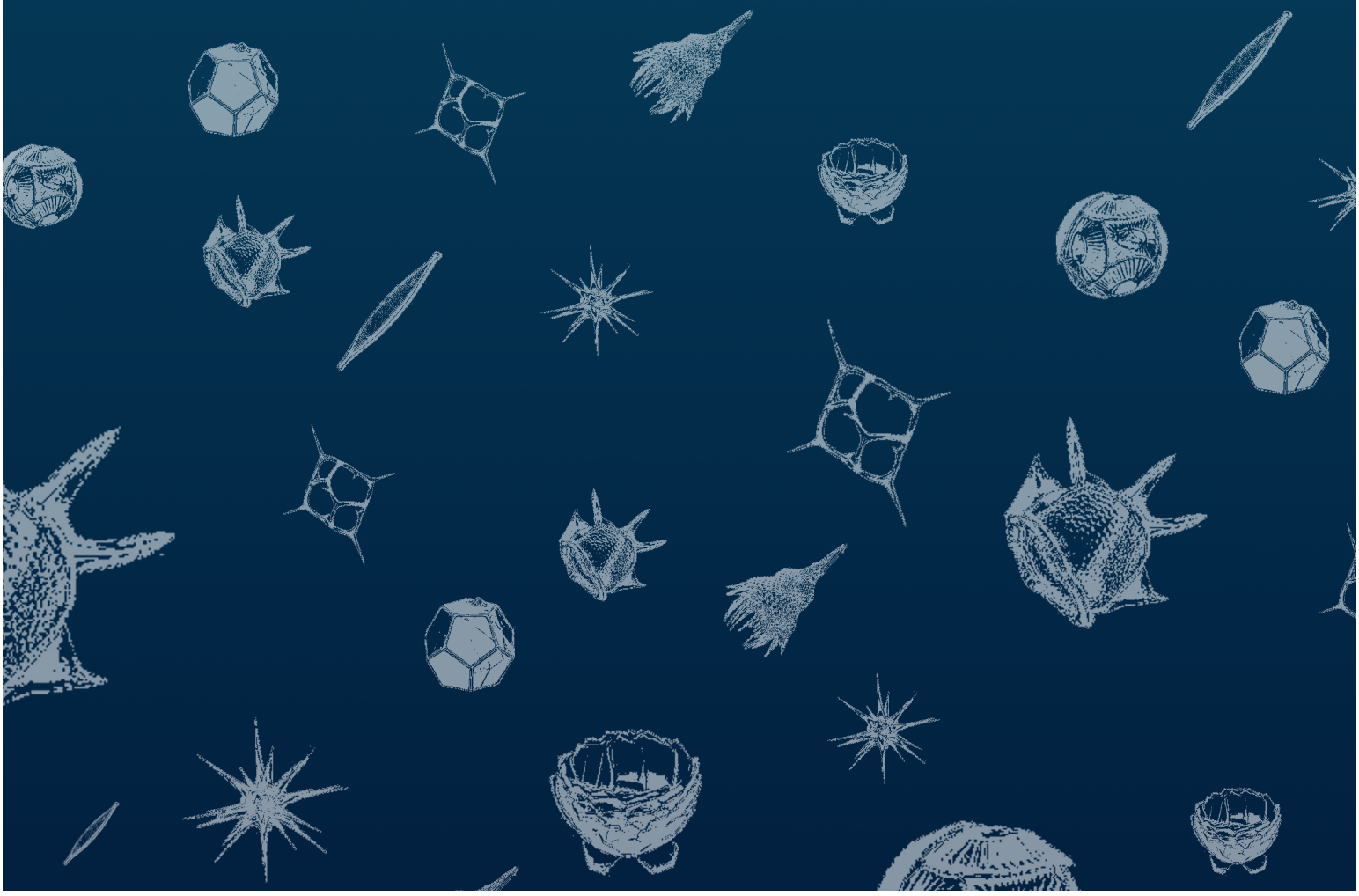
Xia, L.C., Steele, J.A., Cram, J.A., Cardon, Z.G., Simmons, S.L., Vallino, J.J. et al. (2011). Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. BMC Systems Biology.

Yentsch, C.S. & Menzel, D.W. (1963). A method for the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. Deep Sea Research 10, 221-231.

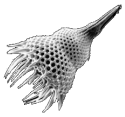
Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30, 614-620.

Chapter 2

Global changes in activity and community structure of marine picoeukaryotes through the water column

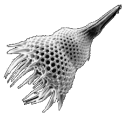


Caterina R. Giner, Vanessa Balagué, Massimo C. Pernice, Aleix Obiol, J.M. Gasol, Ramiro Logares and Ramon Massana



ABSTRACT

Microbial eukaryotes have been described from all marine habitats, however, there is a lack of understanding on how their metabolic activity and diversity change along the marked environmental gradients present in the ocean water column (i.e. from 0 to > 4,000m depth). Here, we analyzed the vertical distribution and metabolic activity of the smallest microeukaryotes, the picoeukaryotes, through the ocean water column at a global scale. Communities inhabiting surface, deep chlorophyll maximum, mesopelagic and bathypelagic realms were analysed using Illumina high-throughput sequencing of the 18S rDNA. Both DNA and rRNA environmental extracts were considered in the analyses. In addition, we used rDNA fragments extracted from metagenomes (a.k.a. miTags) to analyze groups that may be underrepresented in the previous analysis due to primer biases. Our results demonstrate a clear stratification of picoeukaryotic communities along the vertical gradient, with two differentiated assemblages corresponding to the sunlit and the dark ocean. We identified marked changes in the relative abundance of the main picoeukaryotic groups with depth. Furthermore, using the rRNA:rDNA ratio for each individual OTU, we found changes in the metabolic activity of multiple taxa. The mesopelagic was the layer where we identified the highest metabolic activity, whereas in the bathypelagic, most groups displayed the lowest activity. Overall, our results increase our knowledge of picoeukaryotic communities in the global ocean from surface to deep ocean taking special attention in activity changes.



INTRODUCTION

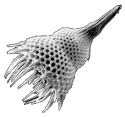
Microorganisms play fundamental roles in the global biogeochemical cycles, however the mechanisms that drive their distribution and community structure at a global scale are still poorly known. The small size of the microorganisms allows them to be carried passively by currents throughout the world oceans, which leads to the idea that microorganisms are ubiquitously dispersed and that environmental selection determines what taxa grows in specific habitats. The latter is summarized by the tenet 'Everything is everywhere but the environment selects' (Baas-Becking 1934). The marine habitat is characterized by two well-marked dimensions with contrasting environmental properties: (i) the horizontal dimension, in which the environment could be very similar throughout large distances (Agogué *et al.* 2011), and (ii) the vertical dimension, where marked environmental gradients occur in short distances, e.g. the light profile, the thermocline or the nutricline. In the latter dimension, active movement can be as important as passive dispersal, and some species regulate the preferred depth by buoyancy or vertical migration (Weiner *et al.* 2012). These two dimensions present different constrains to dispersal and may promote contrasting distributional patterns.

The biogeochemical importance of microbes in the whole marine water column has become noticeable during the last decades (Nagata *et al.* 2010). The water column is divided in different realms: the epipelagic (0-200 m depth), the mesopelagic (200-1000 m) and the bathypelagic (1000-4000 m). The sunlit epipelagic region harbors the photosynthetic microbes and therefore it represents the beginning of the ocean food web. The dark ocean (i.e. >200 meters depth), comprising the mesopelagic and bathypelagic zones is the largest and less known habitat on Earth. Compared to the epipelagic zone, it has higher pressure, lower temperature and a higher content of inorganic nutrient (Arístegui *et al.* 2009). The two realms of the dark ocean are well differentiated in terms of biogeochemical processes. The mesopelagic zone is subjected to strong episodic inputs of organic matter from the euphotic zone and is where most of the sedimenting organic carbon (~90%) is respired back to carbon dioxide (Robinson *et al.* 2010). Furthermore the mesopelagic zone contains layers with very low oxygen concentration, the oxygen minimum zones, which play a key role in the nitrogen cycle (Robinson *et al.* 2010). On the other hand, the bathypelagic is a much more stable environment, rich in oxygen and inorganic nutrients, which receives organic carbon already remineralizated or transformed by microorganisms in the mesopelagic region. Overall, the dark ocean is the largest reservoir of organic carbon in the biosphere (Nagata *et al.* 2010) and contains about 70% of the ocean's microbial cells. Despite their importance, most studies of marine microbes have targeted the sunlit ocean (0-200 m), as

the dark ocean is more difficult to sample. A large part of these efforts have focused on identifying the key microbial players and the factors that constrain their abundance and biogeography.

Protists are recognized as important members of microbial communities playing central roles in marine food webs (Sherr & Sherr 2002) and carbon cycling (Arístegui *et al.* 2009), but their diversity and biogeography throughout the water column has been poorly analyzed. Initial studies were done at a regional scale focusing on the water column (López-García *et al.* 2001; Countway *et al.* 2007; Not *et al.* 2007; Brown *et al.* 2009) or on specific environments like hydrothermal vents (Edgcomb *et al.* 2002), anoxic basins (Stoeck *et al.* 2003) and deep-sea sediments (Edgcomb 2011). These studies highlighted a well-marked difference between epipelagic and deep communities. Consequently, epipelagic assemblages from different sites were more similar than assemblages from the same site but from different depths. Recently, global scale protist studies at the photic (de Vargas *et al.* 2015) and aphotic zones (Pernice *et al.* 2016) have been published, increasing our knowledge about diversity and distribution of marine microeukaryotes. These studies give a fundamental bridge between regional and global views. Within the photic zone, de Vargas et al. (2015) showed that heterotrophic protists were more diverse than phototrophic ones, and that communities were structured according to their oceanographic basin of origin. Within the bathypelagic zone, Pernice et al. (2016) highlighted the importance of water masses as a force structuring deep protists communities, thus adjacent water masses with different environmental properties contained different microbial assemblages.

Most biodiversity studies on microbial eukaryotes so far have been based on sequencing 18S rRNA genes (rDNA) using environmental DNA extracts as templates, which gives information about the organisms present in a given sample. These studies have been sometimes complemented by parallel analyses targeting the same gene but using environmental RNA extracts instead (Not *et al.* 2009; Massana *et al.* 2015). This comparison can first identify taxonomic groups where the ratio of the number of rDNA genes and ribosomes are very different from other taxa at a structural level, as it seems to happen in some alveolate groups (Massana *et al.* 2015). Second, the rRNA:rDNA ratio can be used as a proxy of specific activity, particularly if this ratio varies under specific conditions, as ribosomes are needed to increase the protein synthesis capacity during the growth and acclimation of a given population (Blazewicz *et al.* 2013).



Specifically among microbial eukaryotes, picoeukaryotes (0.2-3 μm in size) have been recognized as the most abundant eukaryotes in the marine plankton, being very diverse, widely distributed and ecologically important (Massana 2011). Here we present the first global survey to investigate changes in picoeukaryotic communities throughout the water column by Illumina sequencing of 18S rRNA genes amplified from DNA and RNA extracts as well as directly extracted from metagenomes. We analyzed seven depths along 13 vertical profiles retrieved in Atlantic, Indian and Pacific Oceans during the Malaspina-2010 Circumnavigation expedition. In this study we aim to respond two principal questions: (i) How the community structure of marine picoeukaryotes change through the water column and what are the environmental drivers? (ii) Do taxonomic groups change their relative abundance and activity with depth?

MATERIALS AND METHODS

Sample collection and nucleic acid extraction

During the Malaspina 2010 Circumnavigation expedition (December 2010 – July 2011), a total of 91 water samples were collected in 13 stations globally distributed across the world's oceans (Fig. 1). Each station was sampled at 7 different depths with Niskin bottles attached to a CTD profiler that had sensors for conductivity, temperature, salinity and oxygen. Each vertical profile included samples at surface (3 m), at DCM (Deep Chlorophyll Maximum), and at 2-3 depths in mesopelagic (200-1000 m) and bathypelagic waters (1000-4000 m). For each sample, typically 12 liters of seawater were prefiltered through a 200 μm nylon mesh to remove large plankton and then sequentially filtered using a peristaltic pump through a 20 μm nylon mesh (at the entrance of the tubing) and 3 μm and 0.2 μm polycarbonate filters of 142 mm diameter (Isopore, Millipore). Filtration time was about 15-20 minutes. The filters were flash frozen in liquid nitrogen and stored at -80ºC until DNA and RNA extraction.

Samples for inorganic nutrients (NO_3-, NO_2-, PO_4^{3-}, SiO_2) were collected from the Niskin bottles, kept frozen, and measured spectrophotometrically using an Alliance Evolution II autoanalyzer (Grasshoff *et al.* 1983). Bacterial and picoeukaryotic abundance were estimated by a combination of flow cytometry as explained in (Pernice *et al.* 2015) and epifluorescence microscopy. Along the cruise, different water masses were sampled. The proportion of the different water masses in each deep ocean sample was inferred from its temperature, salinity and oxygen concentration (Catalá *et al.* 2015).

Prior to nucleic acid extraction, filters were cut in small pieces and cryogrinded with a Freezer-Mill 6770 (Spex) for 3 cycles of 1 minute. Then, RNA and DNA were extracted simultaneously using the Nucleospin RNA kit (Macherey-Nagel) plus the NucleoSpin RNA/DNA buffer Set (Macherey-Nagel) procedures. The existence of residual DNA in RNA extracts was checked by PCR with universal eukaryotic primers and, if detected, was subsequently removed using the Turbo DNA-free kit (Applied Biosystems). RNA was reverse transcribed using the SuperScript III reverse Transcriptase (Invitrogen) and random hexamers. DNA and RNA extracts were quantified with a Qubit 1.0 (Thermo Fisher Scientific).

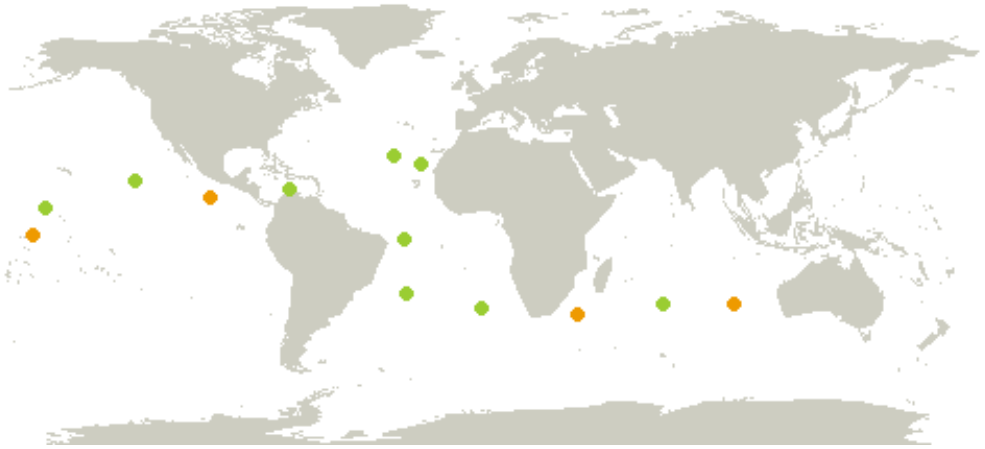
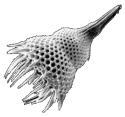


Fig. 1. World map showing the location of the Malaspina stations sampled for this study. Green dots indicate stations only with amplicon sequencing whereas orange dots indicate stations with amplicon sequencing and metagenomes (miTags).

Sequencing and processing of picoeukaryotic community composition

Eukaryotic diversity was assessed by amplicon sequencing of the V4 region of the 18S rDNA gene (~380 bp) using the Illumina MiSeq platform and paired-end reads (2x250 bp). PCR amplifications with the eukaryotic universal primers TAReuk454FWD1 and TAReukREV3 (Stoeck et al. 2010) and amplicon sequencing were carried out at the Research and Testing Laboratory (Lubbock, TX, USA; http://www.researchandtesting.com). Illumina reads obtained from both DNA and RNA extracts (rDNA and rRNA samples, respectively) were processed together following an in-house pipeline (Logares 2017) at the Marine Bioinformatics Service (MARBITS) of the Institut de Ciències del Mar (ICM-CSIC) in Barcelona. Briefly, raw reads were corrected using BayesHammer (Nikolenko et al. 2013) as indicated by Schirmer et al. (Schirmer et al. 2015). Corrected paired-end reads were subsequently merged with PEAR (Zhang *et al.*



2014) and sequences longer than 200 bp were quality-checked and dereplicated using USEARCH (Edgar 2010). OTU clustering at 99% similarity was done using UPARSE v8 (Edgar 2013). Chimera check and removal was performed both de novo and using the SILVA reference database (Quast et al. 2013). Taxonomic assignment was obtained by a BLAST search against three reference databases, PR2 (Guillou et al. 2013) and two in-house marine protist databases (available at https://github.com/ramalok) based in a collection of Sanger sequences from molecular surveys (Pernice et al. 2013) and on 454 reads from the BioMarKs project (Massana et al. 2015). Metazoan, Charophyta and nucleomorphs OTUs were removed. The final OTU table contained 79 rDNA samples (12 samples were removed due to failing PCR or sequencing reactions) and 91 rRNA samples. To enable comparisons between samples, the OTU table was randomly subsampled down to the minimum number of reads per sample (22,379 reads) using the *rrarefy* function in the *Vegan* package (Oksanen *et al.* 2015).

Statistical analyses

Statistical analyses were performed with the R Statistical Software (R Core Team 2015) and *Vegan* package (Oksanen *et al.* 2015). Bray-Curtis dissimilarities were used as an estimator of beta diversity between communities, which were then clustered using non-metric multidimensional scaling (NMDS). In NMDS, the differences between predefined groups were statistically tested with ANOSIM using 1000 permutations. PERMANOVA analyses were performed to determine the proportion of the variation in community composition that was explained by the measured environmental variables. Shannon index (H') and richness (number of OTUs) were calculated as an estimator of eukaryotic diversity.

Comparison with metagenomes (miTags)

For 4 of the 13 vertical profiles we had metagenomic samples for comparison (Fig. 1), obtained with slight modifications of the filtration set-up and DNA extraction. Filtration used the same 200 µm-prefiltered seawater and peristaltic pumping, except that the 3 and 0.2 µm filters were of 47 mm diameter and filtration time was about 2 hours. DNA extraction was not done with a kit but followed the phenol-chlorophorm protocol as explained elsewhere (Pernice *et al.* 2016). Metagenomes were sequenced at the CNAG (http://www.cnag.crg.eu/) with Ilumina HiSeq 2000, yielding about 30 Gb of sequencing information per metagenome.

In order to recruit 18S rDNA metagenomic reads (miTags), a reference database of the V4 region was created combining OTUs from our metabarcoding analysis with SILVA sequences for the

groups absent in the PCR approach. This V4 reference dataset was clustered at 97% and inspected to discard chimeras, yielding a final database of 9,733 sequences. MiTags from 25 metagenomes were extracted following the published protocol (Logares *et al.* 2014b) with minor modifications. Quality-filtered Illumina reads longer than 70 bp were mapped to the V4 database in USEARCH (Edgar 2010) at 97% identity, 90% minimum query coverage and the *top_hits_only* option to retrieve all hits with the same highest score. Retrieved miTags were categorized as (i) those with only one hit to the V4-database (ii) those with 2 or more hits to sequences of the same taxonomic group and (iii) those with 2 or more hits to sequences of different taxonomic groups. MiTags from the first and second cases were kept (97.8%) in the final OTU table. In order to compare the metagenomic and metabarcoding approaches, taxonomic groups represented only in metagenomes were removed, resulting in a final OTU table of 40,222 miTags (1,608 per sample on average; minimal value 470).

RESULTS

The complete dataset (rDNA and rRNA) contained 11,712,170 reads clustered into a total of 45,175 OTUs featuring a minimum of 99% similarity. This dataset describes picoeukaryotic diversity along 13 vertical profiles in the three main oceans (Fig. 1). The individual rarefaction curves (not shown) indicated that most of the samples did not show saturation with the used sequencing effort. When pooling samples from the same depth layer, rarefaction curves from mesopelagic and bathypelagic layers were more saturated, whereas surface and DCM were still far from saturation (Fig. 2a), possibly due to the different number of samples from these two groups (13 samples for surface and DCM and about 30 samples for mesopelagic and bathypelagic). A sample based accumulation curve revealed that OTUs increased rapidly with the first ten samples and after that the discovery of new OTUs was slower (Fig. 2b). The increase in the number of OTUs per sample depended on depth, being surface and DCM the layers with faster increase. This indicated that globally, picoeukaryotic diversity was higher in the euphotic zone than in the deep dark ocean.

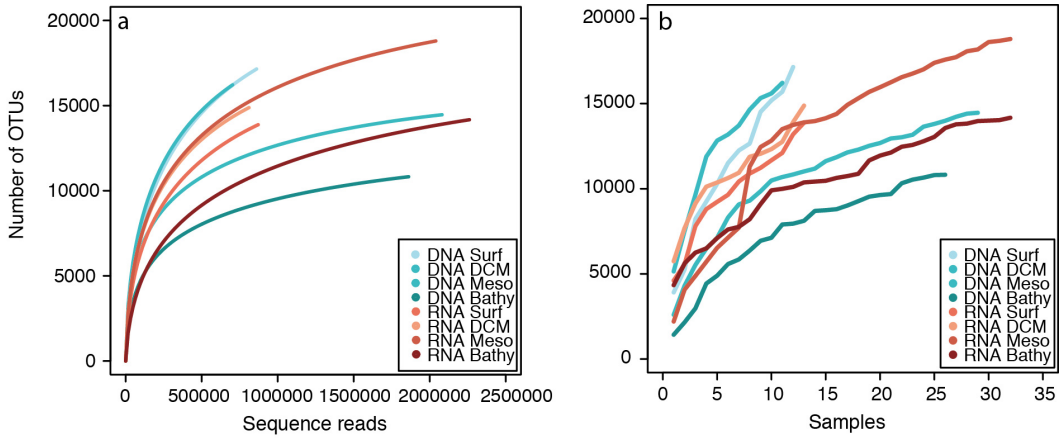


Fig. 2. Saturation of the molecular survey (based on DNA or RNA extracts) in samples grouped into the main four vertical layers. (a) Rarefaction curves relating the number of OTUs detected in function of the sequencing effort. (b) Accumulation curves relating the number of OTUs to the number of samples analyzed.

For further analyses, the OTU table was subsampled down to 22,379 reads per sample, yielding a final table retaining 38,343 OTUs and 3,804,430 reads. As indicated by rarefaction plots, diversity estimates followed a directional water column trend, with richness and Shannon Indices being highest in surface waters and decreasing with depth (Fig. 3). Thus, the lowest diversity values were observed at the bathypelagic layer, with a median richness below 1000 OTUs. The differences in richness between the epipelagic and the deep ocean layers was significant (Wilcoxon test $p<0.05$)

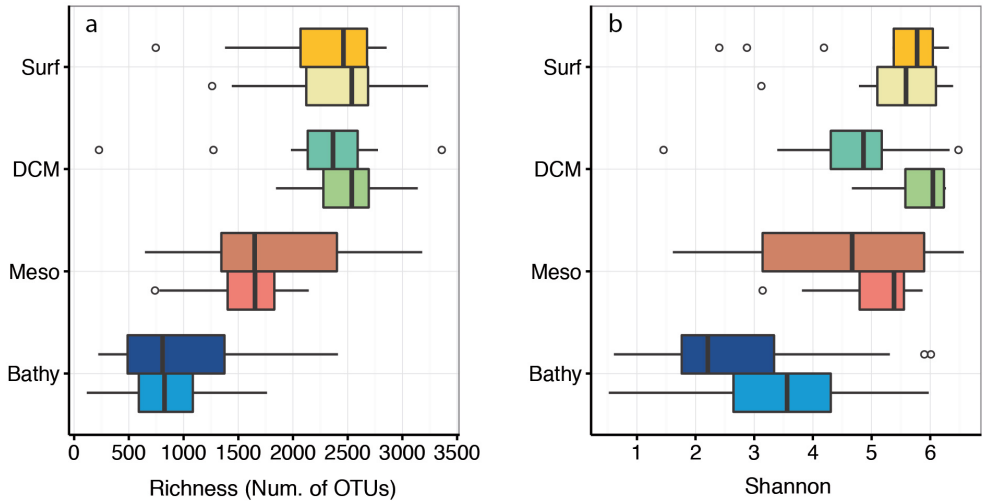


Fig. 3. Changes in alpha diversity in the different water layers using the rRNA (upper boxplots) and rDNA (lower boxplots) datasets. (a) OTU richness and (b) Shannon Index (H'). Significant differences were found between photic and aphotic layers.

Structure of picoeukaryotic communities in the water column

We observed a clear segregation of picoeukaryotic communities based on the water layer, with a striking differentiation among photic (surface and DCM) and aphotic (meso- and bathypelagic) communities in the NMDS (Fig. 4a, $ANOSIM_{photic\text{-}aphotic}$ R=0.513, p<0.001). Within them, the two aphotic layers did not form clear groups ($ANOSIM_{mesopelagic\text{-}bathypelagic}$ R=0.151, p<0.004) and tended to be intermixed in the NMDS plot (Fig. 4b), whereas photic samples from surface and DCM formed two differentiated groups ($ANOSIM_{surface\text{-}DCM}$ R=0.541, p<0.001) (Fig. 4b). When analyzing samples depending on their location, we did not see a clear geographic pattern, as samples from different oceans appear intermixed, although we observed a tendency of Indian Ocean samples to cluster tighter than samples from other oceans. Furthermore, rDNA and rRNA samples formed different clusters, each one containing photic and aphotic samples, showing that both molecular surveys are providing different views of picoeukaryotic diversity (Fig. 4c).

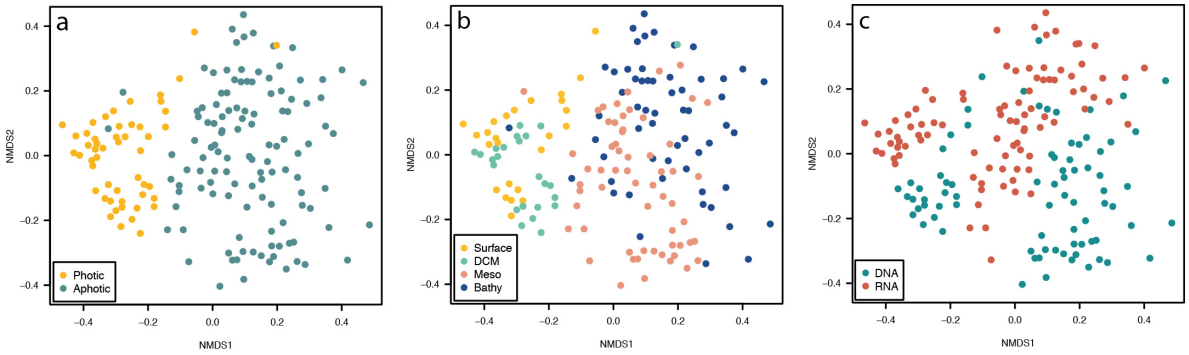


Fig. 4. Clustering of all picoeukaryotic samples on a Non-metric multidimensional analysis (NMDS) based on Bray-Curtis dissimilarities. Each sample is colored according the availability of light (a), the specific depth layer (b) and the use of rDNA or rRNA templates in the molecular survey (c).

Along the vertical profile, there was a variation in environmental conditions that could drive species sorting, such as a marked decrease in temperature and an increase of inorganic nutrients with depth (Fig. 5). Salinity slightly decreased with depth, while oxygen was minimal at the mesopelagic. This physico-chemical setting was affecting the abundance of several microbial components, such as picoeukaryotes, prokaryotes and viruses, which also decreased along the vertical profile (Fig. 5). We performed PERMANOVA analyses to determine what percentage of the variation in community composition could be explained by the measured environmental variables (Table S1). On a global scale, light and the ocean geography (atlantic, indic, pacific) explained together 29.4% of the variance (p<0.001), while 60% of the community

variance along the water column was not explained by the measured variables. As epipelagic and deep-ocean zones have dramatically different conditions, we performed individual PERMANOVA tests for each zone (Table S1). Within epipelagic zone, no variable explained more than 10% of the variability, whereas 32% of the variability in the deep-ocean was explained by the water mass.

