

Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing

Javier del Campo,^{1*} Diego Mallo,² Ramon Massana,³ Colomán de Vargas,^{4,5} Thomas A. Richards^{6,7} and Iñaki Ruiz-Trillo^{1,8,9}

¹*Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.*

²*Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Galicia, Spain.*

³*Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.*

⁴*CNRS, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, UMR 7144, Roscoff, France.*

⁵*Station Biologique de Roscoff, UPMC University Paris 06, Roscoff UMR 7144, France.*

⁶*Geoffrey Pope Building, Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK.*

⁷*Canadian Institute for Advanced Research, CIFAR Program in Integrated Microbial Biodiversity, Toronto, Canada.*

⁸*Departament de Genètica, Universitat de Barcelona, Barcelona, Catalonia, Spain.*

⁹*Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.*

Summary

The opisthokonts are one of the major super groups of eukaryotes. It comprises two major clades: (i) the Metazoa and their unicellular relatives and (ii) the Fungi and their unicellular relatives. There is, however, little knowledge of the role of opisthokont microbes in many natural environments, especially among non-metazoan and non-fungal opisthokonts. Here, we begin to address this gap by analysing high-throughput 18S rDNA and 18S rRNA sequencing data from different European coastal sites, sampled

at different size fractions and depths. In particular, we analyse the diversity and abundance of choanoflagellates, filastereans, ichthyosporeans, nucleariids, corallochytreans and their related lineages. Our results show the great diversity of choanoflagellates in coastal waters as well as a relevant representation of the ichthyosporeans and the uncultured marine opisthokonts (MAOP). Furthermore, we describe a new lineage of marine fonticulids (MAFO) that appears to be abundant in sediments. Taken together, our work points to a greater potential ecological role for unicellular opisthokonts than previously appreciated in marine environments, both in water column and sediments, and also provides evidence of novel opisthokont phylogenetic lineages. This study highlights the importance of high-throughput sequencing approaches to unravel the diversity and distribution of both known and novel eukaryotic lineages.

Introduction

The monophyletic Opisthokonta clade represents one of the five 'super groups' of the eukaryotes (Adl *et al.*, 2012) and include two of the most well-known multicellular forms of life: the animals (Metazoa) and the Fungi. Two morphological synapomorphies have been suggested for opisthokonts: a single posterior flagellum and flat mitochondrial cristae (Cavalier-Smith, 1987). Opisthokonts also share a molecular synapomorphy, a 12-amino-acid insertion in the translation elongation factor 1 α gene (Baldauf and Palmer, 1993). Recent phylogenetic and phylogenomic analyses have shown that the opisthokonts also include a wide range of unicellular lineages separated into two clades: the Holomycota and the Holozoa (Fig. 1) (Steenkamp *et al.*, 2006; Shalchian-Tabrizi *et al.*, 2008; Ruiz-Trillo *et al.*, 2008; Brown *et al.*, 2009; Liu *et al.*, 2009; Torruella *et al.*, 2012; for a review see Paps and Ruiz-Trillo, 2010). The Holomycota comprise Fungi and its unicellular relatives, Nucleariida and *Fonticula alba*. On the other hand, the Holozoa comprise Metazoa and its unicellular relatives: the choanoflagellates, the filastereans, the ichthyosporeans and *Corallochytrium limacisporum* (Cavalier-Smith and Chao, 2003; Ruiz-Trillo *et al.*, 2004;

Received 6 October, 2014; revised 18 December, 2014; accepted 18 December, 2014. *For correspondence. E-mail javier.delcampo@botany.ubc.ca; Tel. +1 (604) 822 2845; Fax +1 (604) 822 6089. †Present Address: University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada.