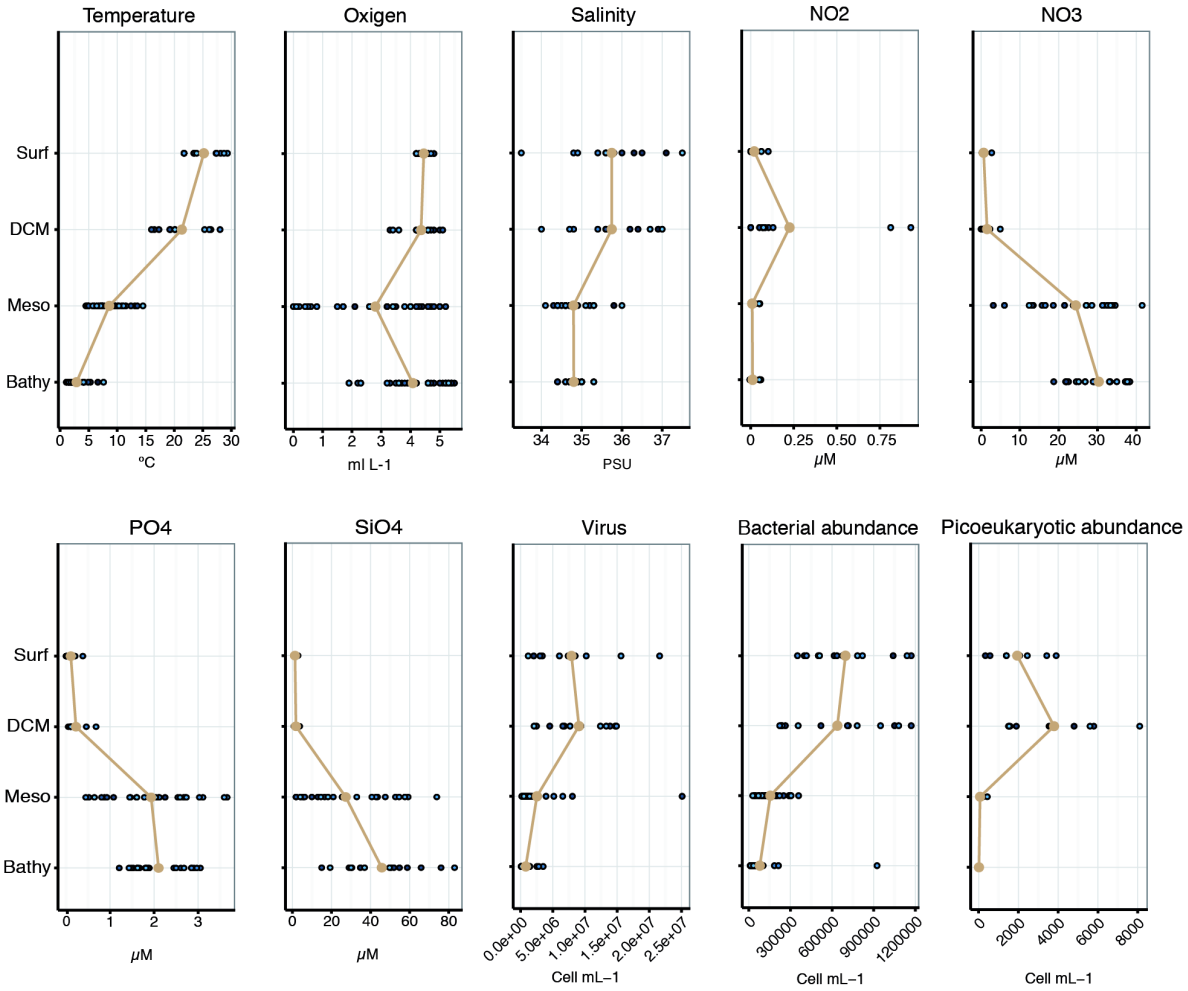


Fig. 5. Averaged values of environmental variables and microbial counts in the four layers of the water column (actual values as black dots and averaged values as brown dots).

Community similarity and OTUs dispersion in the different layers

To address the similarity in community composition along the water column, we calculated the Bray-Curtis dissimilarities among all samples from a given depth layer (Fig. 6a). The photic-zone, both surface and DCM, showed the highest similarity (median Bray-Curtis of 0.6-0.7),

while both aphotic layers were more dissimilar (Bray-Curtis of 0.8-0.9). Interestingly, the bathypelagic had the widest range of Bray-Curtis dissimilarities, indicating that community composition in these waters was highly heterogeneous, ranging from similar (Bray-Curtis close to 0.1) to very different communities (Bray-Curtis close to 1.0).

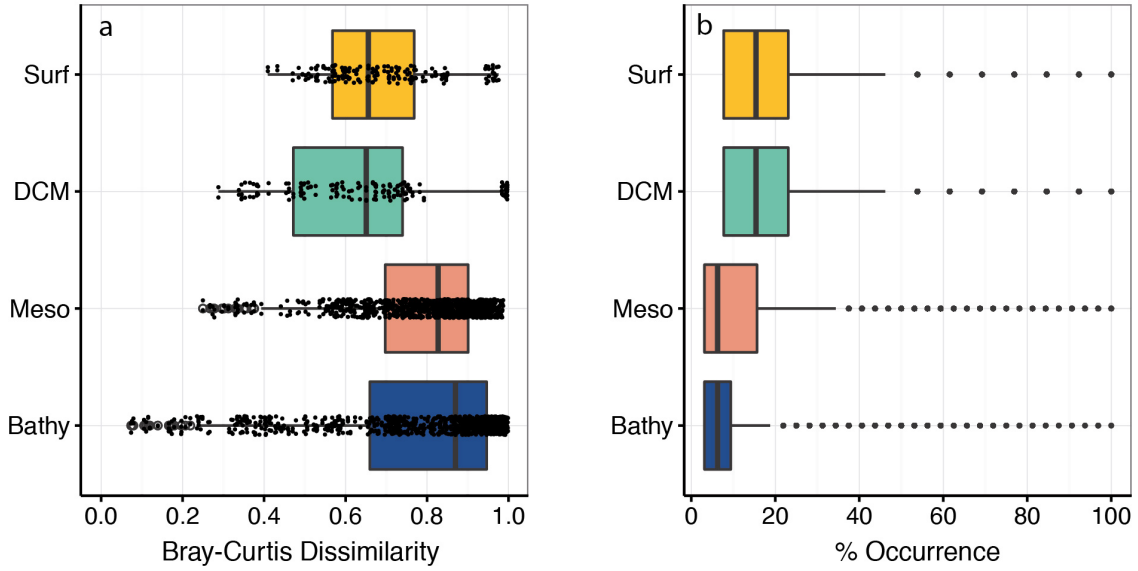


Fig. 6. Community similarity and OTU dispersal among all samples within of each water layer based on the rRNA assay. (a) Distribution of Bray-Curtis dissimilarities values among all samples from a given layer. (b) Distribution of the occurrence of OTUs (percentage of samples where they occur) within a given layer (b).

This variability in community composition along the vertical dimension could be related to a differential dispersal capability of the OTUs in the different depths. To assess this, we calculated the occurrence of all OTUs within each specific depth layer, as the percentage of samples where the OTU was detected. Surface OTUs were the most cosmopolitan, with OTUs being present in 17% of the samples (Fig. 6b) and 3.3% of OTUs in all samples (Table 1). At the other extreme, bathypelagic OTUs showed the lowest dispersion, with the OTUs found in only 7% of the samples (Fig. 6b) and only 0.3% of OTUs found in all samples (Table 1). This agrees with the previous data: layers with higher dispersion (i.e. surface) also display higher community similarity. Furthermore, as expected, OTUs displaying the highest occurrence were also the most abundant (data not shown).

	0-20%	20-40%	40-60%	60-80%	80-100%
Surface	65.4	20.0	5.8	5.4	3.3
DCM	66.3	20.6	5.7	5.0	2.4
Mesopelagic	82.6	10.7	4.5	1.4	0.8
Bathypelagic	89.4	7.4	2.3	0.6	0.3

Table 1. Distribution (in percentage) of all OTUs within a given water layer according to their occurrence in the samples from the layer (in percentage).

In order to evaluate the potential vertical dispersal of OTUs in the water column, we evaluated how many OTUs where unique within a given water layer (Table 2). Interestingly most of the OTUs were shared among different depths, specifically between the mesopelagic and bathypelagic, with 3,596 OTUs shared, and between surface and DCM with 1,794 OTUs shared. Nevertheless, unique OTUs averaged 47.5% of the total OTUs. Interestingly, the mesopelagic presented the highest percentage of unique OTUs (39.4%), which also accounted for the majority of reads, whereas the bathypelagic presented the lowest number of unique OTUs.

	% Unique-OTUs	% reads
Surface	24.52	17.40
DCM	20.14	12.36
Mesopelagic	39.44	57.12
Bathypelagic	15.90	13.12

Table 2. Percentage of OTUs unique in each water layer, together with the percentage of reads that they represent.

Relative abundance of taxonomic groups along the vertical profile

We aimed to identify the changes in the abundance of picoeukaryotes in the different layers of the water column at a global scale. For that, we had data from the two complementary rDNA and rRNA datasets. We first compared the abundance of each taxonomic group in these two datasets at a broad scale (in the 79 samples where both markers were available), in order to detect major differences (Fig. 7). A general view of plotting all groups together showed an acceptable correlation of rDNA and rRNA relative abundances. In a finer detail, some groups were clearly overrepresented in the rDNA dataset (i.e. number of rDNA reads much higher than rRNA reads), and these included three MALV lineages (-I, -II, and -V) and Polycystinea. Other groups, on the contrary, were overrepresented in the rRNA dataset, including some of the groups with largest relative abundance such as Chrysophyceae, Dinoflagellata, Ciliophora, Pelagophyceae and MALV-III. Overall, most groups were overrepresented in the rRNA dataset.

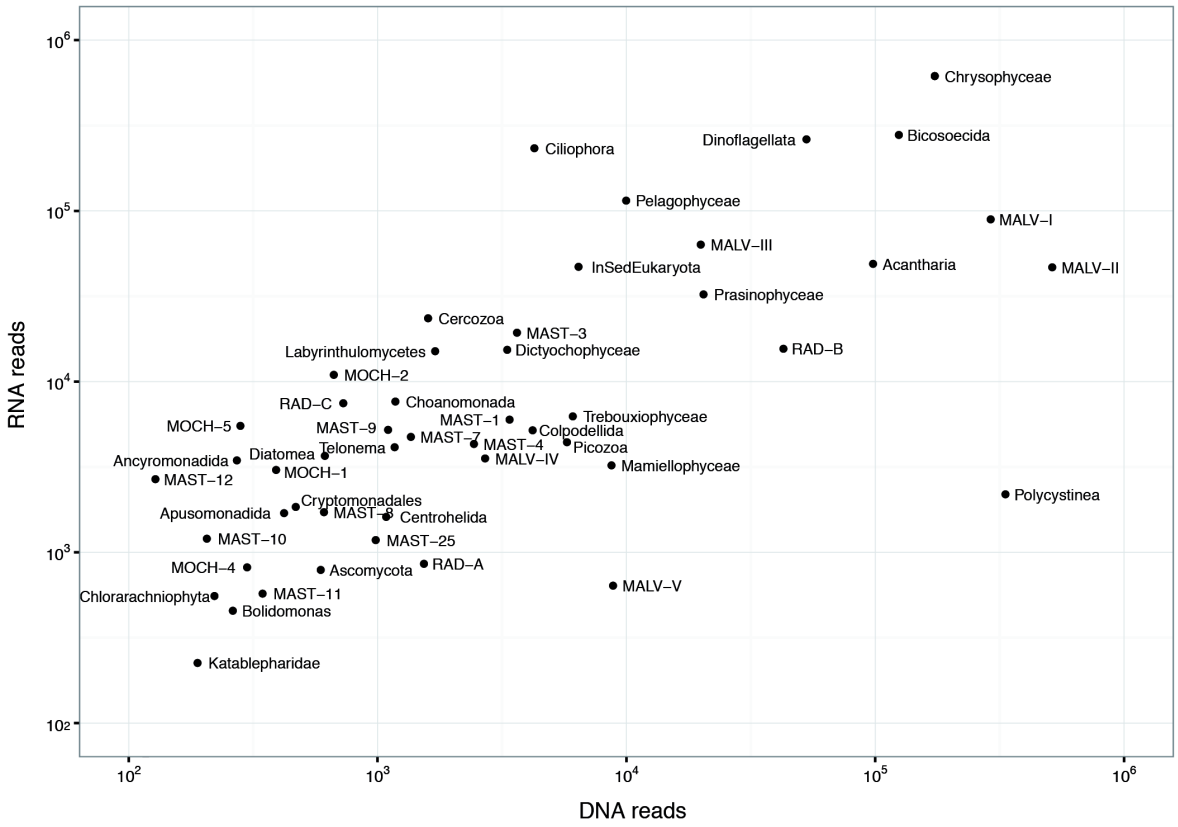


Fig. 7. Broad representation of all phylogenetic groups based in the total abundance in the rRNA and rDNA surveys (only groups with abundances >0.01% are shown).

To assess changes in the vertical profile, we investigated the relative abundance of the each taxonomic group in the four water layers as derived from both datasets (Fig. 8). Based on the values obtained from the rRNA dataset, we observed four main different behaviors: (i) groups that constantly increased their abundance with depth, such as Chrysophyceae, Bicosoecida and RAD-B, (ii) groups that decrease their abundance with depth, such as Dinoflagellata, Ciliophora, and all MAST (except MAST-9) and MOCH lineages, (iii) groups that peak in the mesopelagic (e.g. Cercozoa, Labyrinthulomycetes, RAD-C, MAST-9 and MALV-IV) and (iv) groups that peak at the DCM (e.g. Pelagophyceae, Telonema and green algae). As a consequence of these distributions, the dominant groups at the different depths of the water column were markedly different: Ciliophora and Dinoflagellata dominated in surface waters (42% of reads), Pelagophyceae and Dinoflagellata at the DCM (46% of reads) and Chrysophyceae and Bicosoecida in the dark ocean (40% of reads in mesopelagic and 73% of reads in bathypelagic).

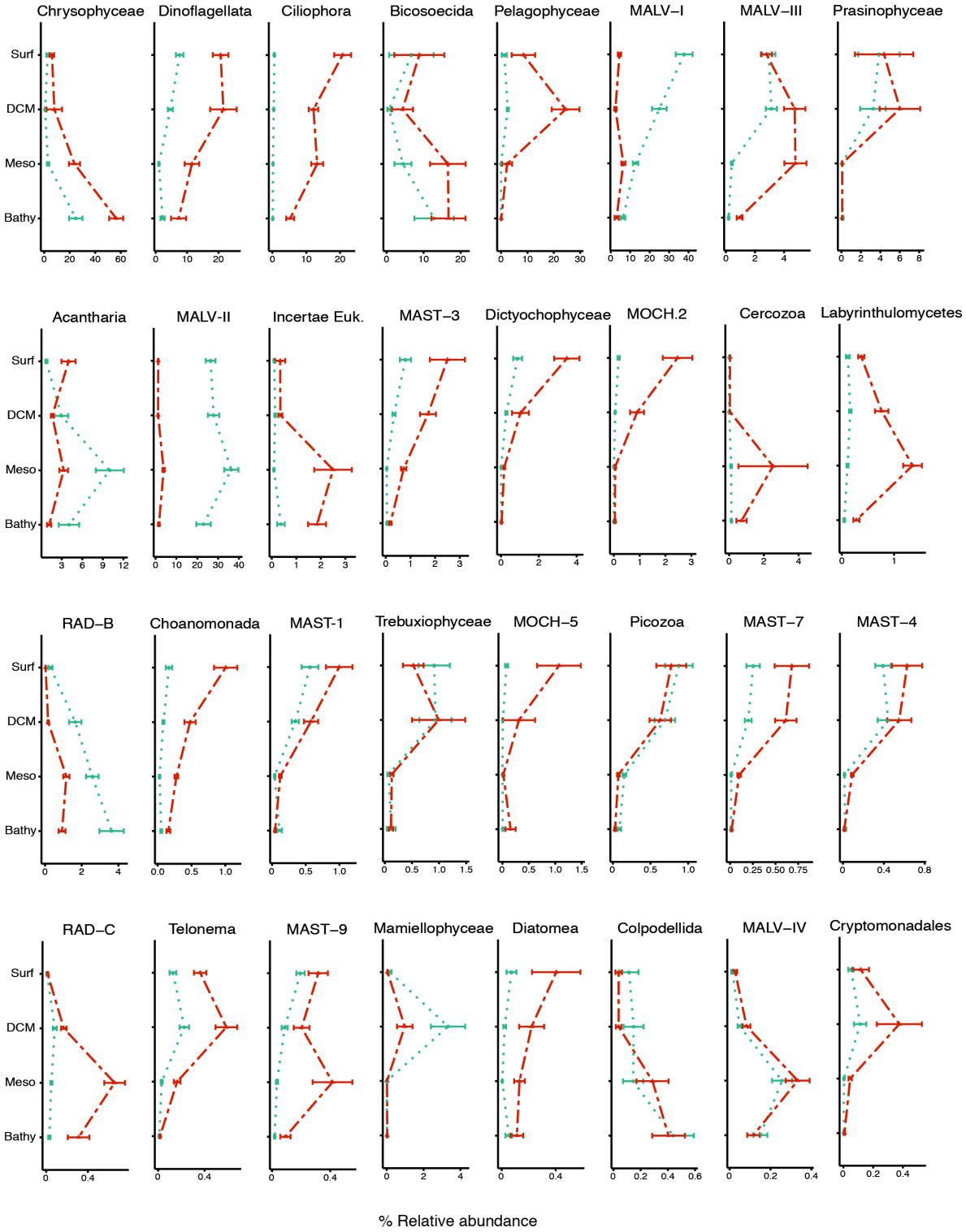


Fig. 8. Averaged relative abundance of the main taxonomic groups in the four water column layers derived from rRNA (red line) and rDNA (blue line) surveys. Groups are ordered based on their rRNA abundance.

miTags perspective of the taxonomic groups in the water column

We further explored the relative abundance of the taxonomic groups in four vertical profiles by the PCR-free metagenomic approach. This data was compared group by group with the rDNA dataset (Table S2). Most taxonomic groups showed a good correlation between the two approaches, and the ones with poor correlations where the least abundant (they accounted about 1-2% of the signal). Specifically, some of the most abundant groups in the metabarcoding like Polycystinea, MALV-II, Acantharia or Dinoflagellata were equally represented by miTags (those accounted for ~60% of the signal in both approaches). Other groups were overrepresented by metabarcoding (i.e. MALV-I, Bicosoecids and Prasinophyceae; together implying 22% of metaB and 12% of miTags) while others were underrepresented by metabarcoding (i.e., Chrysophyceae, RAD-B and Pelagophyceae; together implying 13% of metaB and 29% of miTags, see table S2).

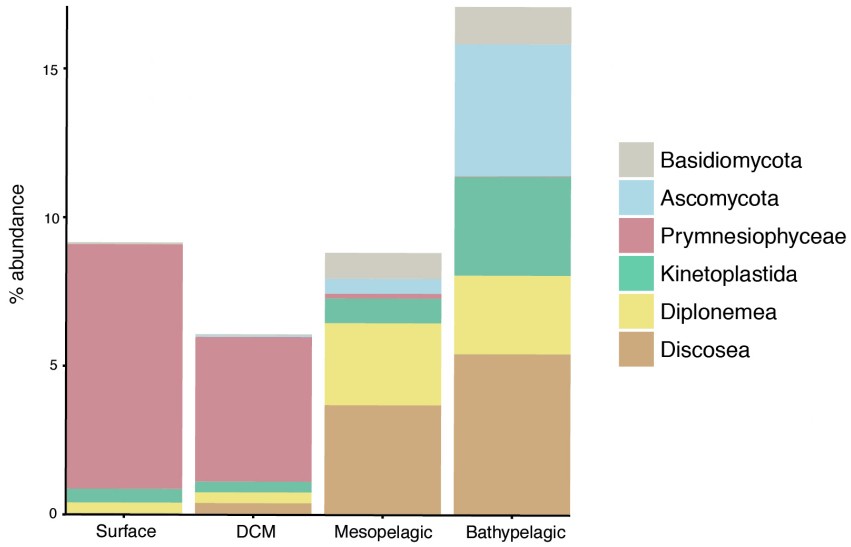


Fig. 9. Relative abundance of miTags in the four layers for taxonomic groups virtually absent in the metabarcoding dataset.

Nevertheless, the most striking output from the miTags was the detection of some groups, which were absent or very little represented in the PCR-based metabarcoding approach. These included the Prymnesiophyceae, the Excavata lineages Kinetoplastida and Diplonemea, the Amoebozoa lineage Discosea and the fungal groups Ascomycota and Basidiomycota. Altogether, these groups accounted for a substantial fraction of the metagenomic reads in all water column

layers. Thus, at surface and DCM the Prymnesiophyceae accounted for 5-8% of miTags, whereas in deeper layers the other five groups (Discosea, Diplonemea, Kinetoplastida and Ascomycota and Basidiomycota) represented 10% of the mesopelagic and 18% of the bathypelagic miTags (Fig. 9). As contrast, these six groups accounted only for 0.02% of metaB reads in the surface, 0.01% at the DCM, 0.04 at the mesopleagic and 0.12% at the bathypelagic.

Stratified activity across the water column

To determine changes in the metabolic activity of each OTU we calculated the ratio of rRNA vs. rDNA reads in all samples where both measures were available. So, each ratio obtained provided an indication of the metabolic activity of a given OTU in a given sample. Altogether we estimated 31,866 ratios. We first explored the global activity of each water layer by representing all ratios together (Fig. 10). The three upper layers had a ratio very close to 1, meaning no specific activity pattern in these, whereas the bathypelagic layer had lower values, indicating a general decrease of activity with depth.

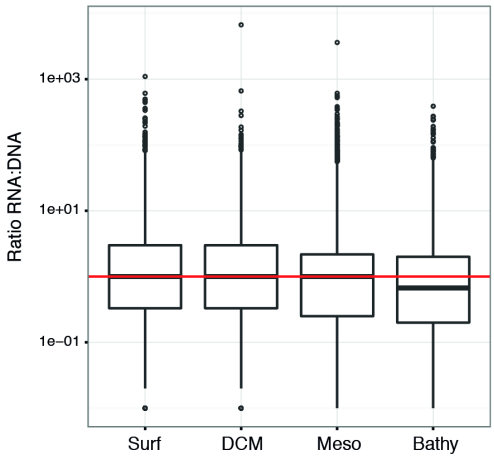


Fig. 10. Distribution of the activity ratios for all OTUs within a given depth layer. The red line indicates a ratio of 1.

We then determined the activity ratio of OTUs belonging to the main taxonomic groups along the vertical profile, to display differential activity of a given group related to depth (Fig. 11). In general, and consistent with the previous results, most taxonomic groups displayed the lowest activity ratios in the bathypelagic, although three groups, Dyctyochophyceae, Cercozoa and RAD-B, showed the lowest activity at the DCM. Regarding the highest activity, it was above the

1000m. Surprisingly, most of the identified groups have higher activity ratios in the mesopelagic (e.g. MALV-I, Dictyochophyceae, Cercozoa, Labyrinthulomycetes, among others, see Fig. 11)

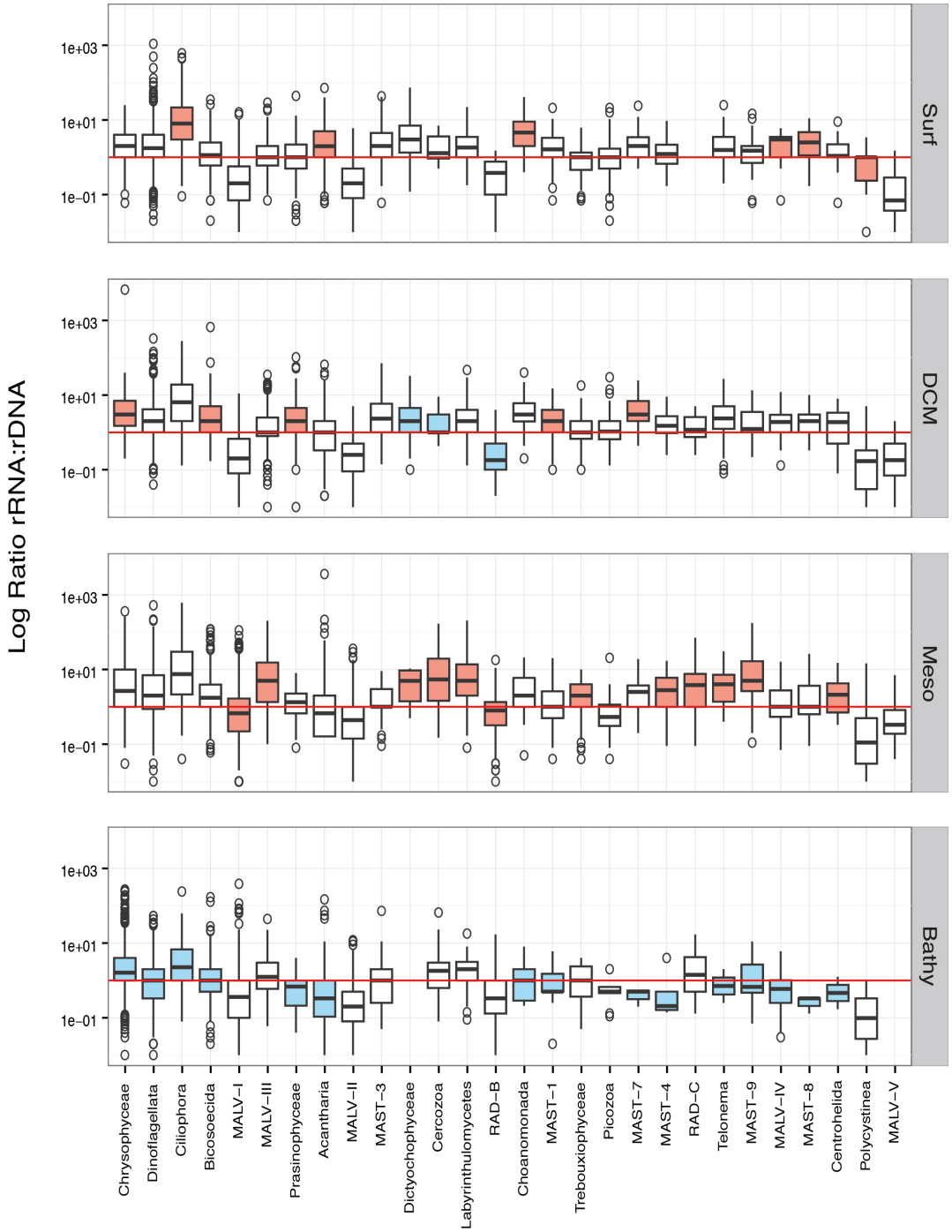


Fig. 11. Distribution of the activity ratios for all OTUs of each major taxonomic group within a given depth layer. The red line indicates a ratio of 1. For each taxonomic group, layers of high or low activity, when clearly contrasted, are marked in red and blue, respectively.

DISCUSSION

Most of the picoeukaryotic surveys done until today have been based in rDNA approaches, targeting rDNA genes in the nuclear genomes, while a few studies also included the rRNA approaches, targeting the rRNA at ribosomes (Not *et al.* 2009; Logares *et al.* 2014a; Massana *et al.* 2015; Hu *et al.* 2016). Both surveys provide complementary views on changes in the diversity of picoeukaryotic assemblages along different vertical and horizontal gradients. Moreover, the comparison among the two may also provide insights on specific metabolic activity, an essential parameter to understand their role in biogeochemical cycles. Indeed, there is still little knowledge about how active members of the picoeukaryotic community respond to different environmental conditions, being depth a critical factor. Our study provides an intensive assessment of the changes in the activity and distribution of picoeukaryotes in the whole water column, with a particular effort in the mesopelagic and bathypelagic depths of the world's main oceans.

Contrasting activity in the different depth layers

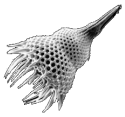
Environmental rDNA and rRNA approaches are giving different views of the taxonomic composition of the community (Stoeck *et al.* 2007; Not *et al.* 2009; Massana *et al.* 2015). This can be partly explained because some rDNA reads could derive from extracellular DNA from dead organisms. Furthermore, differences in the rDNA copy number among different taxa (Prokopowich *et al.* 2003), and differently sized cells (Zhu *et al.* 2005) will also generate differences in rRNA and rDNA views, particularly when considering the relative abundance of the taxa retrieved. In addition, and as a third explanation, the rRNA data may represent the metabolically active and alive taxa and will provide a better perspective of the active community, as rRNA is assumed to be absent from the extracellular pool and, within a single cell, the rRNA copy number will vary depending on its metabolic state (Blazewicz *et al.* 2013).

Our results indicated that broadly most taxonomic groups were overrepresented in the rRNA dataset, whereas only a few groups like MALV-I, MALV-II, Polycystinea, Acantharia were overrepresented in the rDNA. A general explanation for this pattern would be that these latter groups have a relatively higher rDNA copy number than the rest of the community. MALV-I and MALV-II are usually dominant groups in rDNA surveys (López-García *et al.* 2001; Bachy *et al.* 2011; de Vargas *et al.* 2015; Massana *et al.* 2015; Pernice *et al.* 2016) whereas they are ten times less abundant in rRNA surveys. It is well accepted that these likely have many rDNA operon

copies (Siano *et al.* 2010) indicating that, in most cases, we are overestimating their real abundance.

When analyzing the changes of taxonomic groups along the vertical profile, remarkably some groups like Chrysophyceae, Bicosoecida, RAD-B and Colpodellida increased their relative abundance with depth. Previous studies also found similar increases with depth for some of these groups (Countway *et al.* 2007; Brown *et al.* 2009). Interestingly the Radiolaria lineages also increased their abundance with depth, with RAD-C peaking in the mesopelagic, which is in agreement to what have been found by previous studies (Edgcomb *et al.* 2002; Not *et al.* 2007; Hu *et al.* 2016), and supports the idea that Radiolaria could be an important component of twilight and dark deep communities. On the other hand, other taxonomic groups followed the opposite trend, reducing their abundance with depth, and those include typically photosynthetic groups such as Pelagophyceae or the green algae and several heterotrophic lineages such as MAST clades or Picozoa. The occasional detection of phototrophic groups in the deep-ocean, sometimes at significative abundance, could be due to sinking particles (Agustí *et al.* 2015), although there is also the possibility of some of the detected taxa being mixotrophs.

Comparing rRNA:rDNA ratios among different taxa can be difficult to interpret due to the potentially large variation of rDNA copy number in different taxonomic groups, as explained earlier. However, it is feasible and much more interesting to compare rRNA:rDNA ratios within the same taxa, thus the rDNA copy number will be the same, in samples taken at different environmental conditions. So, the ratio could be used as a proxy of the activity of the particular taxonomic group under different conditions. Indeed, we found changes in the activity of the different taxonomic groups along the vertical gradient. Surprisingly, most picoeukaryotic groups showed the maximum of activity in the mesopelagic. A plausible explanation could be that the different biogeochemical features of the mesopelagic create 'hotspots' of microbial growth, usually along oxyclines (Edgcomb 2016), which may promote relatively high prey abundances that could be then grazed by active communities of Ciliates, Dinoflagellates and Cercozoans. Interestingly, in the mesopelagic there is the Deep Scatering Layer (DSL) that contains daily accumulation of fish biomass (Irigoien *et al.* 2014) and results in high prokaryotic biomass that again could attract microeukaryotic grazers. Furthermore, it has been observed that the clearance rates of heterotrophic nanoflagellates were higher in mesopelagic than in epipelagic samples (Cho *et al.* 2000). On the other hand, as expected, the majority of taxa were less active in the bathypelagic. In sum, our analysis shed new light on relatively unknown groups, such as the



MALV-III that is active at the mesopelagic, and overall highlight the potential important role of protists in the deep ocean.

Differentiated community assemblages in the water column

Our results indicate the existence of two differentiated assemblages in the water column corresponding to photic and aphotic communities. This strong differentiation had already been reported in previous vertical surveys (Countway *et al.* 2007; Brown *et al.* 2009) and could be partially explained by the dramatic changes in the presence of light. In addition, photic communities at surface and the DCM were overall more different than aphotic communities, which did not form clear clusters. This could be related with the fact that surface and DCM have stronger differences in their physico-chemical properties (light, temperature and inorganic nutrients) than mesopelagic and bathypelagic waters, indicating that when an organism is adapted to cope with the lack of light and high pressure it can be found in both deep layers. Our results support this hypothesis, as more OTUs where shared between the two deep ocean layers (3,596 OTUs) than between the two epipelagic layers (1,794 OTUs). Interestingly the mesopelagic was the layer that contained more exclusive OTUs, in agreement with previous observations (Brown *et al.* 2009), maybe due to their specific biogeochemical properties, such as the oxygen minimum zone and the deep scattering layer. However, despite the higher number of unique OTUs in the mesopelagic, this layer was not the most diverse, as richness is higher in surface and decreases with depth (Countway *et al.* 2007; Brown *et al.* 2009), which was not surprising as phototrophic organisms occupy the photic layers, and increase the diversity of these depths.

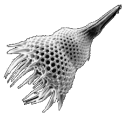
Noticeably, for the horizontal gradient our results showed that epipelagic communities were more similar than deep communities. This is related with the fact that surface currents more easily disperse surface OTUs, which makes the surface communities generally more similar. However, the bathypelagic realm contained both the most different and similar assemblages, and this could be potentially explained by the water-masses, as they explained 32% of the variability in the deep-ocean. It has been shown that different water-masses contain different assemblages, thus two distinct water-masses, even geographically close, will contain different communities, whereas the same water-mass will contain very similar communities (Pernice *et al.* 2016).

miTags vs metabarcoding approach

One of the main issues of molecular surveys is that different biases associated with the PCR step could affect and change the real picture of the diversity present in a sample. To verify that the global picture obtained here of the vertical picoeukaryotic diversity was reliable; we compared the relative abundance of taxonomic groups obtained with metabarcoding with an approach derived by metagenomics (miTags, (Logares *et al.* 2014b)). The miTags approach is PCR-independent, which means that is not subjected to the PCR bias when some phylotypes could be preferentially amplified whereas some could remain undetected due to primer mismatches (von Wintzingerode *et al.* 1997). Overall, most of the groups were well supported with both approaches. However, the important fact highlighted here was that some groups were totally absent (Prymnesiophyceae, Kinetoplastida, Diplonemea) or only slighlty represented (Discosea, Ascomycota and Basidiomycota) in the metabarcoding dataset. The lack of Excavata by metabarcoding was expected and already reported by Pernice *et al*. (2016), due to the long (>400bp) sequence for these groups at the V4 region, which were then not properly amplified during the PCR or did not pair after the pair-end Illumina sequencing. However, the lack of fungi in the metabarcoding data was unexpected, and could be explained due to the different protocols used here for DNA extraction, as fungal sequences have been retrieved from the deep-ocean in other studies that amplify the same 18S region (Pernice et al. 2016). Overall, the groups only found by mitags represent a small albeit substantial percentage in each of the depths, so the metabarcoding data provides a reliable image of the picoeukaryotic diversity in the water column.

CONCLUSIONS

This study provides the first insight of changes in diversity and metabolic activity of picoeukaryotes along the whole water column in the global ocean. In summary, we have shown that picoeukaryotic diversity has strong stratification along the water column, with two main communities corresponding to the epipelagic and the deep ocean. Additionally we have observed changes in the metabolic activity of the different taxonomic groups with depth, being the mesopelagic layer where most taxonomic groups had the highest metabolic activity, and also being the realm with more unique OTUs.



ACKNOWLEDGEMENTS

This research was funded by Spanish projects FLAME (CGL2010-16304, MICINN), ALLFLAGS (CTM2016-75083-R, MINECO) and INTERACTOMICS (CTM2015-69936-P, MINECO/FEDER, EU). CRG was supported by a Spanish FPI grant. We thank all the scientists that sampled for nucleic acid in the different legs of the cruise. We also thank the crew of the R/V BIO Hespérides.

SUPPLEMENTARY MATERIAL

Table S1. Results of the PERMANOVA test for all samples of the whole water column, or for separate sets of epipelagic and deep-ocean samples

Whole water column		
	R2	p-value
Light	0.1489	0.001
Temperature	0.0253	0.002
Salinity	0.0166	0.015
Ocean	0.1461	0.001
Depth	0.0600	0.001

	Epipelagic		Deep-ocean	
	R2	p-value	R2	p-value
Temperature	0.0672	0.026	0.0382	0.010
Conductivity	0.0685	0.024	0.0388	0.007
Salinity	0.0415	0.022	0.0241	0.068
Oxygen	0.0464	0.201	-	-
Bacterial abundance	0.0811	0.008	0.0149	0.450
Water mass	-	-	0.3200	0.001
NO_3	-	-	0.0127	0.538
PO_4	-	-	0.0321	0.029
SiO_4	-	-	0.0173	0.238

Table S2. Comparison of metabarcoding (metaB) and metagenomics (miTags) signal for each phylogenetic group, estimated by plotting its relative abundance in all individual samples in both surveys and calculating the slope and the R2 coefficient. Groups are classified according to their representation in both surveys and then groups without correlation.

Groups overrepresented by metaB (slope above 1.2)				
	% miTags	% metaB	R^2	slope 0
MALV-I	8.91	13.65	0.764	1.614
Bicosoecida	1.70	4.32	0.795	2.116
Prasinophyceae	0.47	2.44	0.951	5.374
Mamiellophyceae	0.62	0.75	0.906	1.200
Trebouxiophyceae	0.08	0.57	0.391	4.490
MALV-V	0.25	0.40	0.713	1.205

Groups equally represented (slope between 0.8 and 1.2)				
	% miTags	% metaB	R^2	slope 0
MALV-II	23.15	27.43	0.213	1.094
Polycystinea	24.34	25.87	0.902	1.012
Acantharia	5.75	6.10	0.579	0.976
Dinoflagellata	3.56	3.53	0.702	0.895
MALV-III	1.01	1.23	0.733	1.184

Groups underrepresented by metaB (slope below 0.8)				
	% miTags	% metaB	R^2	slope 0
Chrysophyceae	16.19	8.90	0.867	0.590
RAD-B	5.35	1.77	0.603	0.352
Pelagophyceae	1.66	0.55	0.816	0.328
Picozoa	0.35	0.31	0.400	0.775
MAST-1	0.54	0.19	0.783	0.307
MAST-3	0.65	0.19	0.615	0.351
Ciliophora	0.48	0.18	0.319	0.330
MALV-IV	0.30	0.14	0.208	0.337
Dictyochophyceae	0.34	0.14	0.561	0.406
MAST-4	0.72	0.13	0.743	0.158
RAD-A	0.37	0.09	0.906	0.238
MAST-7	0.20	0.08	0.625	0.262
Telonema	0.26	0.06	0.498	0.195
Choanomonada	0.18	0.05	0.755	0.251
RAD-C	0.12	0.05	0.288	0.331
MAST-25	0.18	0.04	0.364	0.116
MOCH-2	0.13	0.04	0.261	0.195
Cryptomonadales	0.09	0.04	0.537	0.360
Diatomea	0.12	0.03	0.894	0.235
MOCH-1	0.05	0.03	0.304	0.381
MOCH-5	0.03	0.03	0.582	0.388
MOCH-4	0.07	0.02	0.948	0.273
MAST-11	0.08	0.02	0.834	0.209
Katablepharidae	0.09	0.01	0.417	0.079
MAST-10	0.03	0.01	0.769	0.221
Ichthyosporea	0.01	0.00	0.959	0.062

Groups with a poor correlation (R^2 below 0.2)				
	% miTags	% metaB	R^2	slope 0
Colpodellida	0.16	0.20	-0.183	0.579
Labyrinthulomycetes	0.48	0.14	-0.425	0.187
Centrohelida	0.05	0.07	-0.027	0.456
Cercozoa	0.42	0.06	0.063	0.084
MAST-9	0.03	0.05	0.069	0.781
MAST-8	0.07	0.04	-0.182	0.294
Bolidomonas	0.02	0.01	-0.202	0.013
MAST-12	0.02	0.01	-0.195	0.158
Chlorarachniophyta	0.04	0.01	0.119	0.108
Ellobiopsidae	0.04	0.01	-0.101	0.081
Ancyromonadida	0.19	0.00	0.105	0.006
Apusomonadida	0.01	0.00	-0.132	0.000
MAST-2	0.00	0.00	-0.073	0.000
Marine Opisthokonts	0.02	0.00	-0.038	0.011

REFERENCES

Agogué, H., Lamy, D., Neal, P.R., Sogin, M.L. & Herndl, G.J. (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Molecular ecology*, 20, 258-274.

Agustí, S., Gonzalez-Gordillo, J.I., Vaque, D., Estrada, M., Cerezo, M.I., Salazar, G. *et al.* (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nature communications*, 6, 7608.

Arístegui, J., Gasol, J.M., Duarte, C.M. & Herndl, G.J. (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnology and Oceanography*, 54, 1501-1529.

Baas-Becking, L.G.M. (1934). *Geobiologie of Inleiding tot de Milieukund*. WP Van Stockum and Zoom (in Dutch): The Hague, The Netherlands.

Bachy, C., Lopez-Garcia, P., Vereshchaka, A. & Moreira, D. (2011). Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the north pole at the end of the polar night. *Frontiers in microbiology*, 2, 106.

Blazewicz, S.J., Barnard, R.L., Daly, R.A. & Firestone, M.K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME journal*, 7, 2061-2068.

Brown, M.V., Philip, G.K., Bunge, J.A., Smith, M.C., Bissett, A., Lauro, F.M. *et al.* (2009). Microbial community structure in the North Pacific ocean. *The ISME journal*, 3, 1374-1386.

Catalá, T.S., Reche, I., Fuentes-Lema, A., Romera-Castillo, C., Nieto-Cid, M., Ortega-Retuerta, E. *et al.* (2015). Turnover time of fluorescent dissolved organic matter in the dark global ocean. *Nature communications*, 6, 5986.

Cho, B.C., Na, S.C. & Chol, D.H. (2000). Active ingestion of fluorescently labeled bacteria by mesopelagic heterotrophic nanoflagellates in the East Sea, Korea. *Marine Ecology Progress Series*, 206.

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

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R. *et al.* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460-2461.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*, 10, 996-998.

Edgcomb, V.P. (2016). Marine protist associations and environmental impacts across trophic levels in the twilight zone and below. *Current opinion in microbiology*, 31, 169-175.

Edgcomb, V.P., Kysela, D.T., Teske, A., de Vera Gomez, A. & Sogin, M.L. (2002). Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 7658-7662.

Grasshoff, K., Ehrhardt, M. & Kremling, K. (1983). *Methods on seawater analysis*. 2n ed. Winheim: Verlag Chemie.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L. *et al.* (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic acids research*, 41, D597-604.

Hu, S.K., Campbell, V., Connell, P., Gellene, A.G., Liu, Z., Terrado, R. *et al.* (2016). Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS microbiology ecology*, 92, fiw050.

Irigoien, X., Klevjer, T.A., Rostad, A., Martinez, U., Boyra, G., Acuna, J.L. *et al.* (2014). Large mesopelagic fishes biomass and trophic efficiency in the open ocean. *Nature communications*, 5, 3271.

Logares, R. (2017). Workflow for Analysing MiSeq Amplicons based on Uparse v1.5. https://doi.org/10.5281/zenodo.259579

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R. *et al.* (2014a). Patterns of rare and abundant marine microbial eukaryotes. *Current biology : CB*, 24, 813-821.

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H. *et al.* (2014b). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental microbiology*, 16, 2659-2671.

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

Massana, R. (2011). Eukaryotic picoplankton in surface oceans. *Annual review of microbiology*, 65, 91-110.

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C. *et al.* (2015). Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental microbiology*, 17, 4035-4049.

Nagata, T., Tamburini, C., Arístegui, J., Baltar, F., Bochdansky, A.B., Fonda-Umani, S. *et al.* (2010). Emerging concepts on microbial processes in the bathypelagic ocean – ecology, biogeochemistry, and genomics. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 1519-1536.

Nikolenko, S.I., Korobeynikov, A.I. & Alekseyev, M.A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, 14 Suppl 1, S7.

Not, F., del Campo, J., Balague, V., de Vargas, C. & Massana, R. (2009). New insights into the diversity of marine picoeukaryotes. *PloS one*, 4, e7143.

Not, F., Gausling, R., Azam, F., Heidelberg, J.F. & Worden, A.Z. (2007). Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environmental microbiology*, 9, 1233-1252.

Oksanen, J., Blanchet, F.G., Kindt, R. & al., e. (2015). Vegan: Community Ecology Package.

Pernice, M.C., Forn, I., Gomes, A., Lara, E., Alonso-Saez, L., Arrieta, J.M. *et al.* (2015). Global abundance of planktonic heterotrophic protists in the deep ocean. *The ISME journal*, 9, 782-792.

Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M. *et al.* (2016). Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *The ISME journal*, 10, 945-958.

Pernice, M.C., Logares, R., Guillou, L. & Massana, R. (2013). General patterns of diversity in major marine microeukaryote lineages. *PloS one*, 8, e57170.

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

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41, D590-596.

Robinson, C., Steinberg, D.K., Anderson, T.R., Arístegui, J., Carlson, C.A., Frost, J.R. *et al.* (2010). Mesopelagic zone ecology and biogeochemistry – a synthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 1504-1518.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res*, 43, e37.

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

Siano, R., Alves-de-Souza, C., Foulon, E., Bendif, E.M., Simon, N., Guillou, L. *et al.* (2010). Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. *Biogeosciences Discussions*, 7, 7391-7419.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W. *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular ecology*, 19 Suppl 1, 21-31.

Stoeck, T., Taylor, G.T. & Epstein, S.S. (2003). Novel Eukaryotes from the Permanently Anoxic Cariaco Basin (Caribbean Sea). *Applied and Environmental Microbiology*, 69, 5656-5663.

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.

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

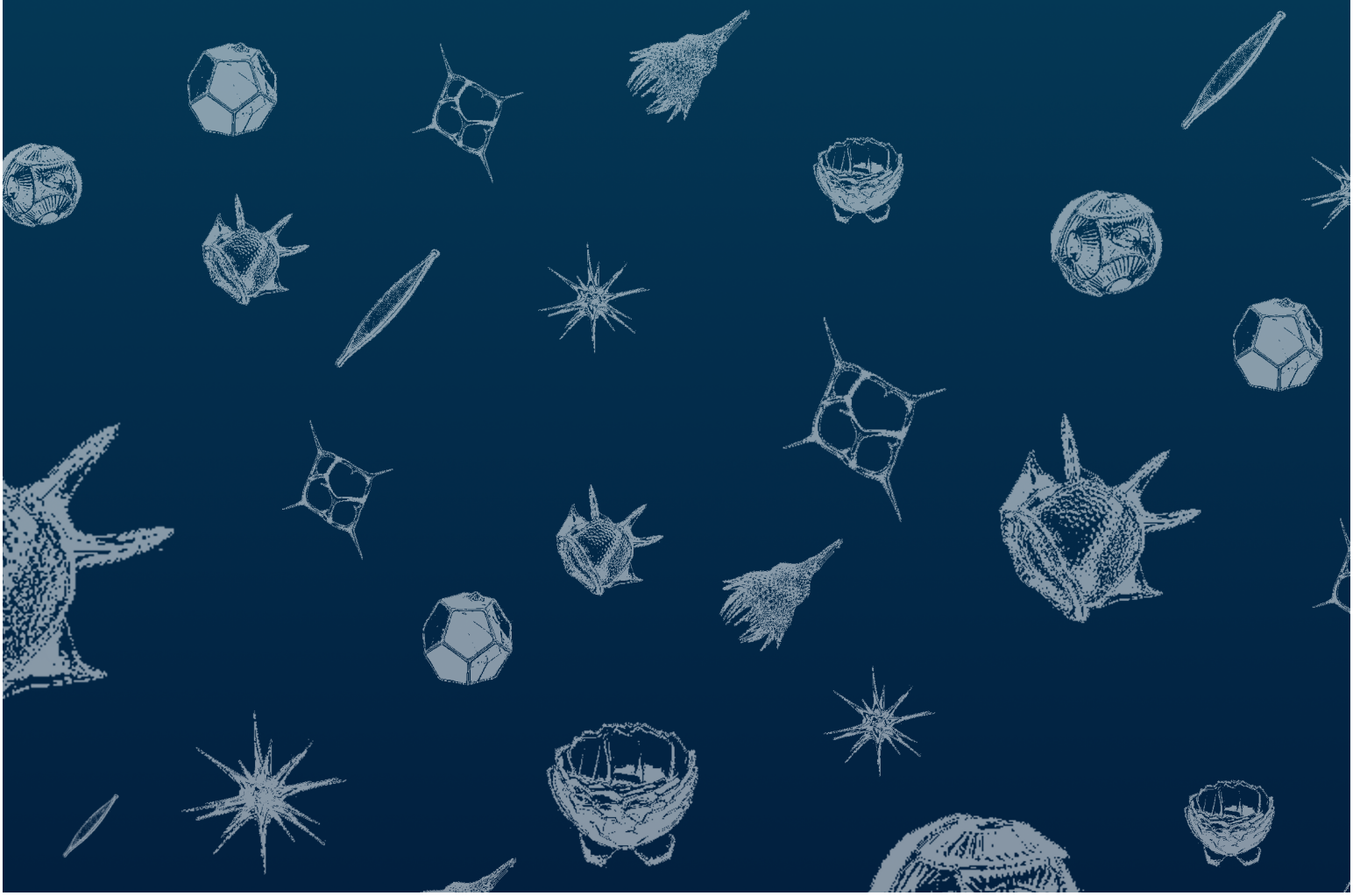
Weiner, A., Aurahs, R., Kurasawa, A., Kitazato, H. & Kucera, M. (2012). Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. *Molecular ecology*, 21, 4063-4073.

Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30, 614-620.

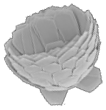
Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS microbiology ecology*, 52, 79-92.

Chapter 3

Chemotactic response of natural protists communities towards various stimuli

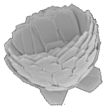


Caterina R. Giner, Irene Forn, Jean-Baptiste Raina, Roman Stocker, Justin R. Seymour, Ramiro Logares and Ramon Massana



ABSTRACT

Chemotaxis is the ability of an organism to respond to a chemical stimulus, but it is not clearly understood how marine protist communities respond to the different gradients of chemical substances. Furthermore, most of the experiments done until date are generally based on laboratory cultures, which could not represent what is occurring in natural assemblages. In this work we used an '*In Situ* Chemotactic Assay' (ISCA) to identify the global response of individual protist cells within a mixed community towards different attractants: ammonia, sugars (glucose, galactose and arabinose) and bacterial exudates from three different bacterial species (*Roseobacter*, *Alteromonas* and *Flavobacteria*), based on epifluorescence microscopy observations. Most of the tested compounds triggered a chemotactic response, which was strongest with the bacterial exudates of *Roseobacter* and *Flavobacteria*. Furthermore we also observed differences in the response between phototrophic and heterotrophic cells and regarding the protist cell size: higher chemotactic index was detected in phototrophic cells and cells smaller than 5 µm. Particular groups easily identified under the microscope, such as dinoflagellates, choanoflagellates and ciliates, also eludidated an specific response. Ciliates showed the strongest response towards *Roseobacter* and *Flavobateria* exudates. Overall our results suggest contrasted responses of the natural community to different stimuli, with a general preference towards bacterial exudates than other chemical cues.



INTRODUCTION

Most protists and bacteria individuals are motile and have the capacity to have a directional movement response in response to a simulus, a process generally known as taxis. This movement can be driven by very different stimulus e.g. by light (phototaxis), magnetic fields (magnetotaxis), pH (pH-taxis), oxygen (aerotaxis), or chemical cues (chemotaxis). Among the different taxis, the capacity to sense and respond according to chemical gradients (chemotaxis) has been the best studied (Stocker & Seymour 2012). In *sensu stricto*, the change in the movement and swimming speed in response to a chemical stimulus is 'chemokinesis' whereas 'chemotaxis' refers to the detection of the chemical gradient. However, in most studies both terms are mixed under the same label 'chemotaxis' (Fenchel & Blackburn 1999; Stocker & Seymour 2012), and here we will use this denotation.

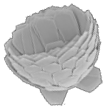
In the ocean, organic and inorganic nutrients are limited and heterogeneously distributed, with microscale hotspots found throughout the water column (Azam 1998). The microenvironment of a nonmotile bacterium, archaea or protist is defined by its cell size, whereas the microenvironment of a swimming cell is largely defined by its motility range (Stocker 2012). Motile populations are able to explore volumes of water, detecting different chemical gradients and moving directly towards detected hotspots. For instance, microscale patches of organic matter can originate by different processes, such as exudation of phytoplankton, excretion by zooplankton or cellular lysis (Mitchell *et al.* 1985; Blackburn *et al.* 1998). It is recognized that most motile bacteria are capable of detecting hotspots of organic matter through chemotaxis towards amino acids, sugars, sulfur compounds or inorganic nutrients (refs,Dennis), creating at the same time hotspots of bacterial activity. Phagotrophic protists, which include strictly heterotrophic and mixotrophic taxa, are significant grazers of bacteria and phytoplankton (Jurgens & Matz 2002). Therefore, these grazers may have detection systems to swim towards bacteria, as the efficient detection of the prey will improve the feeding potential.

The chemosensory and behavioral abilities of unicellular microbes have been recognized since the beginning of the last century (Jennings 1906). Bacterivorous protists are commonly observed to detect patches of high prey density and feed at hotspots of bacteria (Fenchel 1982; Mitchell *et al.* 1988; Blackburn & Fenchel 1999), and their selective feeding have been recognized as an important mechanism for the structuring of planktonic food web (Strom & Loukos 1998). However their mechanisms still remain poorly understood. Previous studies highlighted that heterotrophic protists have the capacity to respond to surrounding stimuli, and cells can regulate their position to find optimal conditions (Fenchel & Blackburn 1999; Wolfe

2000). They can utilize chemosensory mechanisms to move towards an attractant that might act as a proxy for prey cells (Snyder 1991; Fenchel & Blackburn 1999). For instance, in aquatic environments they respond to sources of dissolved organic matter and are capable of congregating at them within few minutes from distances of up to several centimeters. Various chemicals have been shown to affect the movement and, in some cases, the grazing rate of phagotrophic protist, both positively and negatively (Verity 1991; Wolfe 2000). Some positive chemoattractants for protists include ammonia, amino acids, fatty acids and bacterial and phytoplankton preys (Sibbald *et al.* 1987; Bennett *et al.* 1988; Fenchel & Blackburn 1999; Martel 2006), but is not clear which molecular mechanisms are responsible of these motile responses by protists predators. In addition, protistan response to a stimulus can be complex and affected by the physiological state of the cell (Fenchel 2002). Interestingly, some evidences exist suggesting that signaling pathways associated with chemotaxis appear to be conserved throughout eukaryotes (Roberts *et al.* 2011). Moreover, this is surely taxa-specific, as different taxa will have different responses.

In the literature, most experiments studying chemotaxis are 'grazing experiments' or 'predator-prey response experiments', and these had increased our knowledge about the preferences of different bacterial prey (Fenchel 1990; González *et al.* 1993; Pfandl *et al.* 2004; Ayo *et al.* 2009). Nowadays it is known that planktonic protists can control or influence different attributes of prey populations, thus they can discriminate between similar-sized preys based on differences in prey cell-surface composition. The different characteristics of the prey cells e.g. the release of chemical cues, prey motility, prey biogeochemical composition, prey cell surface characteristics and their cell-size influence the selective feeding by protists (Montagnes *et al.* 2008). Interestingly, prey size is the most influential factor. However, most grazing experiments have been done with monocultures, and could not be a good extrapolation for prey selectivity and preferences among different preys.

Fewer studies have been done to investigate direct response to specific chemical cues, i.e. 'pure' chemotactic experiments (Sibbald *et al.* 1987; Lee *et al.* 1999; Fan *et al.* 2002). For instance, Strom and collaborators (Strom *et al.* 2007) studied the response of a tintinnid ciliate to different amino acids and showed that its feeding responses was inhibited in the presence of the amino acids. On the contrary, Ayo and collaborators (Ayo *et al.* 2010) also analyzed the response towards several amino acids and showed that young protists were attracted to them. Understanding the role of chemical mediated prey location in prey selection, which involves response and attraction to chemical cues, remains still in its infancy. It is important to remark



that all experiments performed until now have been done using laboratory cultures, which could react different from natural environmental conditions, because laboratory cultures are acclimated to saturating levels of optimal prey, and this could not reflect real *in situ* conditions and reactions (Montagnes *et al.* 2008).

Our aim was to test the chemotactic response of the individual cells within a natural protists community to different types of stimulus. We used the ISCA (*In Situ* Chemotactic Assay) to identify the response of protists to ammonia, different types of sugars and three bacterial exudates. To our knowledge this is the first chemotactic experiment using protist natural assemblages. We want to assess if there is a significant chemotactic response of the protist cells of the community towards the different attractants tested. In particular, we wanted to address the following questions: do bacterial exudates exhibit a stronger chemotactic response than other attractants? Are larger protists more responsive than smaller ones? Do specific cells, such as ciliates and dinoflagellates, have a chemotactic preference?

MATERIALS AND METHODS

Incubation and microbial cell counts

The study was conducted using surface water of Glebe Bay (Sydney, Australia) in September 2015. Prior to the experiment, ~3 L of seawater were pre-filtered through 200 µm mesh and incubated in the dark during 3 days in order to increase the abundance of uncultured heterotrophic flagellates in an unamended incubation (Massana *et al.* 2006). Every day, samples for microscopic counts were taken in order to follow the dynamics of heterotrophic and phototrophic flagellates. For this, 25 mL of the seawater were fixed with glutaraldehyde (1% final concentration) and left for 3-4 hours at 4ºC. Then 20 mL were filtered through 0.6 µm pore-size polycarbonate black filters and stained with DAPI (4'-6-diamidino-2-phenylindole) at 5 $\mu g \cdot mL^{-1}$. Filters were mounted on a slide and stored at -20ºC until processed.

The total cell abundance of microbial eukaryotes in each sample was estimated by inspecting DAPI-stained filters. Cells were counted with an epifluorescence microscope (Zeiss AxioImager Z2) at 1,000X under UV excitation, changing to blue light excitation to verify the presence or absence of chlorophyll autofluorescence (to identify phototrophic and heterotrophic cells). Cells were classified in different size classes: 1-3 µm, 3-5 µm, 5-10 µm, >10 µm. Furthermore, cells

within recognizable taxa, such as ciliates, dinoflagellates and choanoflagellates, were also identified and counted.

In situ Chemotaxis Assay (ISCAs)

We performed the chemotaxis experiment using the ISCA (*In Situ* Chemotaxis Assay), a newly developed microfluidic-based platform (Seymour *et al.* 2008). The ISCA was designed to create a high-throughput method for developing chemotaxis quantification *in situ*, allowing to test multiple chemoattractants at the same time under identical conditions. It is made out of two layers of acrylic and a gasket layer made out of rubber that binds them. Each ISCA consists in a matrix of 25 cylindrical wells (5 wells per row with 5 rows), and each well have two holes of 0.8 mm and 0.6 mm diameter, which connect the interior of the chamber with the exterior. The internal volume of the well is ~110 µl. Each of the individual wells was filled with the tested chemoattractant using a syringe.

The chemoattractants tested were Ammonia (1 mM), sugars (glucose, galactose and arabinose at 1 mM each), and three bacterial exudates (*Roseobacter*, *Alteromonas* and *Flavobacteria*). These cultured bacterial strains were grown overnight in 5 mL of marine broth 1%. Exudates used for the experiment were obtained by filtering the bacterial culture twice through a 0.22 µm filter to remove any bacterial cells and keep only the exudates. Chemical chemoattractants were prepared in 0.2 µm filtered seawater to ensure that the chemical characteristics of the background solution in the ISCA were similar to the seawater. The control condition contained 0.2 µm filtered seawater, and allowed to measure the number of cells that randomly swim into the ISCA wells.

For the experiment, the ISCAs were filled with the different attractants and submerged within a tray containing 1.6 L of the unamended incubation for 1h, in order to avoid the replication of the protists inside the wells (Fig. 1). During this time, the chemoattractant was gradually leaking outside the well into the external seawater through the hole via molecular diffusion, creating a gradient in the surrounding seawater that derived the chemotactic response of the protists into the wells. After the incubation, the volume of chemoattractant or filtered seawater for the control condition contained in the wells was recovered using a pipette (~100 µl per well were recovered), fixed with glutaraldehyde (1% final concentration) and left for 24 hours at 4ºC. Lately, as the volume to filter was small (~100 µl), it was stained with DAPI and two 100 µl drops (corresponding to two ISCA wells) were filtered in the same 0.6 µm pore-size polycarbonate black filters without touching each other. Each chemoattractant condition was

replicated three times providing three samples for cell counting per condition. The total number of cells in each of the ISCA wells was counted with an epifluorescence microscope (Zeiss AxioImager Z2) at 1,000X under UV excitation. In order to make sure that all the cells were counted, all the diameter of the drop in the 0.6 µm filter was counted. The size of each observed cell was measured and the presence or absence of chlorophyll was also annotated.