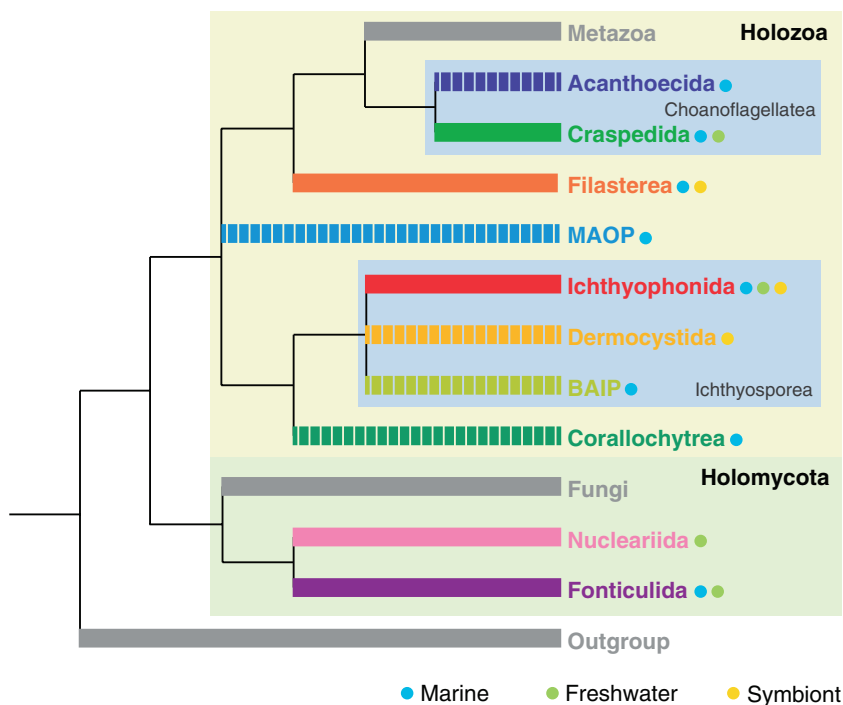


Fig. 1. Opisthokonts summary phylogenetic tree based on phylogenomic analyses (Torruella *et al.*, 2012). Clades with no phylogenomic information available are plotted based on 18S DNA data (del Campo and Ruiz-Trillo, 2013) and are indicated with discontinuous lines. Environmental distribution of the different groups is based on the results from this work and from del Campo and Ruiz-Trillo, 2013.

Steenkamp *et al.*, 2006; Torruella *et al.*, 2012; Paps *et al.*, 2013) as well as several uncultured lineages (del Campo and Ruiz-Trillo, 2013).

In parallel, our knowledge about the biology and lifestyles of the different opisthokont lineages is also improving. Several studies have been published describing the life cycle of some of those organisms such as the ichthyosporeans (Mendoza *et al.*, 2002; Marshall *et al.*, 2008; Marshall and Berbee, 2010; Glockling *et al.*, 2013; Suga and Ruiz-Trillo, 2013), the filasterean *Capsaspora owczarzaki* (Sebé-Pedrós *et al.*, 2013) or *F. alba* (Brown *et al.*, 2009) and peering deeper into the biology of the choanoflagellates (Fairclough *et al.*, 2010; Levin and King, 2013).

Microbial eukaryotes are one of the most abundant forms of life in the oceans and play key roles in marine ecosystems as primary producers as well as consumers and act as a link between the microbial loop and higher levels of the trophic web (Massana, 2011). Despite the importance of microbial opisthokonts from an evolutionary (Ruiz-Trillo *et al.*, 2008) and environmental point of view (Arndt *et al.*, 2000; Gozlan *et al.*, 2014), our current understanding of their diversity and ecology remains limited, especially among unicellular taxa. Molecular ecology approaches using high-throughput sequencing (HTS) of the 18S rDNA marker gene (Stoeck *et al.*, 2010) appear to be a powerful tool for boosting our understanding of abundance and diversity of eukaryotic microbes and characterizing the 'rare biosphere' (Pedrós-Alió, 2006). Some molecular studies have addressed this

question for Metazoa and Fungi (Fonseca *et al.*, 2010; Bik *et al.*, 2012; Richards *et al.*, 2012), but little work has been done for the rest of the opisthokonts. In this regard, recent meta-analyses of published environmental data provided some clues into the diversity of choanoflagellates and ichthyosporeans (del Campo and Massana, 2011; del Campo and Ruiz-Trillo, 2013) by showing that these two groups are better represented in marine environments in terms of richness than previously known and revealing the presence of novel groups such as the marine opisthokonts (MAOP). However, there has never been a systematic HTS analysis to investigate the diversity of non-fungi and non-metazoan unicellular opisthokonts in the marine environment. Therefore, little is known about the diversity and relative abundance of the filastereans, nucleariids, *F. alba* and Corallochytreia (*C. limacisporum*) lineages.

To fill this gap, we surveyed a large dataset (139 samples) of both 18S rRNA and rDNA tag sequences from six separate coastal locations in Europe: the North Sea (Oslo, Norway), the English Channel (Roscoff, France), the Bay of Biscay (Gijón, Spain), the Mediterranean Sea (Barcelona, Spain and Naples, Italy) and the Black Sea (Varna, Bulgaria) (Fig. S1). We used 454 sequencing methodology to obtain data from three size fractions in the plankton, picoplankton (0.8–3 µm), nanoplankton (3–20 µm), micro/mesoplankton (20–2,000 µm) and from sediments (Massana *et al.*, 2014). By using a manually curated opisthokont reference database (del Campo and Ruiz-Trillo, 2013) and a phylogenetic approach, we

annotated the 18S rDNA and 18S rRNA reads to identify the diversity and relative abundance of the unicellular opisthokonts and analyse their distribution across these compartments. We believe that this knowledge will not only provide clues into the ecological relevance of the different opisthokont clades in marine environments but will also help to establish a greater understanding of the ecological and phylogenetic diversity of the opisthokonts. This will provide an improved context for understanding the origin and diversification of both the fungal and meta-zoan lineages.