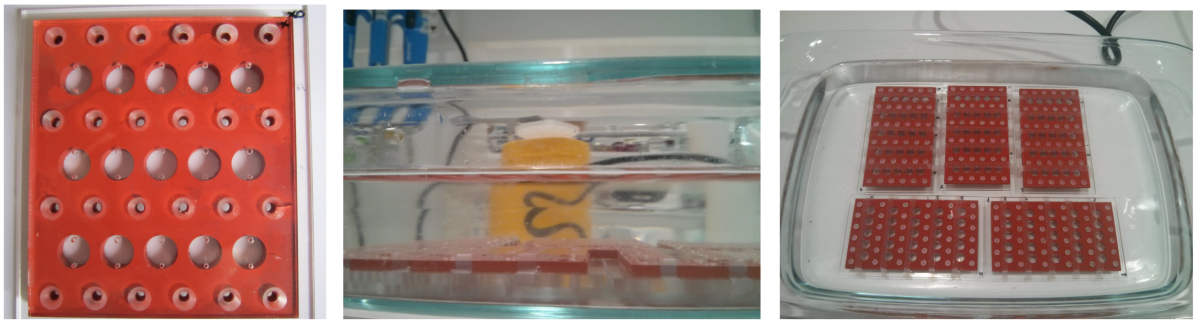


Fig. 1. Photography of the ISCA and its disposition during the experiments. It was placed on a tray and each well was filled with a different chemoattractant.

Analysis of the data and chemotactic index

The accumulation of protists in response to the chemoattractants in each condition was expressed in terms of total number of cells and in terms of a chemotactic index, Ic (based on Tout et al. (Tout *et al.* 2015)). This Ic was calculated by normalizing the number of cells responding to the specific chemoattractant to the number of cells responding to the filtered seawater control. Ic=1 indicates no response, thus the attractant had the same number of cells than the control. Chemotactic responses were compared using a T-test.

RESULTS

The experiment performed with the ISCA allowed to analyze the chemotactic behavior of individual cells within natural protists assemblages, and gave the possibility of testing different attractants at the same time under the same experimental conditions (Fig. 1). Most of the chemoattractants tested elicited a strong chemotactic response of natural protists, evidenced by the higher number of cells within the wells of the ISCA containing the different attractants (Fig. 2). Cell counts in the control-wells were low (~60 cells), indicating that only a small amount of

cells fall into the ISCA by randomly swimming. Overall, all the bacterial exudates exhibited significantly higher chemotaxis (Ic from 4.0 to 8.6; $p<0.05$), being *Roseobacter* exudates the ones with the strongest response, reaching cell concentration 8 times higher than the control (Ic= 8.6). Among the sugars, galactose exhibited a significant chemotactic response ($p<0.05$), whereas glucose and arabinose did not show a significant chemotaxis ($p>0.05$), mainly due to the high inter-replicate variability, since the number of cells in these conditions was higher than in the control, which indicated a positive response to the chemical signal. Interestingly ammonia did not display significant response (Ic=1.81; $p>0.05$) and the number of cells was not significantly different to those in the control with filtered seawater.

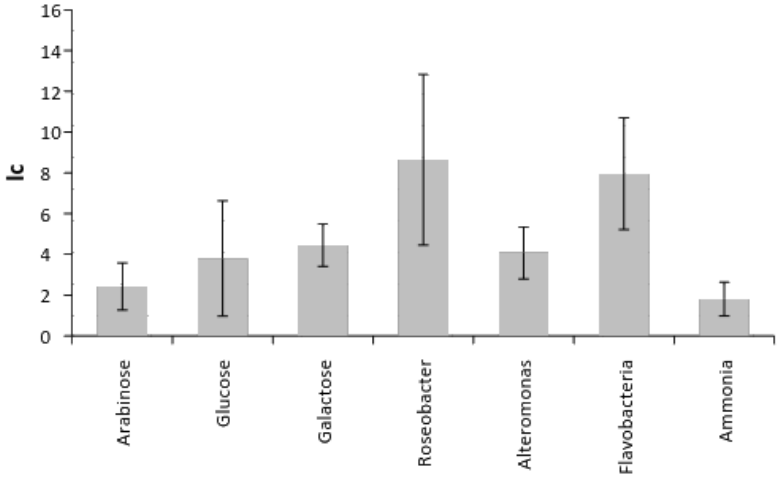


Fig. 2. Chemotactic index (Ic), of protist community responding to the different attractants tested.

We wanted to assess if phototrophic and heterotrophic cells exhibited a different response towards the attractants. In the unamended community used for the experiment, ~60% of the cells were heterotrophic and ~40% were phototrophic, while in all experimental conditions the proportion of heterotrophs was higher than in the unamended. That was more marked in the conditions with galactose and with exudates of *Roseobacter* and *Flavobacteria*, where ~80% of the attracted cells were heterotrophic (Fig. 3a). On the other hand, when analyzing the chemotactic index (Ic), phototropic cells always showed a significantly higher Ic than heterotrophic cells for all conditions (Fig. 3b). This apparent contradiction is mainly due the low number of phototrophic cells found in the control. The significant chemotactic response of phototrophic cells could indicate that they may have the capacity of being mixotrophic and

reacting to bacterial exudates and the other chemical cues. Regarding the heterotrophic subcommunity, all conditions generally triggered a chemotactic response. Noticeable, conditions with exudates of *Roseobacter* and *Flavobacteria* showed the highest Ic for both phototrophic and heterotrophic cells, being between 8-14 times higher than the control ($p>0.05$), indicating a strongest reaction of protists towards bacterial complex exudates instead of single sugars. Both phototrophs and heterotrophs exhibited a small but sigificative response with *Alteromonas* exudates and galactose (Ic from 3.9 to 7), whereas with arabinose and glucose there was a positive trend of phototrophic cells (Ic= 4.1 and Ic= 6.5) but not significant.

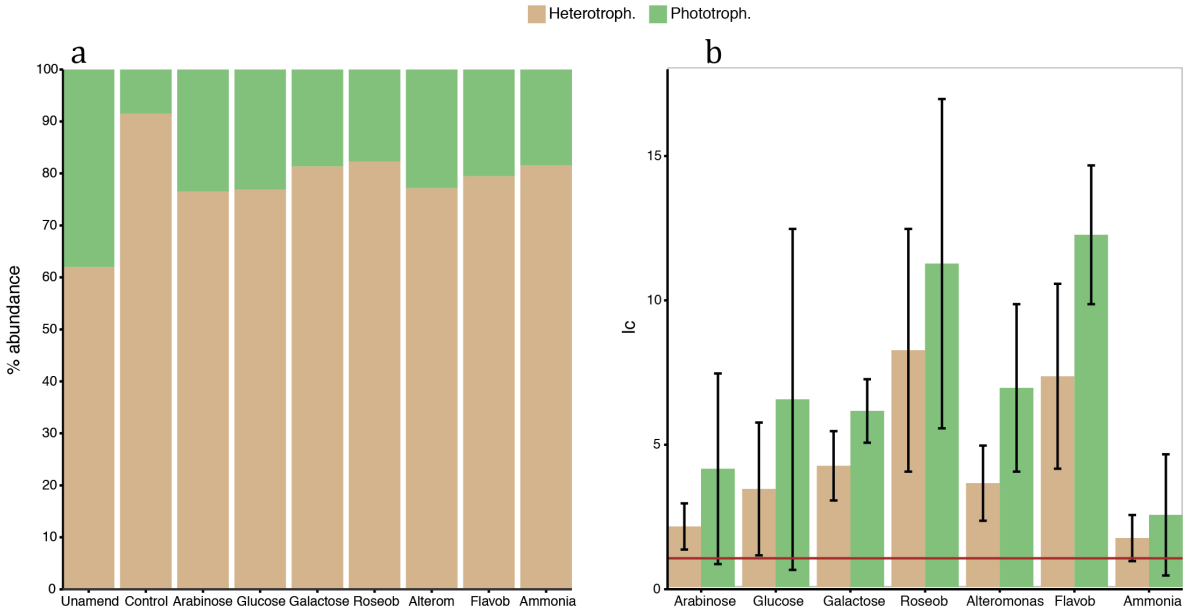


Fig. 3. Heterotrophic and phototrophic cells responding to the different attractants. (a) Percentage of the number of cells in each of the conditions including the unamended incubation. (b) Chemotactic index (Ic). Protists concentrations of the attractants have been normalized to concentrations in the filtered seawater control.

The protist community has cells from a variety of sizes and our results indicated that protists from these size classes reacted differently (Fig. 6). In the unamended community most of the cells (70%) belonged to the 1-3 µm size class, 24% of the cells to 3-5 µm, 5% to 5-10 µm and <1% to cells to >10 µm size class (Fig 4a). Interestingly, the size distribution of the cells reacting to the tested conditions was very different among them. In many conditions cells larger than 3 µm were the most reactive (Fig. 4a) specially the 3-5 µm class in arabinose and glucose, or cells even larger than 10 µm in *Roseobacter* exudates and glucose (about 6% of cells >10 µm). As

contrast, *Flavobacteria* exudates resulted in a size class distribution similar to the unamended incubation whereas the ammonia treatment triggered the response of cells between 1-3 µm, which ended being about ~55% of the cells attracted. When looking at the chemotactic index (Fig. 4b), smaller cells (1-3 µm) from *Roseobacter* and *Flavobacteria* exudates exhibited the highest positive response (Ic=10-13). In galactose and *Alteromonas* exudates, cells from 1-3 µm also displayed the highest Ic. Interestingly in *Roseobacter* exudates and galactose, all the different cell sizes showed a significant response ($p<0.05$). Only significant differences between size classes were found in *Flavobacteria* and *Alteromonas* exudates, where the response of 1-3 µm cells was significantly higher than the rest of the size classes. It is important to remark that overall, in all the wells of the ISCA filled with attractants, there were more cells larger than 10 µm as compared with the unamended incubation and in the filtered control, indicating a response of bigger cells towards the different chemical cues.

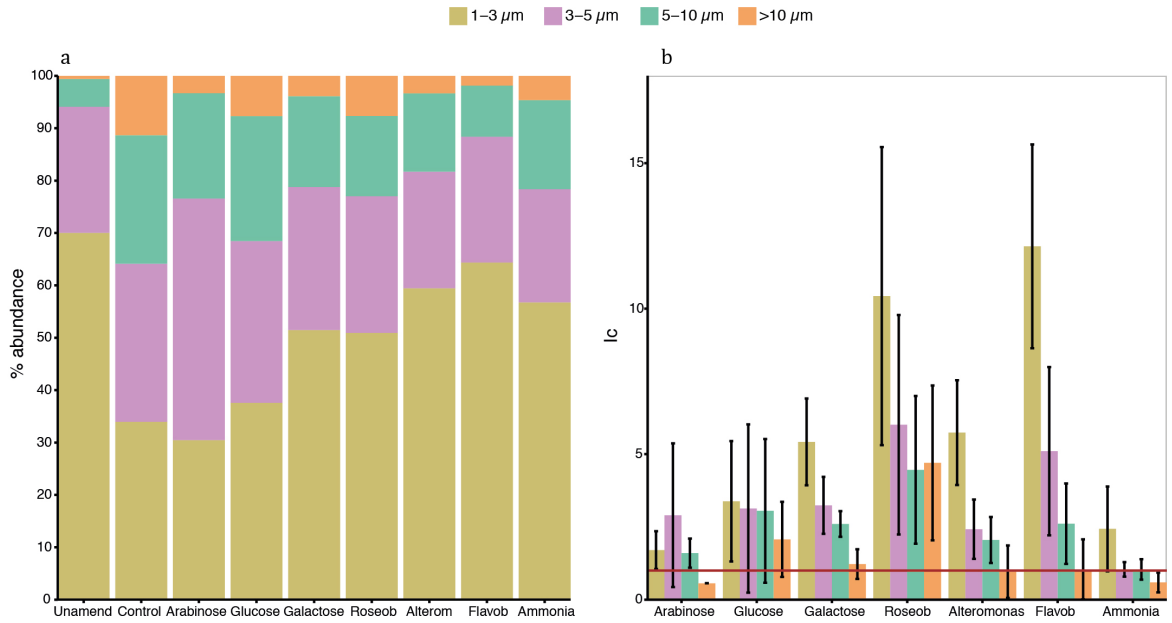


Fig. 4. Different cell sizes of the cells responding to the different attractants. (a) Percentage of the number of cells in each of the conditions including the unamended incubation. (b) Chemotactic index (Ic). Protists concentrations of the attractants have been normalized to concentrations in the filtered seawater control.

Under the microscope we were able to clearly identify cells within three taxonomic classes, i.e. ciliates, dinoflagellates and choanoflagellates, so we determined if they had preference for any of the attractants tested. Dinoflagellates and ciliates showed the highest chemotactic response to

Roseobacter exudates, with Ic >12 (Fig. 5). Ciliates also displayed significant chemotaxis towards *Flavobacteria* exudates (Ic=11) and glucose (Ic=6). Overall, dinoflagellates were the ones that showed most chemotaxis to most of the attractants, having a significant response to glucose (Ic=8), galactose (Ic=5.7) and *Flavobacteria* exudates (Ic=4.7). On the other hand, choanoflagellates were generally the least responsive of the three groups and had a weak chemotactic response only towards *Roseobacter* and *Flavobacteria* exudates (Ic~3.5).

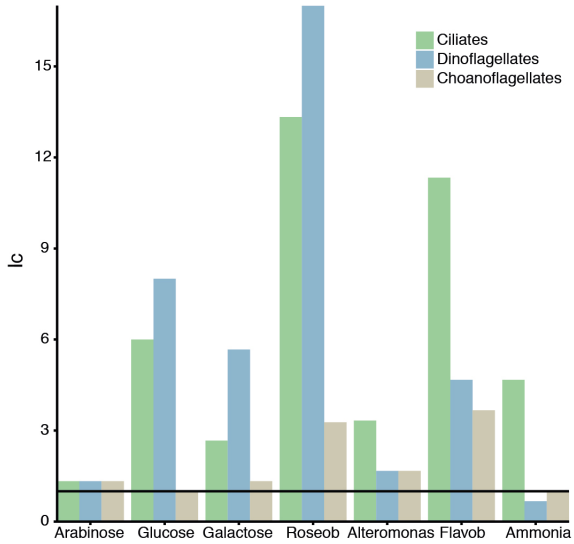


Fig. 5. Chemotactic Index (Ic) of ciliates, dinoflagellates and choanoflagellates towards the different attractants. Their concentrations were normalized to their specific concentrations in the filtered seawater control.

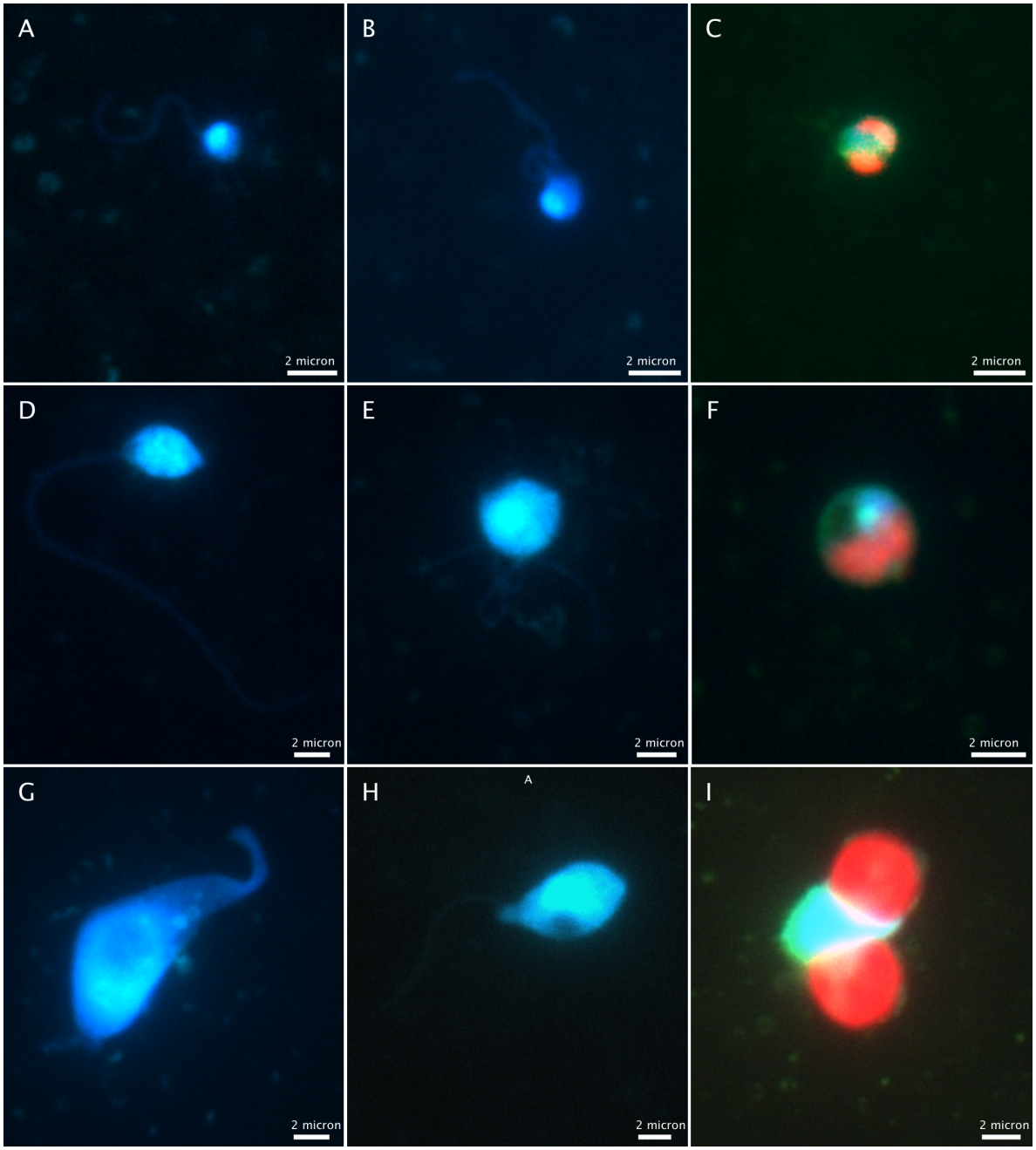


Fig. 6. Epifluorescence images showing different cell shapes and sizes, the presence of chlorophyll and flagella. The blue signal corresponds to the DAPI-stained nucleus, and the red signal to the chlorophyll. Scale bar represents 2 μm.

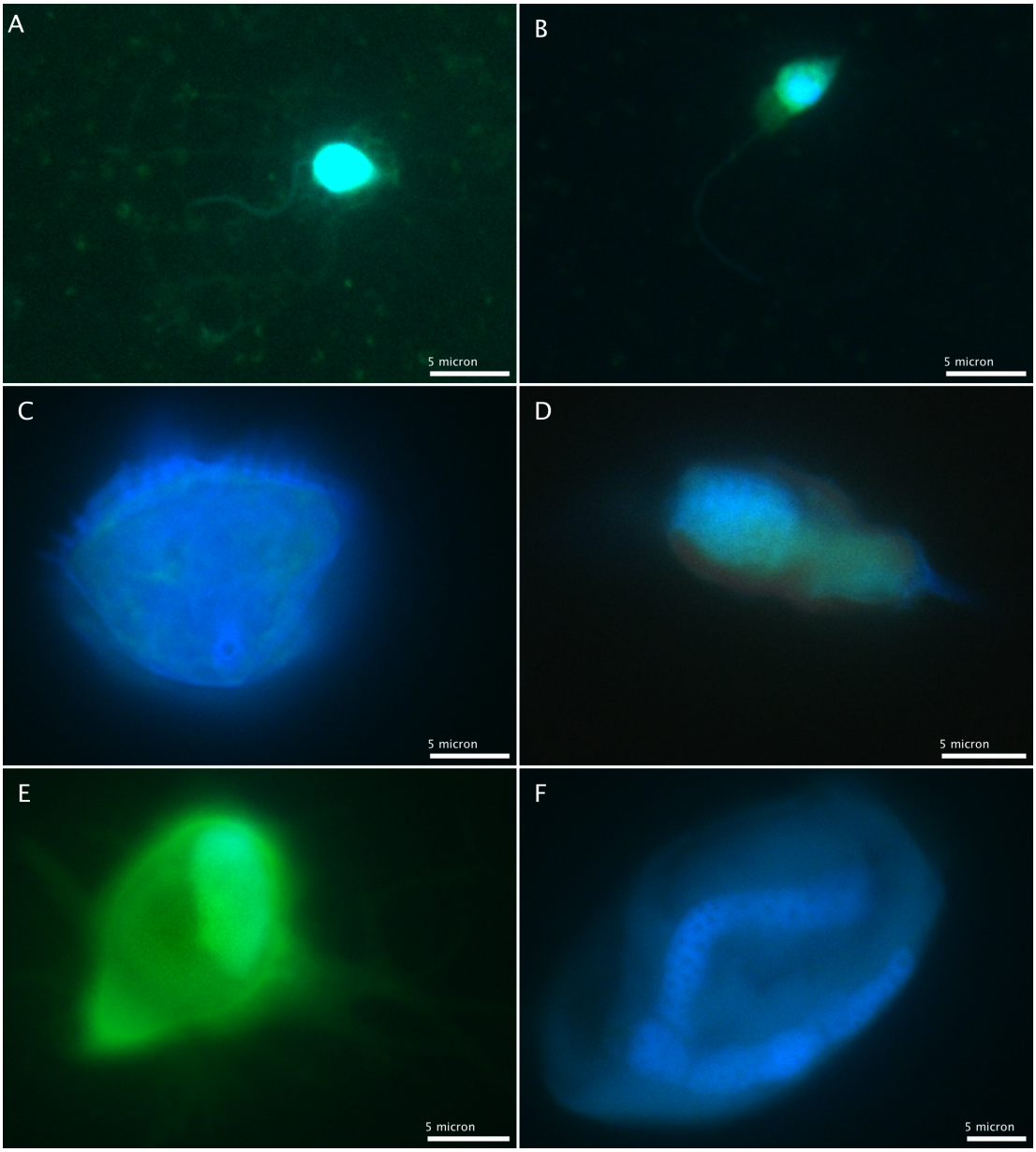


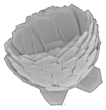
Fig. 7. Epifluorescence images of choanoflagellates, ciliates and dinoflagellates observed in the ISCA wells. Scale bar represents 5 μm.

DISCUSSION

To our knowledge, our data provided the first chemotactic data for individual protist cells within natural communities, being thus the first experimental report with natural mixed asssemblages. Overall we demonstrated high levels of chemotactic response of the protist assemblages towards different attractants. The ISCA device gives us the possibility of doing *in situ* experiments using different concentrations of different chemoattractants at the same time.

Furthermore, the ISCA also allows the possibility of doing the experiment in the laboratory with more controlled conditions, and in our case it gives the possibility to manipulate the community prior to the experiment. In this case we have performed an unamended incubation in the dark, to increase the abundance of heterotrophic flagellates (Massana *et al.* 2006) in order to have higher number of cells and get a stronger reaction. These *in situ* experiments add an advance in the understanding of the behavior of protist communities, since our knowledge obtained in previous chemotactic experiments derived from cultures. Culture experiments have several limitations in this experimental context. First, it is well known that only a small number of species have the potential of being cultured, and usually some of them do not represent the dominant species present in the community (del Campo *et al.* 2013). Second, cultured species may have adapted to a confortable condition, usually living without limited nutrients, and their response to chemical cues could not be the same that they would perform in natural communities. In fact, cultured cells could even have had changes in their genome, which could affect their response towards the chemical cues (Montagnes *et al.* 2008). Finally, experiments with cultured species only assay one species at a time and do not capture the full extent of microbial plot of biological interactions.

The lower number of cells present in the control wells compared to the attractant wells revealed the existence of a specific response of the protist cells, with marked preferences as each chemoattractant promoted a different response within the community. Bacterial exudates generally elucidated the highest chemotactic response, indicating a strongest attraction for bacterial cues than for the other chemicals tested. From them, bacterial exudates of *Roseobacter* and *Flavobacteria* exhibited the strongest response. Probably the three bacterial species exuded different compounds, which generated different chemical cues that could explain the different preferences of the protist community. Positive response to bacterial exudates was already observed in cultures of heterotrophic flagellates (Sibbald *et al.* 1987), and interestingly when a bacteria is washed the attraction disappears, indicating that the chemical cues that bacteria release are more important than the bacteria themselves (Bennett *et al.* 1988). Another explanation for the differences between the three bacterial exudates could be that, as it has been shown in bacterivorous ciliates, protist cells could be more attracted to chemical compounds derived from bacteria that had already been part of their nutrition in contrast to bacteria with no prior exposure (Verity 1991). So, perhaps the *Alteromonas* strain used for the experiment was not frequent in the community where we did the experiment, as our bacterial cultures were isolated from a different area.



Sugars elicit a positive chemotactic response, but weaker than the bacterial exudates. From the three sugars tested, only galactose exhibited a significant chemotaxis. It is important to highlight that for glucose two out of the three replicates were very similar, whereas one had a drastic reduction in the number of cells in the ISCA. For this reason, most of the responses to glucose were not significant, even though a trend was observed. Marine phytoplankton release amount of sugars that attract bacteria, so this reaction to sugars could be acting as a proxy for detecting bacteria (Wolfe 2000). Finally, and interestingly, no response was found towards ammonium even it is know that bacteria respond to patches of ammonium (Dennis *et al.* 2013) and previous culture experiments had shown a positive response of different protist (Govorunova & Sineshchekov 2005).

With respect to the different cell sizes present in the community, overall smaller cells were the ones with higher chemotactic index. However cells bigger than 10 μm, mainly ciliates and dinoflagellates, increase their abundance significantly only in *Roseobacter* exudates, indicating that from all the attractants tested is the one that generated the strongest chemical cue for this bigger protists, some of them of more than 30 μm size.

In order to disentangle the protist behavior, it will be very interesting to have access to the genomes of the different organisms reacting, thus this will give the possibility to identify specific receptors involved in the response. For instance, the completed sequencing of the genome of *Chlamydomonas reinhardtii* facilitated the identification of rhodopsin receptor for phototaxis and helped in explaining their photophobic response (Govorunova & Sineshchekov 2005). Nowadays the existence of recently developed techniques such as SAGs (Single Amplified Genomes) may facilitate this task.

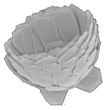
This experiment adds a new light in the behavior of protist communities by comparing completely different attractants at the same time. However different species could have a different reaction depending on the attractant, positive or negative, and whereas some species could be attracted others could be repelled, sometimes based on the history of the individual, thus it has been shown that protist have chemical memory (Verity 1991; Montagnes *et al.* 2008), which makes more difficult the comprehension of all the reactions. In this work we have assessed the protist response carefully characterizing the community using microscopy. However, the ISCA gives us the possibility of recover the sample and do a DNA sequencing which will elucidate which specific taxa of the protist community respond to each attractant, or also metagenomics to identify the different gene expression in the different attractants.

CONCLUSIONS

To our knowledge this is the first chemotactic experiment using protist natural assemblages. It shows a protist response to the different attractants, with a preference to the bacterial exudates above the other chemicals. Overall there were also differences when looking at the morphology of the cells reacting to the different chemical cues, phototrophic cells showed a higher chemotactic index than heterotrophic cells, and according to their size, the smaller cells (1-3 μm) were the ones with strong response.

ACKNOWLEDGEMENTS

This research was funded by Spanish projects ALLFLAGS (CTM2016-75083-R, MINECO) and INTERACTOMICS (CTM2015-69936-P, MINECO/FEDER, EU). CRG was supported by a Spanish FPI grant.



REFERENCES

Ayo, B., Latatu, A., Artolozaga, I., JÜRgens, K. & Iriberri, J. (2009). Factors Affecting Preference Responses of the Freshwater CiliateUronema nigricansto Bacterial Prey. *Journal of Eukaryotic Microbiology*, 56, 188-193.

Ayo, B., Txakartegi, A., Bana, Z., Artolozaga, I. & Iriberri, J. (2010). Chemosensory response of marine flagellate towards L- and D- dissolved free amino acids generated during heavy grazing on bacteria. *International microbiology : the official journal of the Spanish Society for Microbiology*, 13, 151-158.

Azam, F. (1998). Microbial control of oceanic carbon flux: the plot thickens. *Science*, 280, 694-696.

Bennett, S.J., Sanders, R.W. & Porter, K.G. (1988). Chemosensory responses of heterotrophic and mixotrophic flagellates to potential food sources. *Bulletin of marine science*.

Blackburn, N. & Fenchel, T. (1999). Modelling of microscale patch encounter by chemotactic protozoa. *Protist*, 150, 337-343.

Blackburn, N., Fenchel, T. & Mitchell, J. (1998). Microscale Nutrient Patches in Planktonic Habitats Shown by Chemotactic Bacteria. *Science*, 282, 2254-2256.

del Campo, J., Balague, V., Forn, I., Lekunberri, I. & Massana, R. (2013). Culturing bias in marine heterotrophic flagellates analyzed through seawater enrichment incubations. *Microbial ecology*, 66, 489-499.

Dennis, P.G., Seymour, J., Kumbun, K. & Tyson, G.W. (2013). Diverse populations of lake water bacteria exhibit chemotaxis towards inorganic nutrients. *The ISME journal*, 7, 1661-1664.

Fan, K.W., Vrijmoed, L.L.P. & Jones, E.B.G. (2002). Zoospore chemotaxis of mangrove thraustochytrids from Hong Kong. *Mycologia*, 94, 569-578.

Fenchel, T. (1982). Ecology of heterotrophic microflagellates. IV. Quantitative ocurrence and importance as bacterial consumers. *Marine Ecology Progress Series*, 9, 35-42.

Fenchel, T. (1990). Adaptative significance of polymorphic life cycles in Protozoa: responses to starvation and refeeding in two species of marine ciliates. *Journal of Experimental Marine Biology and Ecology*, 136, 159-177.

Fenchel, T. (2002). Microbial behabior in a heterogeneous world. *Science*, 296, 1068-1071.

Fenchel, T. & Blackburn, N. (1999). Motile chemosensory behavior of phagotrophic Protists: Mechanisms for and efficiency in congregating at food patches. *Protist*, 150, 325-336.

González, J.M., Sherr, E.B. & Sherr, B.F. (1993). DIfferential feeding by marine flagellates on growing versus starcing, and on motile versus nonmotile, bacterial prey *Marine Ecology Progress Series*, 201, 257-267.

Govorunova, E.G. & Sineshchekov, O.A. (2005). Chemotaxis in the green flagellate alga Chlamydomonas. *Biochemistry (Moscow)*, 70, 717-725.

Jennings, H. (1906). *Behavior of the lower organisms*. Columbia University Press, New York.

Jurgens, K. & Matz, C. (2002). Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie van Leeuwenhoek*, 81.

Lee, E.S., Lewitus, A.J. & Zimmer, R.K. (1999). Chemoreception in a marine cryptophyte: Behavioral plasticity in response to amino acids and nitrate. *Limnology and Oceanography*, 44, 1571-1574.

Martel, C.M. (2006). Prey location, recognition and ingestion by the phagotrophic marine dinoflagellate Oxyrrhis marina. *Journal of Experimental Marine Biology and Ecology*, 335, 210-220.

Massana, R., Guillou, L., Terrado, R., Forn, I. & Pedrós-Alió, C. (2006). Growth of uncultured heterotrophic flagellates in unamended seawater incubation. *Aquatic Microbial Ecology*, 45.

Mitchell, G.C., Baker, J.H. & Sleigh, M.A. (1988). Feeding of a freshwater flagellate, Bodo saltans, on diverse bacteria. *J. Protozool*.

Mitchell, J.G., Okubo, A. & Furhman, J.A. (1985). Microzones surrounding

phytoplankton form the basis for a stratified marine microbial ecosystem. *Nature*, 316, 58-59.

Montagnes, D.J.S., Barbosa, A.B., Boenigk, J., Davidson, K., Jürgens, K., Macek, M. *et al.* (2008). Selective feeding behaviour of key free-living protists: avenues for continued study. *Aquatic Microbial Ecology*, 53, 83-98.

Pfandl, L., Posch, T. & Boenigk, J. (2004). Unexpected effects of prey dimensions and morphologies on the size selective feeding by two bacterivorous flagellates (Ochromonas sp. and Spumella sp.). *J. Eukaryotic Microbiology*, 51, 626-663.

Roberts, E.C., Legrand, C., Steinke, M. & Wootton, E.C. (2011). Mechanisms underlying chemical interactions between predatory planktonic protists and their prey. *Journal of Plankton Research*, 33, 833-841.

Seymour, J.R., Ahmed, T., Marcos & Stocker, R. (2008). A microfluidic chemotaxis assay to study microbial behavior in diffusing nutrient patches. *Limnology and Oceanography: Methods*, 6, 477-488.

Seymour, J.R., Simo, R., Ahmed, T. & Stocker, R. (2010). Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. *Science*, 329, 342-345.

Sibbald, M., Albright, L. & Sibbald, P. (1987). Chemosensory responses of a heterotrophic microflagellate to bacteria and several nitrogen compounds. *Marine Ecology Progress Series*, 36, 201-204.

Snyder, R.A. (1991). Chemoattraction of a bactivorous ciliate to bacteria surface compounds. *Hydrobiologia*, 215, 205-213.

Stocker, R. (2012). Marine microbes see a sea of gradients. *Science*, 338, 628-633.

Stocker, R. & Seymour, J.R. (2012). Ecology and physics of bacterial chemotaxis in the ocean. *Microbiology and molecular biology reviews : MMBR*, 76, 792-812.

Strom, S.L. & Loukos, H. (1998). Selective feeding by protozoa: model and experimental behaviors and their consequences for population stability. *Journal of Plankton Research*, 20, 831-846.

Strom, S.L., Wolfe, G.V. & Bright, K.J. (2007). Responses of marine planktonic protists to amino acids: feeding inhibition and swimming behavior in the ciliate Favella sp. *Aquatic Microbial Ecology*, 47, 107-121.

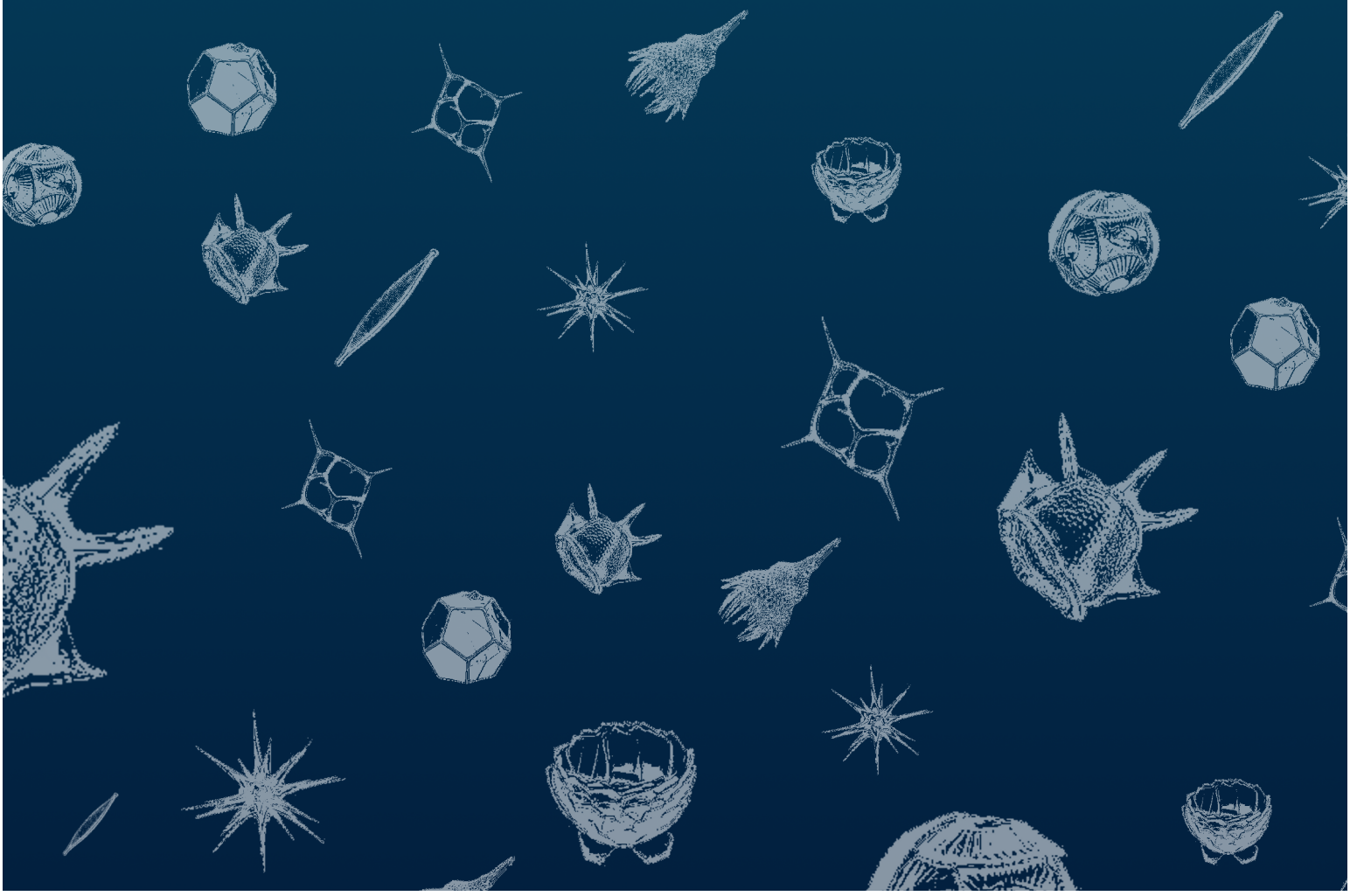
Tout, J., Jeffries, T.C., Petrou, K., Tyson, G.W., Webster, N.S., Garren, M. *et al.* (2015). Chemotaxis by natural populations of coral reef bacteria. *The ISME journal*, 9, 1764-1777.

Verity, P.G. (1991). Feeding in planktonic protozoans: Evidence for non-random acquisition of prey. *J. Protozool*, 38, 69-76.

Wolfe, G.V. (2000). The chemical defense ecology of marine unicellular plankton: constraints, mechanisms, and impacts. *Biol. Bull.*, 198, 224-244.

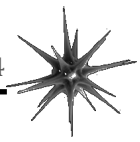
Chapter 4

Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes



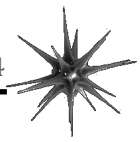
Caterina R. Giner, Irene Forn, Sarah Romac, Ramiro Logares, Colomban de Vargas, and Ramon Massana

Published in *Applied and Environmental Microbiology* (2016). DOI: 10.1128/AEM.00560-16



ABSTRACT

High-throughput sequencing (HTS) is revolutionizing environmental surveys of microbial diversity in the three domains of life by providing detailed information on which taxa are present in microbial assemblages. However, it is still unclear how the relative abundance of specific taxa gathered by HTS correlates with cell abundance. Here, we quantified the relative cell abundance of 6 picoeukaryotic taxa in 13 planktonic samples from 6 European coastal sites using epifluorescence microscopy on tyramide signal amplification-fluorescence in situ hybridization preparations. These relative abundance values were then compared with HTS data obtained in three separate molecular surveys: 454 sequencing the V4 region of the 18S ribosomal DNA (rDNA) using DNA and RNA extracts (DNA-V4 and cDNA-V4), and Illumina sequencing the V9 region (cDNA-V9). The microscopic and molecular signals were generally correlated, indicating that a relative increase in specific 18S rDNA was the result of a large proportion of cells in the given taxa. Despite these positive correlations, the slopes often deviated from 1, precluding a direct translation of sequences to cells. Our data highlighted differences depending on nucleic-acid template or the 18S rDNA region targeted. Thus, the molecular signal obtained using cDNA templates was always closer to relative cell abundances, while the V4 and V9 regions gave better results depending on the taxa. Our data support the quantitative use of HTS data but warn about considering it as direct proxy of absolute cell numbers.



INTRODUCTION

Protists are key components of marine ecosystems, being major players in the global respiration and production budgets (Boenigk *et al*. 2002, Medinger *et al*. 2010) and playing central roles in marine food webs (Sherr *et al*. 2002). Despite their importance and ubiquity, it was only during the past decade that environmental studies, based on molecular (i.e. culture-independent) techniques, revealed an unsuspected protist diversity in a large variety of marine ecosystems (López-García *et al.* 2001, Moon-van der Staay *et al*. 2001, Amaral-Zettler *et al*. 2002, Dawson *et al.* 2002, Stoeck *et al.* 2003, Berney *et al*. 2004, Lovejoy *et al*. 2006, Not *et al.* 2007, Guillou *et al.* 2008, Massana *et al*. 2008). These studies were based on the analysis of 18S rRNA genes retrieved directly from natural assemblages by PCR amplification, cloning and sequencing. Now, the development and use of high-throughput sequencing tools (HTS), e.g. 454 or Illumina, which produce thousands of sequences from a single sample, has revolutionized the field, allowing deeper assessments of diversity (Bik et al. 2012) as well as better estimates of specific relative abundances. One of the main challenges of this approach, however, is to understand the correspondence between the relative abundances and cells that is, how close the specific diversity detected in molecular surveys is to the true species composition of natural assemblages.

Few studies have analyzed the relationship between direct microscopic inspections and sequencing data in protists. One of the first studies compared cloning and sequencing results with an accurate list of protists species (5-100 µm size range) identified by microscopy (Savin *et al.* 2004). In that case, as the sequencing effort was very limited (fewer than 100 clones), few of the protists identified by morphology were detected in the sequencing set. In addition, the few sequences obtained did not represent the dominant observed species, a clear sign of the biases in this molecular approach. More recent comparative studies used HTS, and therefore were not limited by the sequencing effort but focused on specific taxa, in particular marine and freshwater ciliates (Medinger *et al.* 2010, Bachy *et al.* 2012, Santoferrara *et al.* 2014, Stoeck *et al*. 2014). Ciliate species have the advantage of having conspicuous morphological traits that allow proper identification by inverted microscopy. In most cases, the same species were found in microscopic and molecular datasets, but the relative abundance of sequences and morphotypes were not in agreement, so each approach revealed a different community structure. Other studies prepared mock communities, and the results obtained were similar: all individual taxa were detected, but the relative proportion of sequence types was different from cell mixes (Egge *et al*. 2013, Weber *et al*. 2013). Overall, the popularization of HTS now allows a high-resolution

exploration of protist richness present in natural samples; yet, when it comes to evenness, the picture obtained is still limited.

Among protists, picoeukaryotes (protists up to 3 μm in size) are known to be very diverse, widely distributed, and ecologically important in the marine plankton realm (Massana 2011). Picoeukaryotes are counted as a group by epifluorescence microscopy using a general DNA stain (Porter *et al*. 1980) or by flow cytometry (Marie *et al.* 2000), but due to their small size and lack of morphological traits (Potter *et al.* 1997), they cannot be taxonomically identified by these tools. This can be achieved with fluorescence in situ hybridation (FISH), which enables the visualization and quantification of specific cells in natural assemblages by using oligonucleotide probes as phylogenetic stains (Delong *et al.* 1989). FISH has served to identify the cells from novel environmental clades (Not *et al.* 2007, Massana *et al*. 2002, Chambouvet *et al*. 2008), and has been applied in a few marine surveys (Not *et al.* 2004, Massana *et al.* 2006, Siano *et al.* 2010, Lin *et al*. 2012). However, this approach is relatively time consuming and targets only one taxon at a time.

In this study, we assessed the feasibility of using HTS data as a quantitative metric in picoeukaryote diversity studies by comparing relative HTS read abundances with relative FISH cell counts in selected picoeukaryotic taxa. Unlike the previous studies, in which a single taxa (ciliates) or artificial communities were analyzed, this study focused in a set of highly divergent lineages found in geographically separated and unrelated microbial assemblages. Any pattern emerging from this heterogeneous and noisy dataset was expected to be rather robust. We also investigated if there was a difference in community composition assessed by using environmental DNA or RNA extracts as templates (DNA and cDNA reads, respectively), sequencing different regions of the 18S ribosomal DNA (rDNA) (V4 versus V9) or using different HTS platforms (454 versus Illumina). To address these questions, we used published sequencing data sets from several European coastal samples (Massana *et al*., 2015 for DNA/cDNA-V4 and Logares *et al*., 2014 for cDNA-V9) and chose 6 picoeukaryote taxa (<3 μm) for which we had specific FISH probes for quantification.

MATERIALS AND METHODS

Sampling

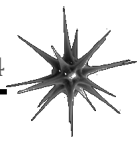
Samples were taken during the BioMarKs project (http://www.biomarks.org) in six European coastal sites near Blanes (Spain, 41° 40’ N, 2° 48’ E), Gijon (Spain, 43° 40' N; 5° 35' W), Naples (Italy, 40° 48’ N, 14° 15’ E), Oslo (Norway, 59° 16’ N, 10° 43’ E), Roscoff (France, 48° 46’ N, 3° 57’ W) and Varna (Bulgaria, 43°10’ N, 28° 50’ E) (Table 1). Seawater was collected with Niskin bottles attached to a conductivity-temperature-depth rosette at surface and deep chlorophyll maximum (DCM) depths. For molecular surveys, ~20 L of seawater was pre-filtered through a 20 µm-pore-size metallic mesh and then sequentially filtered through 3- and 0.8 µm-pore-size polycarbonate filters (142 mm diameter). The 0.8 µm-pore-size polycarbonate filter contained the picoplankton (0.8- to 3-µm size fraction) and was flash frozen and stored at -80ºC. The filtration time was less than 30 minutes to avoid RNA degradation.

Unfiltered seawater was taken for direct cell counts. For total microscopic counts, seawater samples were fixed with glutaraldehyde (1% final concentration) and left for 1 to 24 h at 4°C. Then, aliquots of 20 ml were filtered through 0.6 µm-pore-size polycarbonate black filters and stained with DAPI (4’,6-diamidino-2-phenylindole) at 5 g · ml^{-1}. Filters were mounted on a slide and stored at 20°C until processed. For specific counts with tyramide signal amplification (TSA)-FISH, aliquots of 100 ml were fixed with filtered formaldehyde (3.7% final concentration), incubated for 1 to 24 h in the dark at 4°C, and filtered through 0.6 µm-pore-size polycarbonate filters (25-mm diameter). Filters were kept at -80°C until processed. For flow cytometry counting of photosynthetic picoeukaryotes, aliquots of 1.5 ml were fixed with a mix of paraformaldehyde and glutaraldehyde (1% and 0.25% final concentrations, respectively), frozen in liquid nitrogen, and stored at -80°C until processed.

Sampling site	Date	Depth (m)	Temp. (ºC)	DAPI counts		Flow cytometry	%	%
				Phototrophs	Heterotrophs	Phototrophs	Phototr.[1]	Heterotr.[1]
Blanes	Feb. 2010	1 (Surf.)	12.5	9273	445	9215	48.6	53.7
Gijon	Sep. 2010	1 (Surf.)	20.2	1606	2503	2990	14.5	20.2
Naples	Oct. 2009	1 (Surf.)	22.8	*	*	2714	-	-
		26 (DCM)	22.4	*	*	2049	-	-
	May 2010	1 (Surf.)	19.2	4376	4372	4700	1.1	54.6
		34 (DCM)	15.5	1808	1331	1802	8.3	28.8
Oslo	Sep. 2009	1 (Surf.)	15.0	12342	4470	9540	12.4	21.9
		20 (DCM)	15.0	8773	2807	8930	17.9	38.4
	Jun. 2010	1 (Surf.)	15.0	7727	2893	13295	25.5	7.9
		10 (DCM)	12.5	21523	2823	17900	22.9	40.7
Roscoff	Apr. 2010	1 (Surf.)	9.9	7203	1034	8240	43.9	68.9
Varna	May 2010	1 (Surf.)	21.5	*	*	3861	-	-
		40 (DCM)	9.5	7043	731	9487	24.9	24.6

1. These columns show the percentage of phototrophic and heterotrophic cells targeted by the utilized probes.
*. DAPI counts were not performed, so picoeukaryotes could not be differentiated between phototrophs and heterotrophs. In these samples, total picoeukaryote counts were done on FISH filters and were: 4272 cells ml^{-1} in Naples-2009 Surf, 1834 cells ml-1 in Naples-2009 DCM, and 4656 cells ml^{-1} in Varna Surf. These values were used in the correlations.

Table 1. Planktonic samples analyzed (sampling site, date, depth and seawater temperature) and cell counts (cells ml^{-1}) in these samples. The total picoeukaryote abundance (cells ≤3 μm) was determined by DAPI (phototrophs and heterotrophs), and the photosynthetic picoeukaryote abundance was determined by flow cytometry.



Picoeukaryote cell abundance by DAPI staining and flow cytometry

The total cell abundance of picoeukaryotes was estimated in DAPI-stained filters. Cells were counted with an epifluorescence microscope (Olympus BX61) at 1,000X under UV excitation, changing to blue light excitation to verify the presence or absence of chlorophyll autofluorescence (phototrophic and heterotrophic cells, respectively). A transect of about 13 mm was inspected, and cells were classified in size classes: 2 μm, 3 μm, 4 μm, 5 μm, and >5 μm. All data reported in the study referred to cells within the two smaller size classes (2 to 3 μm), which accounted on average for 82% of the cells.

Cell abundance of photosynthetic picoeukaryotes was determined in a FACSort flow cytometer by using the red fluorescence signal (chlorophyll) after excitation in a 488 nm laser and the side-scattered light of each particle. Fluorescent microspheres (0.95 μm beads) were added as an internal standard (at 10^5 beads · ml^{-1}). Data were acquired for 2 to 4 min with a flow rate of 50 to 100μl · min^{-1} using the settings previously described (Marie *et al*. 1999).

Cell abundance of specific picoeukaryote taxa by TSA-FISH

The specific oligonucleotide probes used targeted several picoeukaryote taxa: NS4 and NS7 targeted the uncultured clades MAST-4 and MAST-7; CRN02 and MICRO01, the species *Minorisa minuta* and *Micromonas spp*.; PELA01, the class Pelagophyceae; and ALV01, the environmental clade MALV-II (Table 2). These probes have been published in other studies (see references cited in Table 2) except NS7. Probe NS7 was designed here with ARB (Quast *et al.* 2013) and targeted 91% of the 192 sequences from MAST-7 available in GenBank; it had 1 mismatch with the remaining MAST-7 sequences and had at least 2 central mismatches with nontarget sequences. Probe NS7 gave a better signal when combined with oligonucleotide helpers contiguous to the probe region (NS7 helper A: AACCAACAAAATAGCAC; NS7 helper B: CCCAACTATCCCTATTAA) that were added to the hybridization buffer at the same concentration as the probe. We tested a range of formamide concentrations to find the best hybridization condition, and we checked that the probe gave a negative signal with a variety of nontarget cultures. Finally, a probe targeting all eukaryotes (EUK502, Lim *et al.* 1999) was also used. All probes were labeled with horseradish peroxidase (HRP).

Probe Name	Target group	Probe sequence (5' – 3')	Probe reference	Num. of reads per Taxa		% reads - probe
				In OTU table	From the raw reads	
NS4	MAST-4	TACTTCGGTCTGCAAACC	Massana *et al.,* 2002	2082	2082	98.0
NS7	MAST-7	TCATTACCATAGTACGCA	This study	2842	2833	95.7
CRN02	*Minorisa minuta*	TACTTAGCTCTCAGAACC	del Campo *et al.*, 2012	1853	1853	99.8
PELA01	Pelagophyceae	ACGTCCTTGTTCGACGCT	Not *et al.*, 2002	4440	3169	98.5
MICRO01	*Micromonas spp.*	AATGGAACACCGCCGGCG	Not *et al.,* 2004	11,166	-	-
ALV01	MALV-II	GCCTGCCGTGAACACTCT	Chambouvet *et al.*, 2008	35,359	29,894	83.0
EUK502	Eukaryotes	GCACCAGACTTGCCCTCC	Lim *et al.*, 1999	-	-	-

Table 2. List of oligonucleotide FISH probes used and effectiveness of the probes against reads from this study (% reads by probe). The table shows the number of 454 reads from each phylogenetic group extracted from the OTU table or from raw reads by local BLAST using seeds. The last column shows the percentage of raw reads in each group that have the probe target region with 0 mismatches.

Hybridizations were performed as previously described (Pernice *et al.* 2015). Filter pieces (about 1/10) of the 0.6 µm-pore-size polycarbonate filters were covered with 20 µl of hybridization buffer (40% deionized formamide [except 30% for probe CNR01], 0.9 M NaCl, 20 mM Tris-HCl [pH 8], 0.01% SDS) and 2 µl of HRP-labeled probes (stock at 50 $ng \cdot l^{-1}$) and incubated overnight at 35°C. After the hybridization, filter pieces were washed twice for 10 min at 37°C with a washing buffer (37 mM NaCl [74 mM NaCl when hybridizing with 20% formamide], 5 mM EDTA, 0.01% SDS, and 20 mM Tris-HCl [pH 8]) and transferred to phosphate-buffered saline (PBS) for 15 min at room temperature. TSA was carried out in a solution (1 PBS, 2 M NaCl, 1 $mg \cdot ml^{-1}$ blocking reagent, 100 $mg \cdot ml^{-1}$ dextran sulfate, and 0.0015% H_2O_2) containing Alexa 488-labeled tyramide (4 $\mu g \cdot ml^{-1}$) by incubating in the dark at room temperature for 30 to 60 min. Filter pieces were transferred twice to a PBS bath in order to stop the enzymatic reaction and air dried at room temperature. Cells were countersained with DAPI (5 $\mu g \cdot ml^{-1}$), and filter pieces were mounted on a slide. Targeted FISH cells were counted by epifluorescence under blue light excitation and checked with UV radiation (DAPI staining) for the presence of the nucleus. Cells labeled with the probe EUK502 were counted using the same size classes as for DAPI counts. Data reported refer to cells of 2- to 3 µm sizes, which accounted on average for 84% of the cells.

High-throughput sequencing by 454 and Illumina

HTS data derive from papers published during the BioMarKs project (http://www.biomarks.eu/). Total DNA and RNA from 13 picoplankton samples were extracted

simultaneously from the same filter. For RNA extracts, contaminating DNA was removed, and RNA was immediately reverse transcribed to cDNA. Data for the 454 sequencing are derived from the work by Massana et al. (Massana *et al*. 2005) and used the eukaryotic universal primers TAReuk454FWD1 and TAReukREV3 (Stoeck *et al*. 2010), which amplified the V4 region of the 18S rDNA (380 bp). Amplicon sequencing from DNA and cDNA templates was carried out on a 454 GS FLX Titanium system (454 Life Sciences, USA) in Genoscope (http://www .genoscope.cns.fr). The complete sequencing data set is available at the European Nucleotide Archive (ENA) under the accession number PRJEB9133. Data for the Illumina sequencing are derived from the work by Logares et al. (Logares *et al*. 2014) and used the eukaryotic universal primers 1398f and 1510r (Amaral-Zettler *et al*. 2009), which amplified the V9 region of the 18S rDNA (130 bp). Paired-end 100-bp sequencing was performed using a Genome Analyzer IIx (GAIIx) system located at Genoscope. Only RNA (cDNA) samples were sequenced with Illumina. Sequences are publicly available at MG-RAST (http://metagenomics.anl.gov) under accession numbers 4549958.3, 4549965.3, 4549959.3, 4549945.3, 4549943.3, 4549927.3, 4549941.3, 4549954.3, and 4549922.3.

Sequence analysis of HTS reads

HTS reads by 454 and Illumina were quality checked following criteria similar to those detailed in the original papers (Massana *et al*. 2015, Logares *et al.* 2014). After the quality control, chimera detection was run with UCHIME (Edgar *et al*. 2011) and ChimeraSlayer (Haas *et al.* 2011) using SILVA108 and PR2 (Guillou *et al.* 2013) as reference databases. The final curated reads were clustered into operational taxonomic units (OTUs) by using UCLUST 1.2.22 (Edgar *et al.* 2010), with similarity thresholds of 97% for V4 reads and 95% for V9 reads. Representative reads of each OTU were taxonomically classified by using BLAST against SILVA108, PR2, and a marine microeukaryote database (Pernice *et al*. 2013). After the taxonomic assignment, metazoan OTUs were removed. From the complete OTU tables for 454 (Massana *et al*. 2015) and Illumina (Logares *et al.* 2014) data sets, the samples targeting the picoplankton were extracted: 13 samples for DNA-V4, 13 samples for cDNA-V4, and 9 samples for cDNA-V9. Then, OTUs corresponding to taxa typically larger than 3 µm (Dinophyceae, Ciliophora, Acantharia, Diatomea, Polycystinea, Raphidophyceae, Ulvophyceae, Rhodophyta, and Xanthophyceae; in this order of relative abundance) were removed. These groups accounted for 8.0% to 87.7% (average, 36.9%) of the 454 data set and 11.5% to 73.5% (average, 33.9%) of the Illumina data set. The read numbers in the final OTU tables of picoeukaryotes were 110,258 for DNA-V4, 77,554 for cDNA-V4, and 1,753,600 for cDNA-V9.

The relative abundance of the picoeukaryotic groups of interest was retrieved from these taxonomically classified OTU tables, by dividing the number of reads of the specific OTUs corresponding to the groups of interest by the total number of reads in the sample. Altogether, the six taxa of interest accounted for 36.4% of the DNA-V4 reads, 23.5% of the cDNA-V4 reads, and 32.4% of the cDNA-V9 reads. In addition to the taxonomic classification of OTUs in the OTU table, we classified the unclustered 454 and Illumina reads to obtain the raw reads for probe checking (see Results) and to double-check the taxonomic classification. For this second classification, we downloaded GenBank sequences representative of each picoeukaryotic group of interest and used this specific taxon database to retrieve HTS reads by local BLAST (sequence similarity, >97%).

RESULTS

An overview of total picoeukaryote counts in marine coastal waters

We estimated the total cell abundance of picoeukaryotes by epifluorescence microscopy and flow cytometry in 13 planktonic samples taken in 6 geographically separated European coastal sites and different depths (Table 1). Total picoeukaryote counts (cells <3 μm) by epifluorescence microscopy of DAPI-stained samples revealed a wide range of cell abundances, from 3,139 cells·ml^{-1} in Naples-2010 DCM to 24,346 cells·ml^{-1} in Oslo-2010 DCM (average in all samples, 10,500 cells·ml^{-1}). Phototrophic and heterotrophic cells were differentiated while counting the DAPI samples. The total abundance of phototrophic cells was generally higher than that of heterotrophic cells (average, 8,200 and 2,400 cells·ml^{-1}, respectively), with the exception of Naples-2010 surface, where the two assemblages had similar abundances. In some cases (Blanes, Oslo-2010 DCM, Roscoff, and Varna DCM), phototrophic cells were 6 times more abundant than heterotrophic cells. Counts of phototrophic picoeukaryotes obtained by flow cytometry correlated well with the microscopic counts in the 10 samples analyzed (linear slope, 0.74; Pearson's r=0.9; P<0.001). When the regression line was forced to intercept at 0, the slope was 0.90.

The general eukaryotic probe EUK502 was also used to estimate total picoeukaryotic abundance. Cell counts by TSA-FISH were always lower than the DAPI counts (60% on average) (Fig. 1). In fact, the sample with the highest total cell abundance was different if estimated by DAPI (Oslo-2010 DCM) or by TSA-FISH (Oslo-2009 surface). The regression between the two

data sets was significant but had a slope very distant from 1 (linear slope, 0.26; Pearson's r=0.74; P<0.05). When the line was forced to intercept at 0, the slope was still very low (0.43). There was some tendency to this discrepancy, as TSA-FISH seemed to underestimate more severely the total cell counts in samples dominated by very small cells. Clearly, DAPI counts provided a better estimate than TSA-FISH counts of total picoeukaryotic abundance; therefore, DAPI counts were used to calculate the relative cell abundances of each of the 6 specific picoeukaryotic groups. TSA-FISH counts of each group were in the numerator, and total DAPI counts were in the denominator.

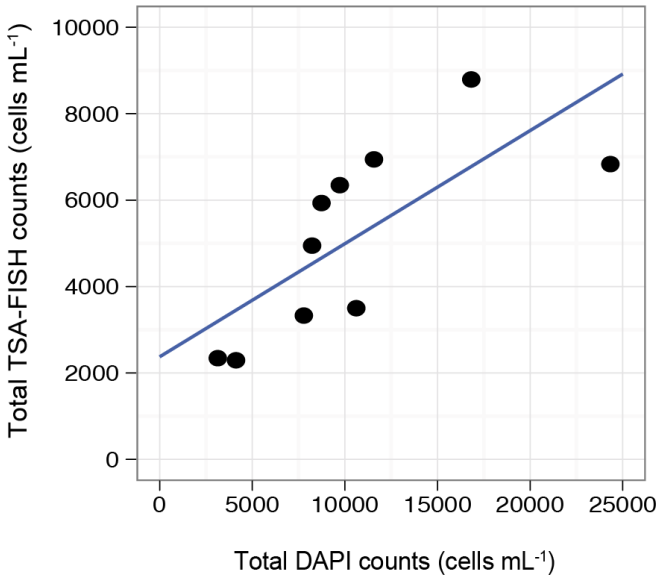


Fig. 1. Comparison of total picoeukaryotic abundance (cells <3 µm) by DAPI counts and FISH counts using the eukaryotic probe EUK502 in all planktonic samples.

Abundance of specific picoeukaryotic taxa

We used TSA-FISH to estimate the total abundance of six groups of picoeukaryotes, chosen because they were well represented in the sequencing data sets of the picoplankton from the studied samples (and poorly represented in the nanoplankton; see Table 3). They belonged to different eukaryotic supergroups: the Stramenopiles (MAST clades and Pelagophyceae), Alveolates (the parasite clade MALV-II), Archaeplastida (*Micromonas spp*.), and Rhizaria (*Minorisa minuta*). The taxonomic coverage of the probes used varied from being very narrow, targeting a species (*Minorisa minuta*) or a constrained phylogenetic clade (*Micromonas spp*. and

the MAST lineages), to being very wide, targeting an algal class (Pelagophyceae) or the diverse MALV-II group (formed by 44 phylogenetic clades). The sum of heterotrophic cells (MASTs, *M. minuta*, and MALV-II) represented, on average, 36% of heterotrophic picoeukaryotes counted by DAPI, whereas the phototrophic cells targeted (*Micromonas* and Pelagophyceae) represented, on average, only 22% of phototrophic picoeukaryotes (Table 1).

	Picoplankton		Nanoplankton	
	DNA	cDNA	DNA	cDNA
MAST-4	0.74	0.65	0.04	0.02
MAST-7	0.68	1.24	0.09	0.13
Minorisa minuta	0.25	1.01	0.02	0.04
Pelagophyceae	0.40	2.62	0.11	0.65
Micromonas spp.	2.98	4.51	0.27	0.24
MALV-II	21.07	2.15	2.83	1.07

Table 3. Proportion of reads of each picoeukaryotic group of interest obtained in the V4-survey from picoplankton and nanoplankton samples. Percentages derive from the original OTU tables (only metazoans removed).

The cell abundances of the six targeted groups varied strongly among the different samples (see Table 4 in the supplemental material). We found that *Micromonas*, MAST-4, MAST-7, and MALV-II were the most abundant taxa (average cell abundances of 1,492, 279, 160, and 127 cells·ml^{-1}, respectively) and were detected in all samples. *Minorisa minuta* was very abundant in some sites but absent in others. In contrast, Pelagophyceae was the least abundant taxon (average cell abundance of 59 cells·ml^{-1}). These cell counts pointed out that each sample contained a different community. *Micromonas* was the most abundant taxon in 7 samples; MAST-4, in 4 samples; and *Minorisa* and MALV-II, in the other two samples (see Table 4).

In silico validation of the FISH probes against raw V4-reads

Before applying TSA-FISH, we evaluated the effectiveness of the probes against the V4 reads obtained from the same samples. This analysis was done with raw reads (extracted from the initial data set by using GenBank sequences of each group as search templates) to take into account all sequence variants. The number of raw reads per group obtained from this way was very similar to the number derived from the OTU table (Table 2). About 1,000 to 3,000 reads were extracted per group (except MALV-II, which had about 30,000 reads). Then, we calculated the percentage of raw reads that had a 100% match with the probes (Table 2). The five specific probes validated this way retrieved a very high percentage of reads, more than 95%, in all cases

except MALV-II (83%). Therefore, the vast majority of reads from these five groups in our samples had the target region of the probes.

The probe targeting *Micromonas* was not designed at the V4 region of the 18S rDNA, so it could not be directly evaluated with V4 reads from this study. Therefore, we took the OTUs affiliating with *Micromonas* (7 OTUs and 11,166 reads), retrieved the closest GenBank complete sequence from these OTUs (nearly identical at the V4 region), and verified the effectiveness of the probe against these 7 GenBank sequences. Only 3 sequences (accounting for 30% of the reads) exhibited a perfect match, whereas the remaining 4 sequences had a mismatch in the first position of the probe. Thus, probe MICRO01 could be improved perhaps by removing the first base, but since this mismatch is located in the first position, it likely does not affect the FISH counts.

Sampling site	Date	Depth (m)	TSA-FISH counts					
			MAST-4	MAST-7	*M. minuta*	Pelagophyceae	*Micromonas*	MALV-II
Blanes	Feb. 2010	1 (Surf.)	76	118	2	339	4167	43
Gijon	Sep. 2010	1 (Surf.)	279	56	89	78	155	81
Naples	Oct. 2009	1 (Surf.)	543	179	130	34	113	109
		26 (DCM)	141	51	9	108	42	61
	May 2010	1 (Surf.)	555	55	1416	31	16	360
		34 (DCM)	156	43	41	136	14	143
Oslo	Sep. 2009	1 (Surf.)	323	337	28	1	1524	292
		20 (DCM)	373	256	50	6	1563	398
	Jun. 2010	1 (Surf.)	66	140	2	1	1973	20
		10 (DCM)	630	481	0	9	4924	38
Roscoff	Apr. 2010	1 (Surf.)	409	300	0	14	3145	3
Varna	May 2010	1 (Surf.)	5	2	0	0	15	52
		40 (DCM)	69	63	0	8	1743	48

Table 4. Cell abundance of the groups of interest determined by TSA-FISH.

Comparison of group specific read abundance and TSA-FISH counts

The relative abundances of 454 V4 reads (from DNA and cDNA templates) and Illumina V9 reads (from cDNA templates) of each group of interest were compared with the relative cell

abundance assessed by epifluorescence microscopy (specific TSA-FISH counts relative to total DAPI counts) in 13 samples for the V4 reads, and in 9 samples for the V9 reads (DCM samples from Naples and Oslo were excluded) (Fig. 2). The statistics of these plots are shown in Table 5. For the DNA-V4 survey, the correlation of the relative abundance of cells and the DNA reads was significant for all groups ($P<0.05$) except for MAST-4 and Pelagophyceae, and the goodness of these correlations varied among groups; goodness was strongest for Minorisa minuta ($R^2=0.97$) and weakest for MALV-II ($R^2=0.29$). Despite these good correlations, linear slopes of the plots were always different from 1 except for MAST-7. In most cases, slopes were below 0.5, indicating an underestimation of cell abundance by 454 reads, while the slope for MALV-II was very high (4.46), indicating a severe overestimation of the molecular signal in this group.

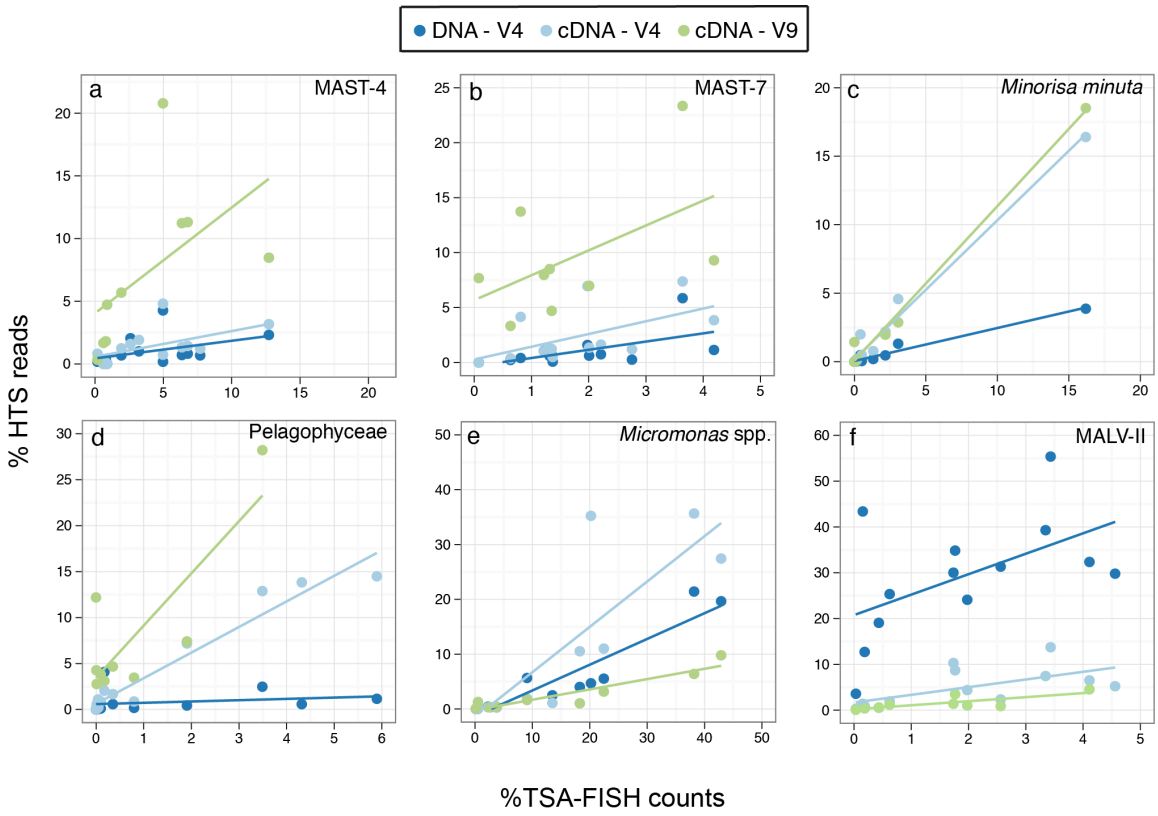
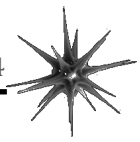


Fig. 2. Comparison of relative abundance of HTS reads against TSA-FISH cell counts in the 13 planktonic samples (9 samples for V9-cDNA reads) for six picoeukaryote taxa: MAST-4 (a), MAST-7 (b), *Minorisa minuta* (c), Pelagophyceae (d), *Micromonas spp*. (e) and MALV-II (f). Dark blue symbols indicate V4-DNA reads, light blue V4-cDNA reads and green V9-cDNA reads. Regression lines are shown, and their statistics are presented in Table 5.



In contrast, the correlations between relative cell and read abundances in the cDNA-V4 survey were generally better for all groups and also were significant for Pelagophyceae and MAST-4 (Table 5). Similar to the DNA-V4 survey, each group had a different slope, but, in this case, there were three taxa (MAST-7, *M. minuta*, and *Micromonas*) with slopes statistically not different from 1, indicating that their relative abundances obtained by cell counts and 454 reads were comparable. In the 6 groups analyzed, the slopes obtained in the cDNA survey were closer to 1 than the slopes derived from the DNA survey, showing a better performance of the cDNA approach.

For the Illumina cDNA-V9 survey, the correlations were slightly worse than for the cDNA-V4 survey (Fig. 2; Table 5), as they were nonsignificant (P>0.05) for MAST-4 and MAST-7. Regarding the linear slopes, the three groups with good performances at the cDNA-V4 survey (*M. minuta*, Pelagophyceae, and *Micromonas*) had slopes statistically different from 1, indicating that, in these groups, the V4 region (and not the V9 region) could be used as a proxy of cell counts. In contrast, MALV-II had a better correlation with the V9-cDNA reads than with the V4 reads, and its slope was not statistically different from 1. This highlights that there is not a best region that applies to all taxa.

	V4 - 454 survey								V9 - Illumina survey			
	DNA				cDNA				cDNA			
	R^2	slope	P value	p1[a]	R^2	slope	P value	p1[a]	R^2	slope	P value	p1[a]
MAST-4	0.18	0.14	ns[c]	-	0.31	0.21	<0.05	<0.001	0.3	0.84	ns	-
MAST-7	0.33	0.75	<0.05	ns	0.31	1.16	<0.05	ns	0.36	2.79	ns	-
Minorisa minuta	0.97	0.24	<0.001	<0.001	0.98	1.01	<0.001	ns	0.99	1.13	<0.001	<0.001
Pelagophyceae	0.06	0.14	ns	-	0.94	2.78	<0.001	<0.001	0.68	5.68	<0.01	<0.01
***Micromonas* spp.**	0.87	0.47	<0.001	<0.001	0.73	0.83	<0.001	ns	0.87	0.2	<0.001	<0.001
MALV-II	0.29	4.46	<0.05	<0.05	0.39	1.68	<0.05	<0.05	0.60	0.89	<0.05	ns

a. p1 compares the slopes against the desired value of 1 (i.e. "ns" indicates that the slope is not significantly different from 1).
ns: no significant

Table 5. Statistics (R^2, slope value, and p-value) of the correlations between relative abundance of reads and cells in the three molecular surveys. The three molecular surveys are 454 DNA-V4 (Fig. 2, dark blue), 454 cDNA-V4 (Fig. 2, light blue) and Illumina cDNA-V9 (Fig. 2, green).

Differences when targeting V4 and V9 regions of the 18S rDNA

To discard the possibility that the differences observed between the V4 and the V9 regions were due to the use of different sequencing platforms (454 for V4 and Illumina for V9), we sequenced

with Illumina (MiSeq platform) the V4 region of one sample of the data set (Oslo-2009 DCM) using both templates (DNA and cDNA). The relative abundances of 60 taxonomic groups inferred from the same targeted region (V4) in the 2 platforms displayed a very good agreement, with an R^2 of 0.97 and of 0.91 (for DNA and cDNA, respectively), and linear slopes of 0.92 to 1.02. Both slopes were not significantly different from 1. Furthermore, this analysis was done in an additional set of 14 samples (from other planktonic size fractions and sediments; data not shown), and the two platforms performed similarly, with R^2 results ranging from 0.57 to 1.00 (average, 0.91) and slopes ranging from 0.73 to 1.21 (average, 0.99). Therefore, sequencing the same 18S rDNA region with 454 or Illumina (MiSeq) gave highly consistent results.

Therefore, the differences outlined above between V4-454 and V9-Illumina sequencing (Table 5) were due to targeting different 18S rDNA regions and not due to the sequencing platform. In order to observe these differences in more detail, we compared the relative abundances of cDNA-V4 reads and cDNA-V9 reads for the six picoeukaryotic taxa studied here (Fig. 3). Clear and consistent differences were identified in each case. As before, the correlations were good and significant, with R^2 results ranging from 0.68 to 0.98 (lower in MALV-II: 0.45), but the slopes deviated significantly from 1 ($P<0.05$). The V9 analysis significantly increased the relative abundance of the stramenopile groups (the two MAST clades and Pelagophyceae), with slopes ranging from 2.3 to 3.4, while the opposite occurred for *Micromonas* and MALV-II, which had slopes of 0.2 and 0.3, respectively, and the same occurred for *Minorisa minuta* (slope, 1.1).

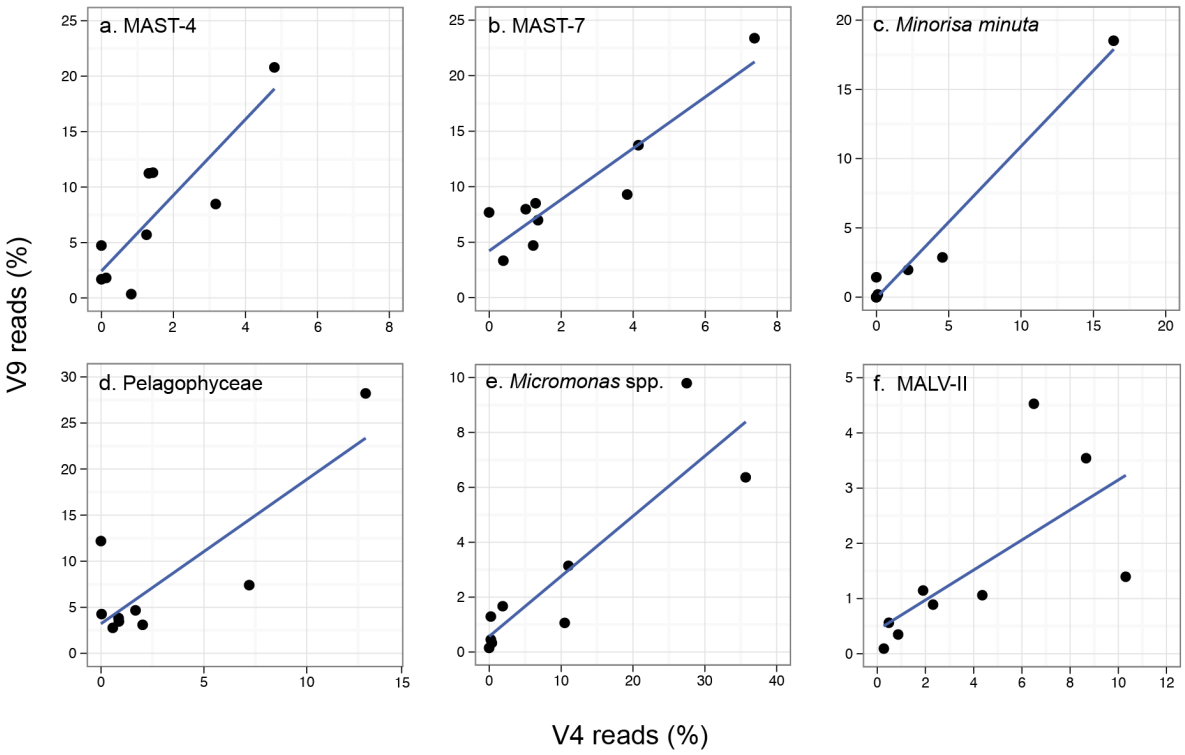


Fig. 3. Comparison of relative abundance of V9-Illumina reads and V4-454 reads (cDNA surveys in both cases) in 9 planktonic samples for six picoeukaryote taxa: MAST-4 (a), MAST-7 (b), *Minorisa minuta* (c), Pelagophyceae (d), *Micromonas spp*. (e) and MALV-II (f).

DISCUSSION

Identifying marine picoeukaryotes by direct microscopy is problematic because of their small sizes, and, as a consequence, there is an increasing interest in using high-throughput sequencing (HTS) technologies to explore their diversity. HTS surveys provide a detailed picture of the taxa present in the community, including rare species in the assemblage (Stoeck *et al*. 2014, Logares *et al*. 2014), and reveal diversity not evident using other methods. However, the interpretation of the HTS signal in terms of total cell abundances is not straightforward. Interestingly, TSA-FISH is able to bridge microscopic and sequencing approaches by using specific phylogenetic probes to estimate true cell abundances (Not *et al*. 2004, Not *et al.* 2002). FISH, besides being very laborious, is limited by the number of taxon-specific probes available as well as by the phylogenetic resolution of the probes (Alonso-Sáez *et al*. 2007). Moreover, TSA-FISH could be inaccurate due to putative mismatches of the probes with the target group, which would result in cell count underestimates. We addressed this issue by evaluating the six probes against sequences obtained from the same samples, and we found an acceptable performance (very

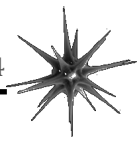
good in four cases: 83% of reads for MALV-II and only one terminal mismatch for *Micromonas*). This validated that the TSA-FISH cell counts performed here were accurate and supported the main objective of this study, which was to evaluate how well the HTS signal estimates community structure in terms of specific abundance.

More sequences imply more cells

Since the HTS signal is always relative (number of reads of a given taxon with respect to the total read number), we needed the total picoeukaryote abundance to calculate relative cell abundances. In principle, using TSA-FISH with a universal eukaryotic probe would be consistent with the study and would also provide an extra layer of certainty, since it allows an easier differentiation of eukaryotic cells from fluorescent particles and large bacteria. However, TSA-FISH counts systematically resulted in fewer cells than direct DAPI counts, and we noticed protists that were not labeled with the EUK502 probe. Moreover, this discrepancy was particularly critical in samples dominated by very small cells. The wide size spectra of protist cells in natural samples implied a large variation in the fluorescent signal, so small cells with dim fluorescence may remain unnoticed when close to large fluorescent cells and may easily fade away while counting a field having many cells with diverse sizes and morphologies. This problem did not happen when using specific probes, since we focused on counting a defined cell type (even with dim fluorescence). Therefore, we used the direct DAPI counts to calculate relative cell abundances.

When comparing the relative abundance of HTS reads against the relative cell abundance obtained by TSA-FISH for the different taxa, we generally found a good correlation between the two methods. The R^2 coefficients of each picoeukaryotic taxon were similar in the three comparisons conducted (DNA-V4, cDNA-V4, and cDNA-V9 versus TSA-FISH), except that there was a very poor correlation for Pelagophyceae in the DNA-V4 survey. Nevertheless, the statistical significance was always better for the cDNA survey than for the DNA survey. These correlations imply that relative read abundance was proportional to relative cell abundance, i.e., an increase in the HTS signal from a particular taxon is the result of an increase of the proportion of targeted cells in the sample. However, the correlation coefficients were far from 1 in most cases, and this noisy signal was probably related to molecular biases plus the large differences in the picoeukaryotic composition of each sample.

Molecular surveys based on a single gene are affected by the widely discussed PCR biases (Wintzingerode *et al*. 1997). During PCR, some phylotypes can be amplified preferentially, some



groups can remain undetected due to primer mismatches (Hong *et al*. 2009), or there could be biases due to the number of PCR cycles (Suzuki *et al*. 1998). Thus, it has been suggested that the relative read abundance can no longer reflect the real composition of the original community, biasing diversity estimates and producing over- or underestimations of specific groups (Medinger *et al*. 2010). Furthermore, sequencing errors may create false or chimeric taxa (Bachy *et al*. 2012, Quince *et al.* 2009, Kunin *et al*. 2010). Our results indicate that PCR biases and putative sequencing artifacts do not affect proportionality between relative read and cell abundance: more reads imply a higher proportion of cells. The significant correlations detected here using this sample data set, where each sample had large differences in the picoeukaryotic composition because they were taken in distant sites and different times of the year, justify the use of relative read abundance as a proxy of community composition for comparative purposes.

Relative abundances of sequences and cells may disagree

Despite the significant correlations discussed above, HTS and TSA-FISH surveys did not give the same quantitative information, as the regression line often was statistically different from 1. Moreover, these slopes varied strongly among the 3 HTS surveys. In order to compare these surveys, we analyzed the relative abundances of the 6 picoeukaryotic groups (among themselves) in the different samples (Fig. 4). This showed a general agreement between TSA-FISH and the two cDNA surveys, but, depending on the composition of the sample, the agreement was better using the V4 region or the V9 region. In samples dominated by Micromonas (e.g., Blanes, Oslo-2010, Roscoff, Varna DCM), the picture obtained with the V4 region better matched the cell abundance, while the V9 region performed better in samples dominated by stramenopiles (MAST-4, MAST-7, Pelagophyceae). In our samples, the cDNA-V4 survey gave a better representation of the true species composition for 5 of the samples, while cDNA-V9 per-formed better in 4 of the samples.

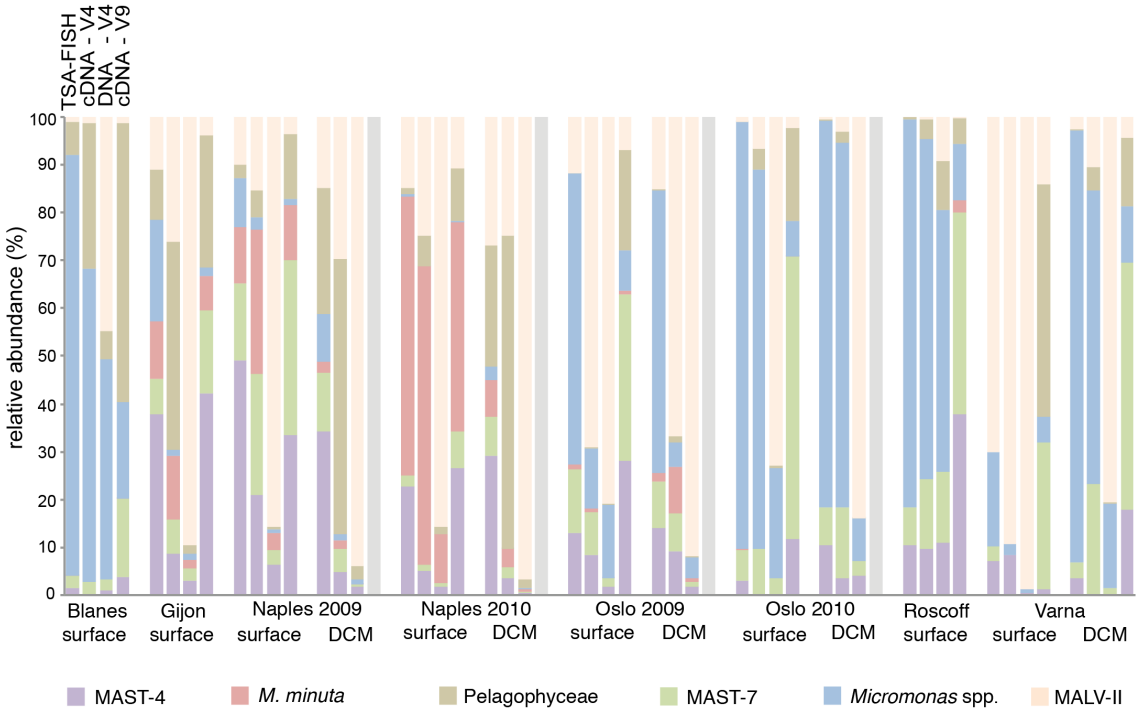


Fig. 4. Relative abundance of the different groups (among themselves) shown by the four approaches (TSA-FISH, V4-cDNA, V4-DNA, V9-cDNA) in all planktonic samples. Gray bars indicate the absence of the sample.

In all cases, the DNA survey gave a more biased perspective of the relative abundance of the 6 picoeukaryotic taxa, being influenced by a very high abundance of MALV-II reads in all samples. This is probably due to a particularly high number of rDNA-operon copies in MALV groups (Medinger *et al.* 2010, Siano *et al*. 2010, Massana *et al.* 2015). The 18S rDNA copy number can vary by orders of magnitude among protist taxa, from a few copies per cell in some green algae (Zhu *et al*. 2005) to about 30 copies in MAST-4 (Rodríguez-Martínez *et al.* 2009) or several thousand copies in some dinoflagellates (Zhu *et al*. 2005), depending on the cell size and genome size (Prokopowich *et al.* 2003). Large differences in the copy number of the targeted gene will affect the abundance estimates in DNA surveys (Medinger *et al*. 2010). Moreover, reads retrieved in DNA surveys could derive from dead organisms or dis-solved extracellular DNA. It is known that dissolved DNA is pre-served in marine waters (Danovaro *et al*. 2005), escaping from degradation and persisting for different periods of time, from hours to days (Nielsen *et al*. 2007). In contrast, reads from cDNA surveys derive from ribosomes and represent metabolically active taxa in the community, as ribosomes are needed to perform the RNA translation in metabolically active cells (Stoeck *et al.* 2007, Not *et al*. 2009). This, in addition to the 18S rDNA copy number, could explain the differences observed between DNA and cDNA surveys. Moreover, our data also

highlighted the impact of targeting different regions of the 18S rDNA gene for estimating relative abundances. For example, the cDNA-V9 survey showed a higher signal (more reads) for MAST taxa and a lower signal for Micromonas compared with cDNA-V4. It is known that the ranges of taxonomic groups detected by V4 and V9 are different (Stoeck *et al*. 2010, Dunthorn *et al*. 2012, Decelle *et al*. 2014) and that some groups can be over- or underrepresented. In particular, in our samples, the V4 region gave good estimates of cell counts for MAST-7 and *Micromonas spp*.; the V9, for MALV-II; and both regions, for *Minorisa minuta*. Thus, the region targeted (and the primers used) is fundamental to interpret any existing molecular data.

Concluding remarks

To our knowledge, this is the first study investigating the correspondence between HTS and cell counts for selected and relevant taxa of marine picoeukaryotes. Indeed, true cell abundances of picoeukaryotic taxa require the TSA-FISH approach, but, as this approach has inherent limitations (it is time consuming, few probes are available, and fine resolution cannot be provided), we see the need to pursue HTS studies. Our results indicate a good correlation between the two methods, implying that more cells result in more sequences, although they give different quantitative information, i.e., the relative read abundance cannot be directly related to relative cell abundance. The cDNA-V4 survey showed the best agreement with TSA-FISH abundance, providing 1:1 relationships in half of the assayed taxa, but the cDNA-V9 was best for other taxa. Thus, the targeted region of the 18S rDNA gene clearly affected the relative abundance of specific taxa. Finally, based on the data mentioned here, we suggest that the sequencing platform used (454 or Illumina) does not produce major biases in diversity. In conclusion, the most quantitative option is to use cDNA templates rather than DNA, while the choice of the targeted region will result in different relative abundances in each particular taxon.

ACKNOWLEDGMENTS

This work was supported by the European project BioMarKs (2008-6530, ERA-net Biodiversa, EU), DEVOTES (FP7-ENV-2012, 308392, EU), MICRO-3B (FP7-OCEAN-2011 287589, EU) and MEFISTO (CTM2013-43767-P, MINECO). We thank Marta Vila for her help on the TSA-FISH counts.

REFERENCES

Alonso-Sáez L, Balagué V, Sà EL, Sánchez O, González JM, Pinhassi J, Massana R, Pernthaler J, Pedrós-Alió C, Gasol JM. 2007. Seasonality in bacterial diversity in north-west Mediterranean coastal waters: Assessment through clone libraries, fingerprinting and FISH. FEMS Microbiol. Ecol. 60:98–112.

Amaral-Zettler LA, Gómez F, Zettler E, Keenan BG, Amils R, Sogin ML. 2002. Microbiology: eukaryotic diversity in Spain's River of Fire. Nature 417:137.

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS One 4:1–9.

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D. 2012. Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. ISME J. 7:244–255.

Berney C, Fahrni J, Pawlowski J. 2004. How many novel eukaryotic "kingdoms"? Pitfalls and limitations of environmental DNA surveys. BMC Biol. 2:13.

Bik HM, Sung W, De Ley P, Baldwin JG, Sharma J, Rocha-Olivares A, Thomas WK. 2012. Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. Mol. Ecol. 21:1048–1059.

Boenigk J, Arndt H. 2002. Bacterivory by heterotrophic flagellates. Antonie Van Leeuwenhoek 81:465–480.

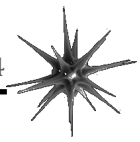
Chambouvet A, Morin P, Marie D, Guillou L. 2008. Control of toxic marine dinoflagellate blooms by serial parasitic killers. Science 322:1254–1257.

Danovaro R, Corinaldesi C, Dell'Anno A, Fabiano M, Corselli C. 2005. Viruses, prokaryotes and DNA in the sediments of a deep-hypersaline anoxic basin (DHAB) of the Mediterranean Sea. Environ. Microbiol. 7(4):586–592.

Dawson SC, Pace NR. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. Proc. Natl. Acad. Sci. U.S.A 99:8324–8329.

Decelle J, Romac S, Sasaki E, Not F, Mahé F. 2014. Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. PLoS One 9: e104297.

Del Campo J, Not F, Forn I, Sieracki ME, Massana R. 2013. Taming the smallest predators of the oceans. ISME J 7:351-358



Delong EF, Wickham GS, Pace NR. 1989. Phylogenetic stains: Ribosomal RNA-based probes for the idenfication of single cells. Science 243:1360–1363.

Dunthorn M, Klier J, Bunge J, Stoeck T. 2012. Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. J. Eukaryot. Microbiol. 59:185–7.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200.

Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B. 2013. 454 Pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: A test for marine Haptophytes. PLoS One 8: e74371.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R. 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41:597–604.

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. 10:3349–3365.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF, Knight R, Birren BW. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21:494–504.

Hong S, Bunge J, Leslin C, Jeon S, Epstein S. 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J. 3: 1365-1373.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. 12:118–123.

Lim EL, Dennett 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.

Lin YC, Campbell T, Chung CC, Gong GC, Chiang KP, Worden AZ. 2012. Distribution patterns and phylogeny of marine stramenopiles in the North Pacific Ocean. Appl. Environ. Microbiol. 78:3387–3399.

Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Gobet A, Kooistra WHCF, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Stoeck T, Santini S, Siano R, Wincker P, Zingone A, Richards TA, de Vargas C, Massana R. 2014. Patterns of rare and abundant marine microbial eukaryotes. Curr. Biol. 24:813–821.

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

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

Marie D, Partensky F, Simon N, Guillou L, Vaulot D. 2000. Flow cytometry analysis of marine picoplankton. In Living Colors: Protocols in Flow Cytometry and Cell Sorting. Diamond, R.A. and DeMaggio, S. (eds). Springer Verlag, Berlin. pp. 421–454.

Marie D, Partensky F, Vaulot D, Brussaard C. 1999. Enumeration of phytoplankton, bacteria, and viruses in marine samples. Current Protocols in Cytometry. John Wiley & Sons, Inc.

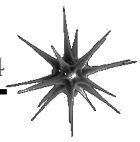
Massana R. 2011. Eukaryotic picoplankton in surface oceans. Annu. Rev. Microbiol. 65:91–110.

Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R, Claverie JM, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Forn I, Forster D, Guillou L, Jaillon O, Kooistra W, Logares R, Mahé F, Not F, Ogata H, Pawlowski J, Ernice MC, Probert I, Romac S, Richards T, Santini S, Shalchian-Tabrizi K, Siano R, Simon N, Stoeck T, Valuot D, Zingone A, de Vargas C. 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ. Microbiol. 17:4035–4049.

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

Massana R, Pedrós-Alió C. 2008. Unveiling new microbial eukaryotes in the surface ocean. Curr. Opin. Microbiol. 11:213–8.

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



Medinger R, Nolte V, Pandey RV, Jost S, Ottenwälder B, Schlötterer C, Boenigk J. 2010. Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol. Ecol. 19:32–40.

Moon-van der Staay SY, De Wachter R, Vaulot D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409:607–10.

Nielsen KM, Johnsen PJ, Bensasson D, Daffonchio D. 2007. Release and persistence of extracellular DNA in the environment. Environ. Biosafety Res. 6:37–53.

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.

Not F, Gausling R, Azam F, Heidelberg JF, Worden AZ. 2007. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. Environ. Microbiol. 9:1233–52.

Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N. 2004. A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. Appl. Environ. Microbiol. 70:4064–4072.

Not F, Simon N, Biegala IC, Vaulot D.. 2002. Application of fluorescent in situ hybridization coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton composition. Aquat. Microb. Ecol. 28:157–166.

Pernice MC, Forn I, Gomes A, Lara E, Alonso-Sáez L, Arrieta JM, del Carmen Garcia F, Hernando-Morales V, MacKenzie R, Mestre M, Sintes E, Teira E, Valencia J, Varela MM, Vaqué D, Duarte CM, Gasol JM, Massana R. 2015. Global abundance of planktonic heterotrophic protists in the deep ocean. ISME J. 9:782–792.

Pernice MC, Logares R, Guillou L, Massana R. 2013. General patterns of diversity in major marine microeukaryote lineages. PLoS One 8: e57170.

Porter KG, Feig YS. 1980. The use of DAPI for identifying aquatic microfloral. Limnol. Oceanogr. 25:943–948.

Potter D, Lajeunesse T. 1997. Convergent evolution masks extensive biodiversity among marine coccoid picoplankton. Biodiv. Conserv. 107:99–108.

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

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic. Acids. Res. 41:590–596.

Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat. Methods 6:639–641.

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

Santoferrara LF, Grattepanche JD, Katz LA, McManus GB. 2014. Pyrosequencing for assessing diversity of eukaryotic microbes: Analysis of data on marine planktonic ciliates and comparison with traditional methods. Environ. Microbiol. 16: 2752-2763

Savin MC, Martin JL, LeGresley M, Giewat M, Rooney-Varga J. 2004. Plankton diversity in the bay of fundy as measured by morphological and molecular methods. Microb. Ecol. 48:51–65.

Sherr EB, Sherr BF. 2002. Significance of predation by protists in aquatic microbial food webs. Antonie Van Leeuwenhoek 81:293–308.7.

Siano R, Alves-de-Souza C, Foulon E, Bendif EM, Simon N, Guillou L, Not F.. 2010. Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. Biogeosciences 8:267–278.

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

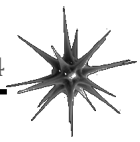
Stoeck T, Breiner HW, Filker S, Ostermaier V, Kammerlander B, Sonntag B. 2014. A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Environ. Microbiol. 16:430–444.

Stoeck T, Epstein S. 2003. Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. Appl. Environ. Microbiol. 69:2657–2663.

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

Suzuki M, Rappé MS, Giovannoni SJ. 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl. Environ. Microbiol. 64:4522–4529.

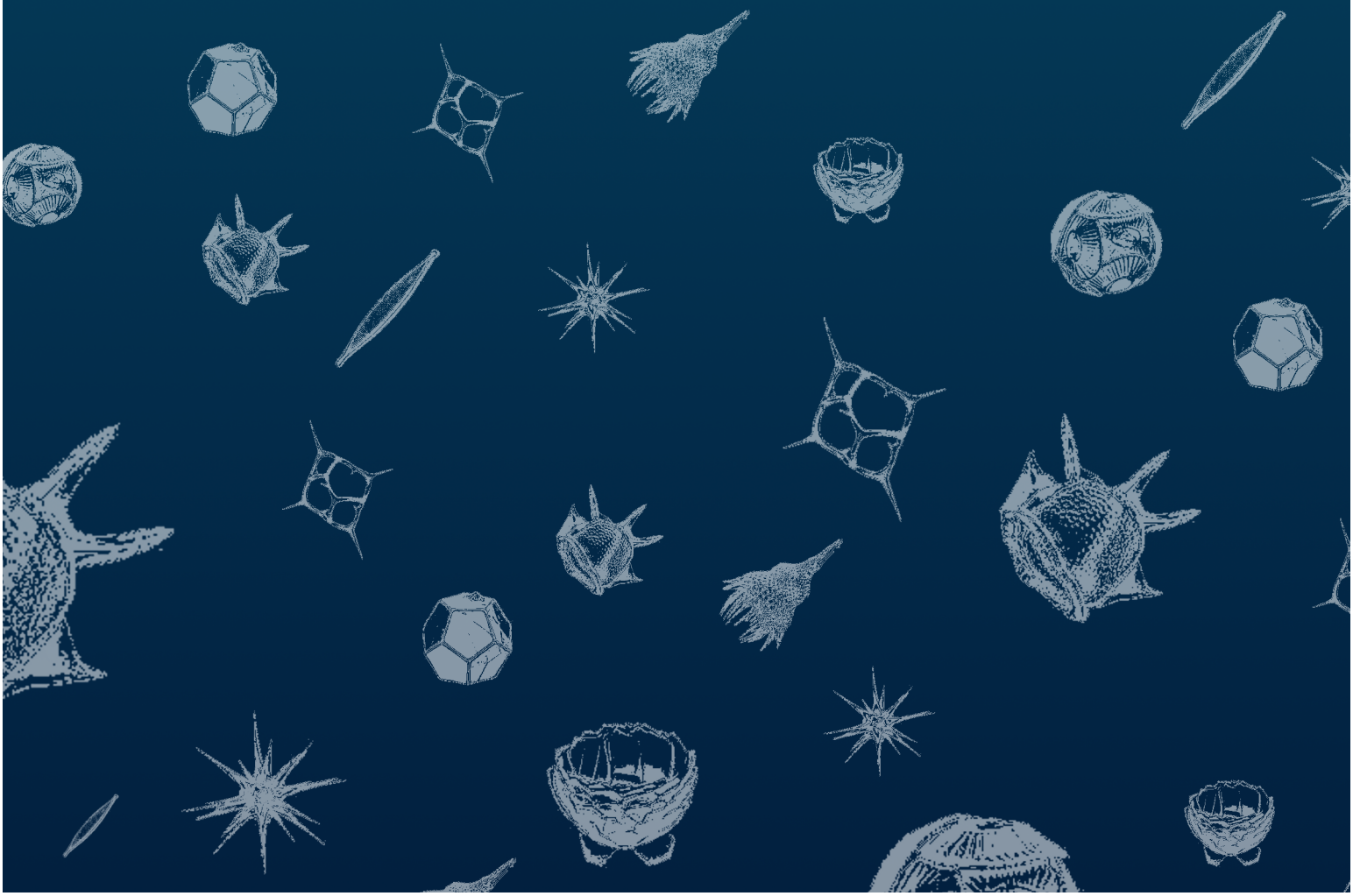
Weber AAT, Pawlowski J. 2013. Can abundance of Protists be inferred from sequence data: A case study of Foraminifera. PLoS One 8:1–8.



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.

General discussion



SYNTHESIS OF RESULTS AND GENERAL DISCUSSION

This thesis is framed within the broad context of investigating the changes in protist community composition in the marine environment. It studies the factors that drive community structure at different dimensions, with the aim of increasing our knowledge on the multiple factors that have an effect on the diversity, distribution and response of marine protists communities. With that purpose in mind, we carried out four studies, three environmental and one experimental, covering different analytical levels: (i) temporal, (ii) spatial, (iii) chemotactic, and (iv) methodological. We have used several methodological approaches to assess community composition, mainly metabarcoding and epifluorescence microscopy (FISH and DAPI techniques).

Changes in community composition through time were analyzed in a concrete area, the north-western Mediterranean, along 10 years (chapter 1) focusing in their seasonal variation. Then, we moved to a much broader area to assess differences in community composition in different depths of the water column in samples distributed along the main oceans (chapter 2). These temporal and spatial studies used sampling methods involving large samples in relation with protists body size (i.e. several liters of water are filtered), so integrating and collapsing all possible microstructures of the microbial world. In our next study we aimed to understand at small scale how protists are influenced by their closer environment and may respond to chemical signaling (chapter 3). We further analyzed the relation between different methodologies typically used to assess the diversity and abundance of microbial assemblages (chapter 4). Altogether, the results of the 4 chapters give an overview of the main factors that drive community composition. As each chapter of this thesis already contains a discussion, in the following section the main results of the chapters will be reorganized and combined with a general discussion, focusing in some common issues. As this thesis uses mainly environmental samples, the sampling (i.e., the number and frequency of samples) is generally subjected to the limitation of funding and time. For this reason, the results obtained have to be addressed and evaluated under the scope of some possible biases.

Seasonal and spatial variability and implications in community structure

Unveiling the temporal dynamics of marine organisms could provide important information about their distributional patterns and about their capacity of adaptation and response to

changing conditions. Long-term perturbations may have large consequences for the functioning of the ecosystem, and have the risk of going unnoticed or being hardly distinguishable from natural factors. In the 10 years time-series studied (chapter 1) we have confirmed that pico- and nanoeukaryotic communities followed a recurrent annual pattern, in agreement with the pattern already found in previous bacterioplankton studies (Gilbert *et al.* 2012; Cram *et al.* 2015; Fuhrman *et al.* 2015), and we differentiated two main configurations corresponding to summer and winter periods, and two transitional states in autumn and spring. Contrary to our expectations, when pulling together the different environmental variables measured, a low percentage of the community variability was explained, only 23.2%, similarly to what has been already seen in other studies (Kim *et al.* 2014; Genitsaris *et al.* 2015). However when analyzing the individual effect of each variable to each OTU, temperature and day length correlated positively or negatively with OTUs representing most of the reads (~47%). This indicates that individual OTUs can react to different variables, showing the necessity of studying the interactions between environmental variables and individual species one by one instead of pulling all together.

As one of the main issues of temporal studies is the lack of an existent methodology to quantify the seasonality present in the community, we moved a step forward and developped a simple index to recognize when a taxa follows a seasonal pattern. This allowed us to quantify the percentage of the assemblage that showed seasonality. Furthermore, we could also characterize the different behaviors present in the community, including conditionally rare and permanently rare taxa.

The sampling strategy performed consisted in one sample per month, as this has been pointed to be a convenient frequency at the Blanes Bay Microbial Observatory (BBMO). Thus, it has been observed that the forces that shape bacterial community structure operate at monthly scales (Schauer *et al.* 2003). However, in a temporal series the frequency of the sampling has a key importance in order to get the right temporal dynamics of the components of the system. We have to take into account that annual series of monthly data are adequate to explore the effects of physical and environmental phenomena that have a clear seasonal trend, as seems to be the case of the BBMO, which has a mediterranean climate and therefore follows the recurrent annual pattern determined by the seasons. In these circumstances, the shape of the function that describes the behavior of the community pattern follows a sinusoidal function. Because of that, the different theorems that apply to those functions (e.g. the waves theorems) could be applied to our community dynamics, as for instance the '*Nyquist-Shannon sampling theorem*'. To guarantee that we are not missing any signal due to under-sampling, we have to make sure that we follow that theorem. Briefly it is based on that samples of two sine waves can be identical

when at least one of them is at a frequency above half the sample rate (Fig. 1), meaning that if we sampled one per month but the community is changing twice per month we will be underestimating their frequency of change and get wrong conclusions.

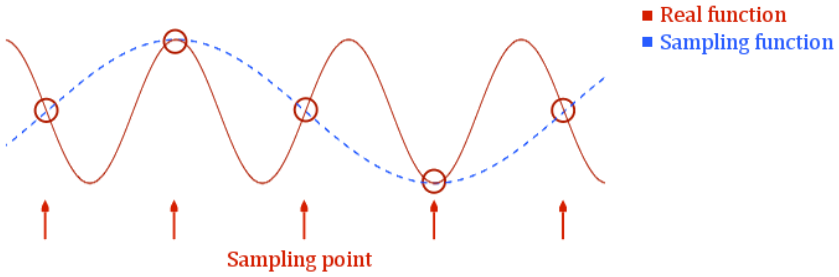


Fig. 1. Schematic representation of the '*Nyquist-Shannon sampling theorem*'. The red line indicate the real function whereas the blue line indicate the seasonal function obtained with the sampling at an unapropiate low frequencing, missing information.

Probably, in our case, conducting a more frequent sampling would not result in the detection of additional temporal patterns of the main community (i.e. the abundant one). Thus, Ward *et al.* (2017) performed a weekly sampling in a temperate site and they observed the same main patterns as the monthly scale. But we might be loosing some patterns of the rare community, e.g. maybe what we detected as 'permanently rare' is actually 'conditionally rare' because we could have missed the peak, which would be observed with a higher sampling frequency. Short-term variability in eukaryotes has already been observed (Mangot *et al.* 2013), and rapid shifts of rare taxa that could become abundant with changing environment, have been reported (Caron & Countway 2009). We have to consider this possibility due to the rapid growth of microeukaryotes, being capable to double every few hours to days (Nolte *et al.* 2010), which allow them to respond rapidly to even minor environmental fluctuations (Countway *et al.* 2005; Caron & Countway 2009). For this reason, studies on short-term dynamics could be critical for obtaining a better understanding of the factors that control and regulate eukaryotic populations, and it expected that the monthly scheme does not accurately capture all the fluctuations of the system. Nevertheless, usually when the system follows a regular annual cycle, the monthly sampling may miss the week with the highest chlorophyll or nutrient concentration but overall give a good picture to characterize the main variability present in the system.

Our temporal study gave a very detailed record of the protist dynamics in a localized area, and it was pertinent to move to a broader area of study to compare different communities and assess the drivers of changes in community structure (chapter 2). We used the data obtained in the Malaspina 2010 circumnavigation to assess how the protist assemblages changed along the water column in 13 vertical profiles. A clear segregation between photic (surface and DCM) and aphotic (meso- and bathypelagic) communities was observed, being community composition more similar within those layers than between them, as previously observed (Not, countway). We found that photic layer was dominated by Ciliophora, Dinoflagellata and Pelagophyceae whereas Chrysophyceae and Bicosoecida dominated the aphotic layer. Furthermore the dispersal capability appeared to be higher in the photic communities than in aphotic ones, consistent with our understanding of ocean circulation.

An important point that we have to consider when developing a spatial study is: are we conducting these studies at the correct scale? Usually the community composition of the different layers of the water column (surface, DCM, mesopelagic and bathypelagic) is the result of a single sampling point in each of the layers. In our study, for most of the stations we have 3 samples representative of the mesopelagic and bathypelagic layers (chapter 2), which is an improvement towards most other studies but that could be not enough. For instance, it has been observed that the DCM is an environmentally diverse layer with well-marked gradients occurring in a short distance, which is also influencing the proportion of the groups living in the upper or lower part of the DCM (Cabello et al., 2016). This may be even more extreme in the mesopelagic, as this layer contains both the DSL (deep scattering layer) and the minimum oxygen zone, intermedial layers with specific environmental characteristics that could drive the community composition (Robinson *et al.* 2010). In fact, this specific environment is likely explaining the high amount of unique OTUs in the mesopelagic and the higher relative activity detected for some taxonomic groups. Thus, depending on the depth sampled and the number of samples in a given layer we could get a biased picture of the community composition. And if the sample is not representative of the community, trying to understand environmental filtering may not be possible.

Single cell responses in the microscale. Which size scale matters?

This leads to the following question: what spatial scale is relevant to microbes? For the temporal (chapter 1) and spatial (chapter 2) study we sample tens of liters to study several microbial features, however, it is well known that the microstructure of the ecosystem influence microbial behavior, and everyday more techniques are allowing to study microbes at their small cell-scale

(e.g. FISH, SAGs, nanoSIMS, microfluidics) and study cells one by one. Due to their small size, motile microorganisms e.g. bacteria, are able to swim ~0.5 µl in 10 min, for this reason, they are not affected for what is happening hundreds of meters not even kilometers away from their position. However their ability to sense a chemical gradient increases their swimming velocity to 6 mm in 10 min (Stocker 2012).

Due to the importance of chemotactic behavior to detect sources of food, we aimed to understand the main preferences of the protists within a community towards different chemicals (chapter 3). Overall, bacterial exudates were the tested compounds triggering the highest chemotactic response. As protists are the main grazers of the ocean, their ability to sense bacterial exudates could explain their high bacterivory. Few previous studies had detected the ability of protists to swim towards bacteria, but none has tried this type of experiments using all the members of the community. This protist chemotactic behavior could explain some of the features observed in chapter 2. We observed that the mesopelagic was the layer with more metabolic activity. This layer is characterized by a higher concentration of nutrients that could generate different chemical gradients detected by bacteria and at the same time these hotspots of bacteria being detected by protists. Protists attached to large particles like aggregates, could be sinking from upper layers and when chemotactically detect a patch of chemical stimuli swim towards it, as it has been seen that some OTUs are shared between different depths (chapter 2).

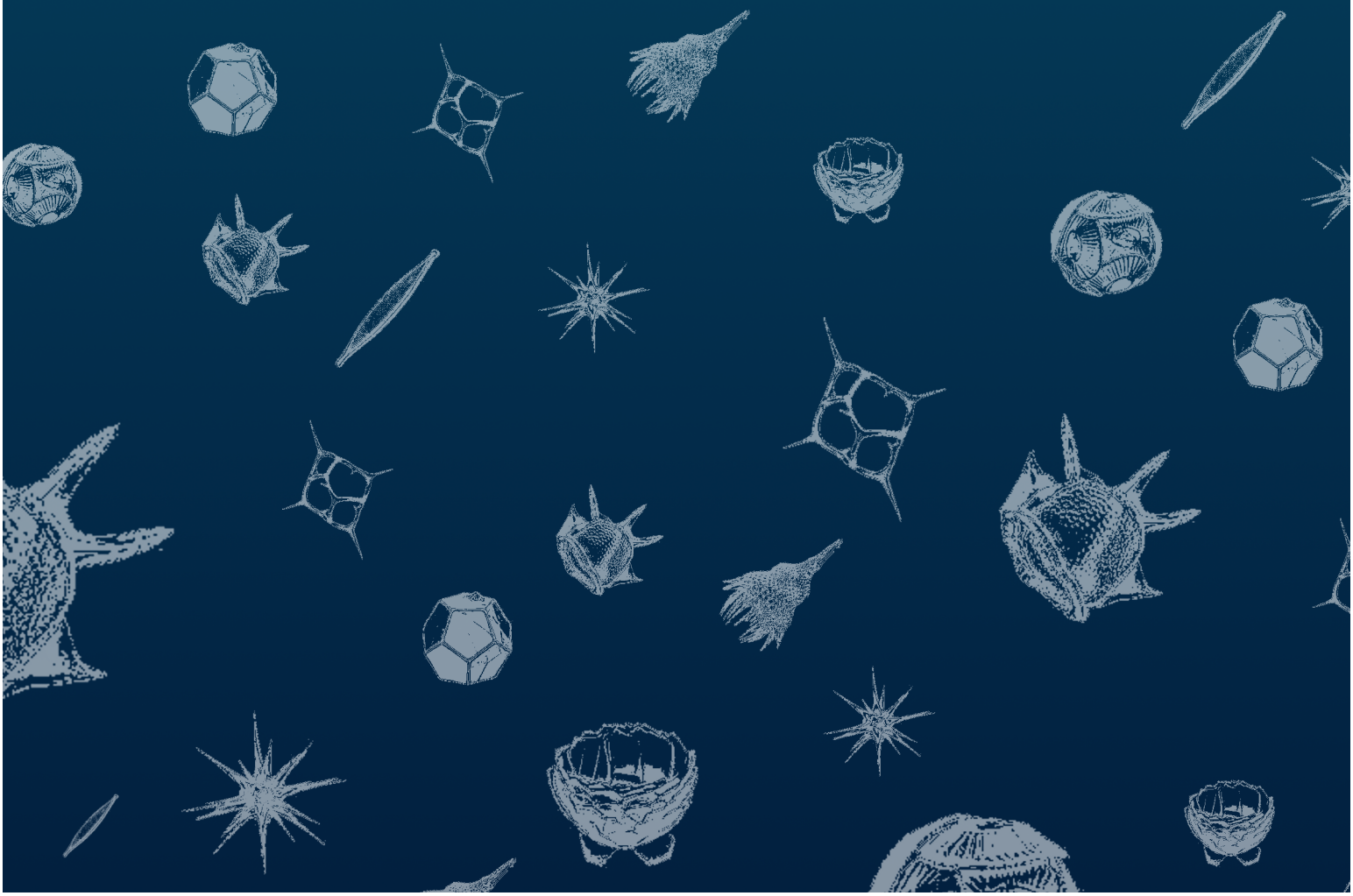
Molecular approaches. A consensus is needed

To address the community composition present in a sample, environmental 18S rDNA genes are sequenced. Due that new high-throughput sequencing technologies do not allow to sequencing the complete 18S rDNA gene, we had to focus in a given region, being the V4 or V9 regions the most frequently sequenced. Initially, the first surveys using high-throughput sequencing (HTS) used the V9 region (Amaral-Zettler *et al.* 2009; Stoeck *et al.* 2009), mainly due to the fact that the V9 is a short region (~ 180bp) and at the beginning HTS only allowed to sequence short regions. The technical advances in HTS allowed to sequence longer reads, at the V4 region (~380bp) started to be sequenced (Stoeck *et al.* 2010). However the information inferred by both regions is not exactly the same (chapter 4). We demonstrated that the relative abundance of specific groups changed depending on the region sequenced but also depending on the template used (e.g. DNA or cDNA). When comparing the relative abundance obtained by HTS with cell abundance, cDNA-V4 gives overall the best information. However we could not establish a

general rule, as some groups performed better in the V9, like the stramenopile tested taxa, whereas other taxa like *Micromonas* performed better in the V4. An advantage of the V4 region is that it has been proved to be a good descriptor of the variability of the entire 18S rRNA (Pernice *et al*., 2013). Nowadays we face the problem that some studies are using the V9 (de Vargas *et al*., 2015), whereas others are using the V4 (Massana *et al*., 2015), and this makes these studies not totally comparable, as the differences observed could be due to the primers used or to real biogeographical variability. For this reason it must be a priority for protistologists to make a consensus on which region should be used at the first term, to make all studies comparable and give the possibility to advance more in the knowledge of protist communities.

During the last years, the marine protist world has suffered a rapid evolution mainly due to the discovery of their huge hidden diversity. Nowadays it is time to go deeper in the study of single cells and interactions among different phyla to determine ecological interactions at the microscale level.

Conclusions



CONCLUSIONS

1. The analysis of long temporal series indicated that, throughout the 10 years, pico- and nanoeukaryotic communities present a recurrent annual pattern with two main states corresponding to warm (summer) and cold (winter) water temperatures. Furthermore, we evidenced two transitional states during spring and autumn months. Communities separated 12 months or their multiples (24, 36, and so on) where most similar, while communities separated 6 months and their multiples were the most dissimilar.

2. The permanently rare sub-community also showed seasonality and the same main and transitional states as the whole-community.

3. The 'Seasonality index' allowed to numerically quantify the degree of seasonality for each taxonomic group or OTU, differentiating between 'Seasonal' ('Strongly seasonal' or 'Moderately seasonal') and 'Non-seasonal' behaviors. In Blanes Bay 13.2% of the OTUs and 22.4% of the taxonomic groups were seasonal representing 39.4% and 35.2% of the abundance (reads) respectively. Among the Seasonal taxa, pulse and wave strategies were identified.

4. Picoeukaryotic communities had a clear vertical segregation with a striking community differentiation between photic and aphotic layers. Richness and dispersal was higher in the sunlit ocean and decreased with depth. Different taxonomic classes were dominant along the water column: Ciliophora and Dinoflagellata in surface waters, Pelagophyceae and Dinoflagellata at the DCM and Chrysophyceae and Biscosoecida in meso- and bathypelagic.

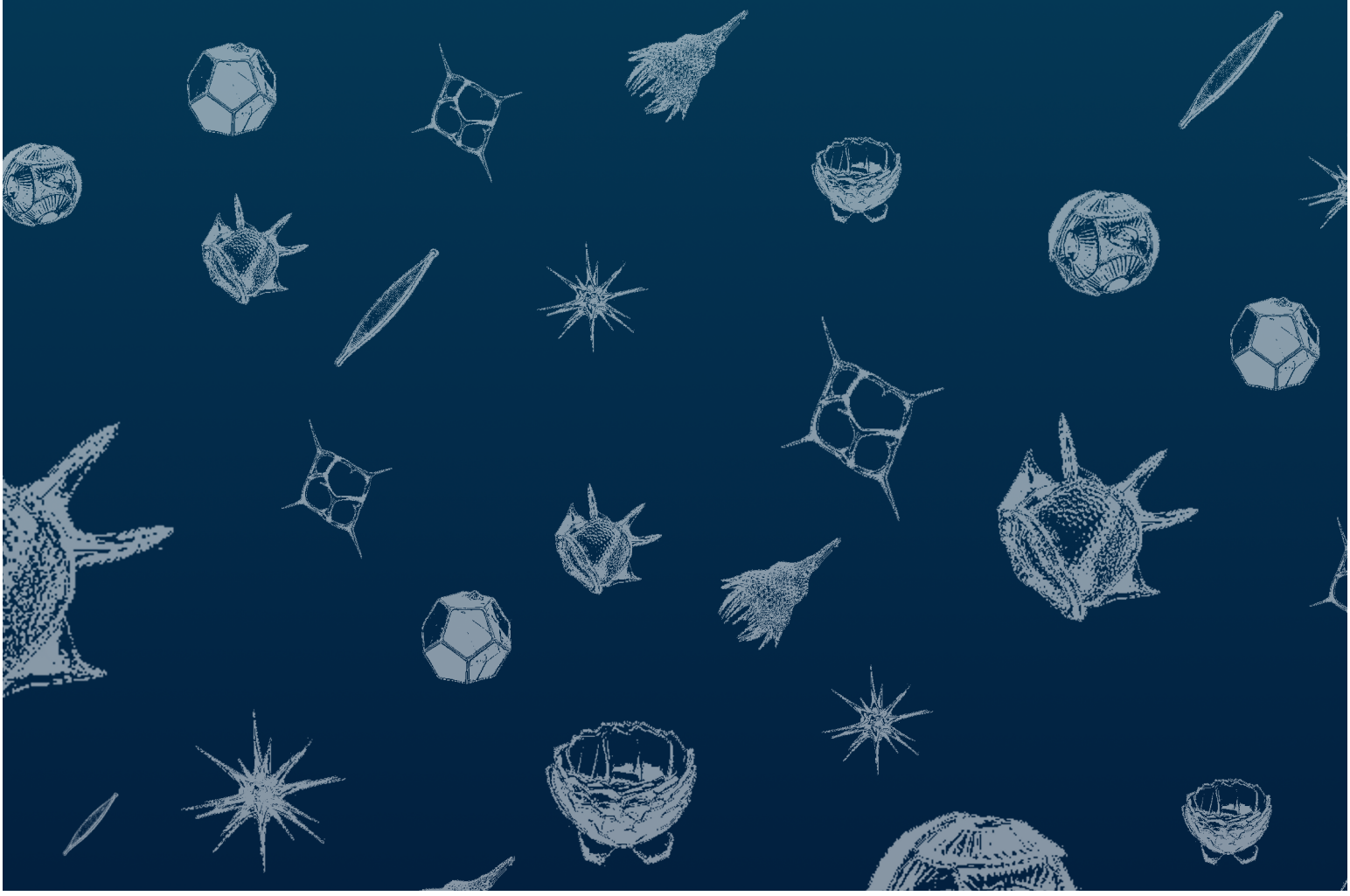
5. In general, the diversity obtained by the metabarcoding (iTags) and metagenomes (miTags) approaches showed a good correlation. But miTags detected 6 groups totally or partially absent from the metabarcoding (Prymnesiophyceae, Kinetoplastida, Diplonemea, Discosea, Ascomycota and Basidiomycota).

6. Most taxonomic groups had their highest relative metabolic activities in the mesopelagic layer, whereas the bathypelagic displayed the lowest activity for the majority of groups.

7. Protists showed stronger chemotaxis response to bacterial exudates (especially from Roseobacter and Flavobacteria) than to other chemical cues such as sugars o ammonia. Specifically, phototrophic cells showed a higher chemotactic index than heterotrophic cells, and regarding size, the smaller cells (1-3 μm) were the ones with stronger response. *Roseobacter* exudates were the attractant that elucidated a higher response in Ciliates and Dinoflagellates.

8. Metabarcoding and microscopic (TSA-FISH) abundance data were generally correlated, implying that more cells in the sample resulted in more 18S rDNA sequences, but slopes often deviated from 1, giving different quantitative information. Usually, the molecular signal obtained using cDNA was closer to relative cell abundances. The targeted region of the 18S rDNA, V4 or V9, clearly affected the relative abundances of taxa, but each region provided better estimates depending on the taxa analyzed.

General references



REFERENCES *(Includes Introduction and General discussion)*

Adl, S.M., Simpson, A.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R. *et al.* (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *The Journal of eukaryotic microbiology*, 52, 399-451.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PloS one*, 4, e6372.

Alonso-Saez, L., Diaz-Perez, L. & Moran, X.A. (2015). The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environmental microbiology*, 17, 3766-3780.

Arístegui, J., Gasol, J.M., Duarte, C.M. & Herndl, G.J. (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnology and Oceanography*, 54, 1501-1529.

Azam, F., Fenchel, T., FIeld, J.G., Gray, J.S., Meyerreil, L.A. & Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, 10.

Azam, F. & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature reviews. Microbiology*, 5, 782-791.

Baldauf, S.L. (2003). The deep roots of eukaryotes. *Science*, 300, 1703-1706.

Baldauf, S.L. (2008). An overview of the phylogeny and diversity of eukaryotes. *J. Syst. Evol.*, 46, 263-273.

Berney, C., Fahrni, J. & Pawlowski, J. (2004). How many novel eukaryotic 'kingdoms'?: pitfalls and limitations of environmental DNA surveys. *BMC Biol*, 2, 13.

Bunse, C. & Pinhassi, J. (2017). Marine Bacterioplankton Seasonal Succession Dynamics. *Trends in microbiology*, 25, 494-505.

Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor perspectives in biology*, 6, a016147.

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S.I., Jakobsen, K.S. *et al.* (2007). Phylogenomics reshuffles the eukaryotic supergroups. *PloS one*, 2, e790.

Campbell, B.J., Yu, L., Heidelberg, J.F. & Kirchman, D.L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 12776-12781.

Caron, D.A. & Countway, P.D. (2009). Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquatic Microbial Ecology*, 57, 227-238.

Caron, D.A., Alexander, H., Allen, A.E., Archibald, J.M., Armbrust, E.V., Bachy, C. *et al.* (2017). Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature reviews. Microbiology*, 15, 6-20.

Caron, D.A., Worden, A.Z., Countway, P.D., Demir, E. & Heidelberg, K.B. (2009). Protists are microbes too: a perspective. *The ISME journal*, 3, 4-12.

Chrétiennot-Dinet, M.J., Courties, C., Vaquer, A., Neveux, J., Claustre, H., Lautier, J. *et al.* (1995). A new marine picoeucaryote: Ostreococcus tauri gen. et sp. nov. (Chlorophyta, Prasinophyceae). *Phycologia*, 34, 285-292.

Countway, P.D., Gast, R.J., Savai, P. & Caron, D.A. (2005). Protistan diversity estimates based on 18S rDNA from sea-water incubations in the Western North Atlantic. *The Journal of eukaryotic microbiology*, 52, 95-106.

Cram, J.A., Chow, C.E., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A. *et al.* (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *The ISME journal*, 9, 563-580.

Cuvelier, M.L., Ortiz, A., Kim, E., Moehlig, H., Richardson, D.E., Heidelberg, J.F. *et al.* (2008). Widespread distribution of a unique marine protistan lineage. *Environmental microbiology*, 10, 1621-1634.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R. *et al.* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348.

del Campo, J., Balague, V., Forn, I., Lekunberri, I. & Massana, R. (2013). Culturing bias in marine heterotrophic flagellates analyzed through seawater enrichment incubations. *Microbial ecology*, 66, 489-499.

del Campo, J. & Massana, R. (2011). Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist*, 162, 435-448.

Diez, B., Pedrós-Alió, C., Marsh, T.L. & Massana, R. (2001a). 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.

Diez, B., Pedrós-Alió, C. & Massana, R. (2001b). 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.

Duarte, C.M. (2015). Seafaring in the 21St Century: The Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin*, 24, 11-14.

Falkowski, P. (2012). Ocean Science: The power of plankton. *Nature*, 483, S17-S20.

Falkowski, P.G., Fenchel, T. & Delong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320, 1034-1039.

Fenchel, T. & Blackburn, N. (1999). Motile chemosensory behavior of phagotrophic Protists: Mechanisms for and efficiency in congregating at food patches. *Protist*, 150, 325-336.

Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P.G. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *science*, 281, 237-240.

Fuhrman, J.A. (2009). Microbial community structure and its functional implications. *Nature*, 459, 193-199.

Fuhrman, J.A., Cram, J.A. & Needham, D.M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature reviews. Microbiology*, 13, 133-146.

Genitsaris, S., Monchy, S., Viscogliosi, E., Sime-Ngando, T., Ferreira, S. & Christaki, U. (2015). Seasonal variations of marine protist community structure based on taxon-specific traits using the eastern English Channel as a model coastal system. *FEMS microbiology ecology*, 91.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P. *et al.* (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PloS one*, 3, e3042.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbruck, L., Reeder, J., Temperton, B. *et al.* (2012). Defining seasonal marine microbial community dynamics. *The ISME journal*, 6, 298-308.

Giovannoni, S.J., Britschgi, T.B. & Moyer, C.L. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345, 60-63.

Johnson, P.W. & Sieburth, J.M. (1982). In situ morphology and occurrence of eukcayotic phototrophs of bacterial size in the picoplankton of estuarine and oceanic waters. *J. Phycol*, 18, 318-327.

Jones, R.I. (2000). Mixotrophy in planktonic protists: An overview. *Freshwater biol.*, 45, 219-226.

Jürgens, K. & Massana, R. (2008). Protist grazing on marine bacterioplankton. In: *Microbial Ecology of the Oceans.* 2n ed.(editor D.L. Kirchman).

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

Karl, D.M. (2007). Microbial oceanography: paradigms, processes and promise. *Nature reviews. Microbiology*, 5, 759-769.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., de Vargas, C., Raes, J. *et al.* (2011). A holistic approach to marine eco-systems biology. *PLoS Biol*, 9, e1001177.

Kim, D.Y., Countway, P.D., Jones, A.C., Schnetzer, A., Yamashita, W., Tung, C. *et al.* (2014). Monthly to interannual variability of microbial eukaryote assemblages at four depths in the eastern North Pacific. *The ISME journal*, 8, 515-530.

Krabberød, A.K., Bjorbækmo, M.F.M., Shalchian-Tabrizi, K. & Logares, R. (2017). Exploring the oceanic microeukaryotic interactome with metaomics approaches. *Aquatic Microbial Ecology*, 79, 1-12.

Li, W.K.W. (1994). Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnology and Oceanography*, 39, 169-175.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R. *et al.* (2014). Patterns of rare and abundant marine microbial eukaryotes. *Current biology : CB*, 24, 813-821.

Logares, R., Mangot, J.F. & Massana, R. (2015). Rarity in aquatic microbes: placing protists on the map. *Research in microbiology*, 166, 831-841.

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

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

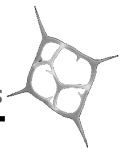
Lovejoy, C., Vincent, W.F., Bonilla, S., Roy, S., Martineau, M.-J., Terrado, R. et al. (2007). Distribution, Phylogeny, and Growth of Cold-Adapted Picoprasinophytes in Arctic Seas. Journal of Phycology, 43, 78-89.

Mangot, J.F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G. *et al.* (2013). Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environmental microbiology*, 15, 1745-1758.

Massana, R. (2011). Eukaryotic picoplankton in surface oceans. Annual review of microbiology, 65, 91-110.

Massana, R. (2015). Protistan diversity in environmental molecular surveys. S. Ohtsuka et al. (eds.), Marine Protists. Springer Japan 2015.

Massana, R., Balague, V., Guillou, L. & Pedros-Alio, C. (2004a). Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. FEMS microbiology ecology, 50, 231-243.



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

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C. *et al.* (2015). Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental microbiology*, 17, 4035-4049.

Massana, R., Guillou, L., Diez, B. & Pedros-Alio, C. (2002). Unveiling the Organisms behind Novel Eukaryotic Ribosomal DNA Sequences from the Ocean. *Applied and Environmental Microbiology*, 68, 4554-4558.

Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwalder, B., Schlotterer, C. *et al.* (2010). Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular ecology*, 19 Suppl 1, 32-40.

Montagnes, D.J.S., Barbosa, A.B., Boenigk, J., Davidson, K., Jürgens, K., Macek, M. *et al.* (2008). Selective feeding behaviour of key free-living protists: avenues for continued study. *Aquatic Microbial Ecology*, 53, 83-98.

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

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

Not, F., del Campo, J., Balague, V., de Vargas, C. & Massana, R. (2009). New insights into the diversity of marine picoeukaryotes. *PloS one*, 4, e7143.

Not, F., Gausling, R., Azam, F., Heidelberg, J.F. & Worden, A.Z. (2007). Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environmental microbiology*, 9, 1233-1252.

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

Partensky, F., Blanchot, J., Lantoine, F., Neveux, J. & Marie, D. (1996). Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. *Deep Sea Research I*, 43.

Pedros-Alio, C. (2012). The rare bacterial biosphere. *Annual review of marine science*, 4, 449-466.

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

Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M. *et al.* (2016). Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *The ISME journal*, 10, 945-958.

Pernice, M.C., Logares, R., Guillou, L. & Massana, R. (2013). General patterns of diversity in major marine microeukaryote lineages. *PloS one*, 8, e57170.

Pomeroy, L.R. (1974). Oceans Food Web, a changing paradigm. *Bioscience*, 24.

Porter & Feig (1980). The use of DAPI for identifying and counting aquatic microflora. *Limnology and Oceanography*.

Robinson, C., Steinberg, D.K., Anderson, T.R., Arístegui, J., Carlson, C.A., Frost, J.R. *et al.* (2010). Mesopelagic zone ecology and biogeochemistry – a synthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 1504-1518.

Rodriguez-Martinez, R., Labrenz, M., del Campo, J., Forn, I., Jurgens, K. & Massana, R. (2009). Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environmental microbiology*, 11, 397-408.

Salazar, G., Cornejo-Castillo, F.M., Benitez-Barrios, V., Fraile-Nuez, E., Alvarez-Salgado, X.A., Duarte, C.M. *et al.* (2016). Global diversity and biogeography of deep-sea pelagic prokaryotes. *The ISME journal*, 10, 596-608.

Sanders, R.W. (1991). Mixotrophic protists in marine and freshwater ecosystems. *J. Protozool.*, 28, 76-81.

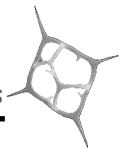
Sanders, R.W., Berninger, U.G., Lim, E.L., Kemp, P.F. & Caron, D.A. (2000). Heterotrophic and mixotrophic nanoplankton predation on picoplankton in the Sargasso Sea and on Georges Bank. *Marine Ecology Progress Series*, 192, 103-118.

Schauer, M., Balague, V., Pedros-Alio, C. & Massana, R. (2003). Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquatic Microbial Ecology*, 21.

Scholin, C.A., Buck, K.R., Britschgi, T., Cangelosi, G. & Chavez, F.P. (1996). Identification of Pseudo-nitzschia australis (Bacillariophyceae) using rRNA-targeted probes in whole cell and sandwich hybridization formats. *Phycologia*, 35, 190-197.

Scholz, M.B., Lo, C.C. & Chain, P.S. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current opinion in biotechnology*, 23, 9-15.

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



Sherr, E.B. & Sherr, B.F. (2000). *Marine microbes: an overview.* . Wiley: New York, In: Kirchman DL. Microbial ecology of the oceans.

Simpson, A.G. & Roger, A.J. (2004). The real 'kingdoms' of eukaryotes. *Current biology : CB*, 14, R693-696.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R. *et al.* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, 103, 12115-12120.

Stocker, R. (2012). Marine microbes see a sea of gradients. *Science*, 338, 628-633.

Stocker, R. & Seymour, J.R. (2012). Ecology and physics of bacterial chemotaxis in the ocean. *Microbiology and molecular biology reviews : MMBR*, 76, 792-812.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W. *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular ecology*, 19 Suppl 1, 21-31.

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M.J., Chistoserdov, A. *et al.* (2009). Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol*, 7, 72.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G. *et al.* (2015). Structure and function of the global ocean microbiome. *Science*, 6237.

Suttle, C.A. (2007). Marine viruses--major players in the global ecosystem. *Nature reviews. Microbiology*, 5, 801-812.

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

Ward, C.S., Yung, C.M., Davis, K.M., Blinebry, S.K., Williams, T.C., Johnson, Z.I. *et al.* (2017). Annual community patterns are driven by seasonal switching between closely related marine bacteria. *The ISME journal*, 11, 1412-1422.

Whitman, W.B., Coleman, D.C. & Wiebe, W.J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 6578-6583.

Wintzingerode, F., Göbel, U.B. & Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, 21, 213-229.

Woese, C.R. (1987). Bacterial evolution. *Microbiol Rev*, 51, 221-271.

Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 5088-5090.

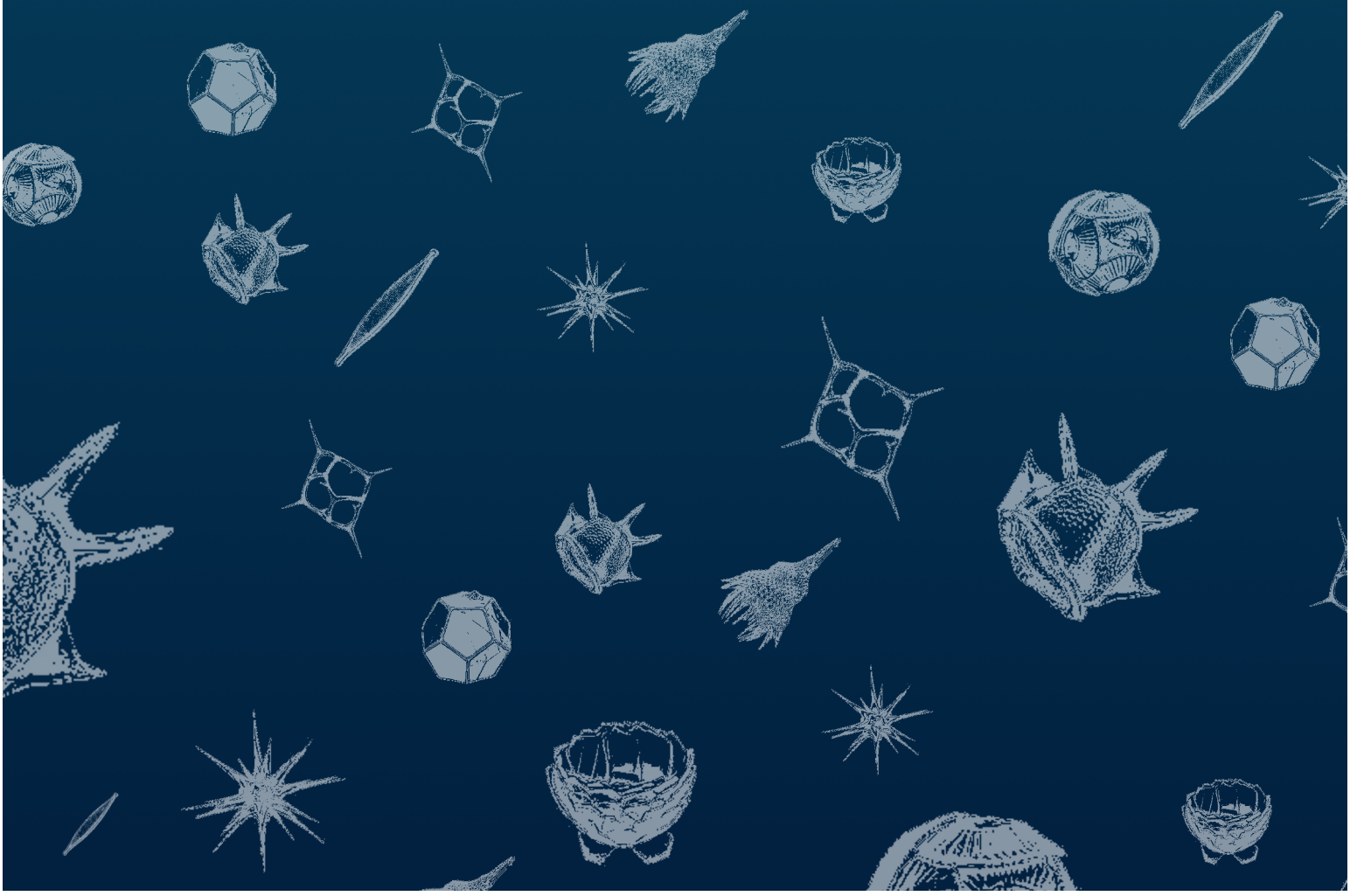
Worden, A.Z. (2006). Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecology*, 43, 165-175.

Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E. & Keeling, P.J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, 347(6223).

Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS microbiology ecology*, 52, 79-92.

Zubkov, M.V. & Tarran, G.A. (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature*, 455, 224-226.

Acknowledgements



Sembla mentida, però pràcticament sense adonar-me'n ha arribat el moment d'escriure els agraïments!! Durant aquests 5 anys he pensat diverses vegades en aquest moment i ara que ha arribat l'hora no sé per on començar! (quan t'hi has de posar és més complicat del que sembla!). Un es para a pensar en tota la gent que d'una manera o altra ha format part d'aquesta etapa i a la que li vol agrair, i els noms es comencen a atropellar un darrera l'altre. Han sigut 5 anys carregats de MOLT bons moments, que m'han permès conèixer gent meravellosa, viure experiències i viatjar per llocs que mai hagués pensat que trepitjaria!! Per aquest motiu espero no deixar-me a ningú.

Primer de tot està clar que res d'això no hagués sigut possible sense els meus directors de tesi! Ramon! Moltíssimes gràcies per confiar en mi (una noia que no venia ni del món marí ni del microbiològic) per fer una tesi en microbiologia marina (al final ha anat prou bé, oi?). Moltes gràcies per la dedicació i l'esforç, i per tenir sempre la porta del despatx oberta per una pregunta. Encara recordo com de perduda anava les primeres setmanes quan tothom parlava de CTDs i campanyes com si fossin el pa de cada dia… Moltes gràcies per transmetre'm el teu coneixement i entusiasme-me pels petits eucariotes, i ensenyar-me a ser encara més meticulosa del que era! Gràcies per donar-me sempre la teva opinió però també deixar-me fer les coses a la meva manera quan tenia les idees clares, i sobretot MOLTES gràcies per aquesta última etapa! Semblava complicat però ho hem aconseguit!

I d'un director a l'altre! Ramiro, muchísimas gracias!!! Gracias por descubrirme el mundo de la bioinformática y de Linux (aún recuerdo el dolor de cabeza de los primeros días! y creo que tú todavía te ríes recordando mi cara). Gracias por tu paciencia infinita cuando volvía con algún error, por compartir tus 'truquishos' (seguro que tienes algunos nuevos), por las charlas ecológicas, por los lunch-meetings, y gracias en esta última etapa por tu apoyo y ayuda (a pesar de que se ha juntado todo!). Pero no todo ha sido trabajo! Gracias también por los Friday-beers, y por ser el mejor embajador del Claror! (¡¡Viva el spinning!!).

Però segurament no hauria arribat fins aquí, si al 2010 Miquel Àngel Arnedo no m'hagués obert les portes del seu laboratori. Moltes gràcies Kele per donar-me la primera oportunitat i ensenyar-me el que és la vida al laboratori. Indirectament una part d'aquesta tesi també és gràcies a tu.

Gràcies als companys de despatx!!! Gràcies per fer que fos el millor despatx on podria haver estat! Francisco, qué te voy a contar que no sepas?! Terminaste hace un año y no sabes como se ha notado tu ausencia. Muuuuuchas muuuuuchas gracias por todos los momentos compartidos dentro y fuera del ICM, por apoyarme y escucharme siempre, por nuestras conversaciones, por los regalitos que me encontraba en la mesa de buena mañana, en definitiva gracias por hacer esta tesis más fácil!!. Estela, he aprendido mucho de ti! gracias por todos lo momentos y conversaciones compartidas, de mayor quiero ser como tu! Maria, gràcies per ser tan alegre i portar aquesta alegria al despatx! i també per l'ajuda estadística (ets una crack!!) Teresa, la italiana del despacho, gracias por estar siempre predispuesta a ayudar, conversar, o lo que hiciera falta! Paula, gràcies per cuidar-nos a tots i ajudar-nos sempre en tot el que pots, Néstor, aunque llegaste hace un año ya eres miembro 100% del despacho, ha sido genial compartir esta última fase juntos! (y gracias por tu ayuda en R!), Elena gracias por el apoyo en esta última etapa, ánimo que ya lo tienes!. En definitiva, gràcies a tots per fer del P39 un lloc especial!!

Però a part de la gent amb la que convius moltes hores dins del teu propi despatx, en altres despatxos hi ha gent imprescindible sense la qual aquesta tesi no seria el mateix! Mireia, moltíssimes gràcies per les mil hores compartides i no només a l'ICM sinó a salsa, al gimnàs, tornant cap a casa en bici, caminant,...puff em venen un munt de records! Ja saps que aquesta tesi no hagués sigut possible sense el teu suport en molts moments (i mira que quasi no ha sigut possible gràcies a Texas, eh?). Una setmana més i tu també ja ho tens!!. Isabel (tronca!) la otra futura-doctora de esta promoción! Realmente terminar a la vez ha hecho el proceso mucho más ligero! Muchas gracias por todos los momentos compartidos dentro y fuera del ICM (bbq, scape-rooms, excursiones,...). Somos las promotoras de salir a comer fuera!!. Sdena (mi brasileña favorita), muchas gracias por estar siempre ahí, por preocuparte y escucharme, y saber que puedo contar contigo para cualquier cosa!

A tota la gent que volta per l'ICM amb els quals he compartit bons moments, i que fan que anar a l'ICM sigui una alegria diària i no hi vulguis marxar: Dorleta, Yaiza, Idaira, Mariri y Marta muchas gracias por aguantarme en esta última etapa! (qué paciencia habéis tenido aguantando a 3 doctorandas que estaban a punto de terminar!). Francisco M. (que gran artistaaaaaaa eres!! Un conciertillo de celebración de tesis no estaría mal, no?), Clara Ruiz (otra artistaza!!), Eli A., Laia, Marta R., Pau, Lucia, Marina, Pablo R., Rachele, Albert R., Adrià, I a dues grans post-docs del departament Isabel F. i Marta S., gràcies per ajudar-me sempre amb qualsevol dubte que he tingut!! i també a molta gent que ja ha marxat: AnaMari (gracias por tu alegría y salero!), Massimo (nunca olvidaré nuestras millones de horas en el lab! gracias por preocuparte por mi!),

Guillem, Carmen, Elisa, Maria de la F., Juancho, Daffne, Rosana, Elena L., Edgar, Sergi, Sarah-Jane, Suso, Pedro, Raquel Rodríguez, Bea F., Roy, Rodrigo, Eva O., Bibiana, me guardo un pedacito de todos y cada uno de vosotros! Gracias!

Bona part de la tesi, i en general, el bon funcionament del departament, és i ha sigut possible gràcies a gent que sempre està disposada a ajudar-te amb un somriure als llavis. Irene, una part d'aquesta tesi es teva! gràcies per totes aquelles hores, dies, setmanes i mesos contant mano a mano al microscopi. Gràcies per la teva bona predisposició i ajuda en tot moment! Vane, a tu també t'he d'agrair bona part d'aquesta tesi. Gràcies per fer que el lab molecular funcioni, i per les hores i hores extraient DNA!. Qui ens anava a dir fa anys que ara seriem tan 'pros' de Linux? Pablo Sánchez, mil gràcies per solucionar els milions de dubtes i problemes bioinformàtics que tenim! però sobretot per no cansar-te mai d'explicar-nos el mateix tots els cops que ens faci falta (ets La Pera!). Clara C. i Eli Sà moltes gràcies a vosaltres també, sou part indispensable del bon funcionament del departament!!

I com no, moltes gràcies als 'jefazos' que ronden pel segon pis i que fan que tot el sistema funcioni: Pep, Dolors, Cèlia (gràcies per deixar-me participar a la Ficaram!), Rafel, Carles Pedrós (tot i que ja no estàs per l'ICM), Silvia A., Cesc, Gràcies per ser 'las mentes pensantes' del departament!

Però durant la tesi també he pogut veure món. I would like to thank Roman Stocker for let me go to his lab, and also to everyone in Stocker's lab that made my months there awesome! Steve, Ben, Jen and Becky, thank you! Pero estar en Boston también me permitió conocer personas maravillosas fuera del laboratorio, y me llevo amigos para siempre. Muchas gracias Mati, Alberto and Julian por hacer que la experiencia americana fuera inolvidable! and of course Ben and Joel! thanks for everything, I will always remember Thanks Giving Day with your family!

And also to the Seymour's Lab. Thank you Justin for hosting me in Sydney! Those three months at the UTS were amazing, I learned a lot! and of course thanks to Marco (thanks to share your bench!), Marco (gracias por todo los ratos en microscopía y momentos fuera del lab), Rendy, Bojana, Rachel, Lauren, Bonnie, Ric (thanks for all the banana breads!) and Caitlin, to make me feel like home!!

Per descomptat no em puc oblidar de la gent amb la qual he compartit una de les millors experiències del doctorat, les campanyes! Gracias Ficarameros-Osobuqueros por esa gran

campaña llena de momentos que serán imposibles de borrar de la memoria!! y como no también a los Aresteros por todas estas salidas al monte, tan necesarias para desconectar!!

Fora de l'àmbit acadèmic hi ha molta gent que crec que encara tenia més ganes que jo de que acabés la tesi. Gràcies, moltes gràcies als 'Ochenterus' per ser la meva via d'escapatòria no científica. Sempre es bo tenir amics amb els quals no es parli únicament de ciència. Aida, Anna, Xexi, Victor, Javi, gràcies per entendre que desaparegués durant èpoques!! I Moltes gràcies Maria per tot el suport que m'has donat ara i abans de començar la tesi. Sé que puc contar amb tu pel que sigui!! i gràcies per venir a dinar a l'ICM quan feia molt que no ens veiem! (llàstima que ja no facin les amanides!).

Brito, te mereces una mención especial! Qué decirte que no sepas?! Mil gracias por estar siempre ahí para lo que necesite a la hora que lo necesite. Gracias por tu cariño, por tu energía positiva, por animarme cuando me hacía falta y por hacerme ver las cosas desde otro punto de vista cuando solo lo veía desde un ángulo. Eres la hermana que no tengo! Si olvidar la fecha del deposito es complicado, coincidiendo con el día de tu cumpleaños será imposible!!

I gràcies a la meva família! Gràcies als meus tiets, i als meus avis Matias i Misi que sense entendre massa el que faig em recolzen i sempre es preocupen perquè faci el que faci em vagi bé. Gràcies a la Lola, una àvia ben tossuda i lluitadora que sempre ha estat present, començant per totes aquelles nits que em quedava dormida llegint i t'encarregaves d'apagar-me el llum, gràcies per preocupar-te sempre per mi. Per descomptat moltíssimes gràcies als meus pares, gràcies recolzar-me sempre des de que de ben petita vaig dir que volia treballar en un laboratori, gràcies per haver-me deixat seguir el meu camí i pel vostre suport incondicional!!! i sobretot gràcies per aguantar estoicament aquesta època de tesis! Res d'això no hagués sigut possible sense vosaltres!

Aquesta tesis és per tots vosaltres!

Gràcies!

¡Gracias!

Thank you!