Results

Unicellular opisthokonts among eukaryotes

Our data show that the unicellular opisthokonts are not abundant compared to the other eukaryotes (discarding metazoans), accounting for up to 1% of the total sequences recovered. However, their presence is still relevant in terms of abundance in comparison with other protist groups present in this dataset. As reference, in the same BioMarKs dataset the Marine Stramenopiles (MAST), considered one of the most abundant heterotrophic protists groups in the ocean, account for approximately 6.5% of all eukaryotes on average in the picoplankton ribonucleic acid (RNA) (Logares *et al.*, 2012), while other groups like Perkinsea represent up to 0.26% of the total eukaryotes in sediments RNA (Chambouvet *et al.*, 2014). In particular, the unicellular opisthokonts represents up to 4% of the eukaryotic tags recovered from the picoplankton and nanoplankton RNA derived sequence surveys in some sites, with a mean value of ~1.5% (Fig. 2A.i, Fig. 2C). Similar values are observed when we look at the deep chlorophyll maximum (DCM) RNA (1.5%) and at the sediments, both RNA (2.7%) and deoxyribonucleic acid (DNA) (1.0%) samples (Fig. 2B). Unicellular opisthokonts are considerably abundant in the Varna samples, where part of the water column and the sediments are anoxic, accounting in some of the samples for up to 20% of the tags (Fig. 2A.iii).

Most unicellular opisthokonts lineages are well represented by high-throughput marine data

The diversity of major groups within the unicellular opisthokonts is, in general, well covered by our coastal sites sequence sampling (Fig. 3). In particular, we retrieve a significant number of OTUs (Operational taxonomic units) from all the analysed groups, except from the Filasterea and the Corallochytraea. In our dataset, the most abundant unicellular opisthokont groups are the Acanthocorbis choanoflagellate group (30.7% of the unicellular opisthokont reads) followed by the novel marine

Fenticulida (MAFO), identified in this study, within the Holomycota (15.8%). The Diaphanoeca choanoflagellate group (9.7%), the marine choanoflagellates group 3 (MACHO3, 8.9%) and the MAOP group 1 (MAOP1, 8.0%) (del Campo and Ruiz-Trillo, 2013) are also quite abundant in number of reads (Table S3). Three of these groups (Acanthocorbis, Diaphanoeca and MACHO3) belong to the Acanthoecida, the most abundant marine choanoflagellates (Tong, 1997a,b). The uncultured MAOP and the Ichthyosporea are dominant in the anoxic water column. We only retrieve a few representatives for the other big group of choanoflagellates, the Craspedida, which is mainly freshwater (del Campo and Ruiz-Trillo, 2013), although there are some marine species described (Jeuck *et al.*, 2014).

Abundance and diversity within the unicellular opisthokonts

Our data allow us to identify OTU distribution patterns with regards size fraction, depth and geographical site within the unicellular opisthokonts. For example, with regards the size, the acanthoecids dominate the smaller pico (84.8% RNA / 64.2% DNA) and nano (87.3% RNA / 78.0% DNA) fractions of oxic waters, while the MAOP dominate the micro-fraction (58.5% RNA / 51.8% DNA), although acanthoecids are quite abundant as well (35.4% RNA / 26.4% DNA) (Fig. 2C). In the case of the anoxic water column, the MAOP dominate the pico (69.2% RNA / 84.6% DNA) and the microplankton (61.5% RNA / 71.4% DNA), whereas the ichthyosporeans dominate the nanoplankton (50.5% RNA / 55.9% DNA) (Fig. 2D). Marine opisthokont is a polyphyletic and heterogeneous group. In our dataset, MAOP1 is the more abundant lineage among the MAOP, whereas the other two groups (MAOP2 and MAOP3) have low read abundance and low OTU diversity (Table S4).

Comparison of the opisthokont diversity between water column and sediments reveals a different pattern. Specifically, these data demonstrate that the oxic water column is clearly dominated by the Acanthoecida, both at the surface (81.6% RNA / 57.1% DNA) and the DCM (87.3% RNA / 78.4% DNA) and both RNA and DNA templates (Fig. 2B). In the case of the oxic sediments, the most abundant unicellular opisthokonts are the novel MAFO clade that dominates also both RNA (67.5%) and DNA (76.3%) samples (Fig. 2D). In the case of the anoxic sediments, only eight reads have been retrieved from the RNA extraction, all belonging to the Acanthoecida. More reads are retrieved in the case of DNA, and all of them belong to the ichthyosporeans (Fig. 2B), most of them to the uncultured marine ichthyosporeans group 1 (MAIP1).

Most of the abundance, however, is explained by a few OTUs. More than 75% of the abundance retrieved is

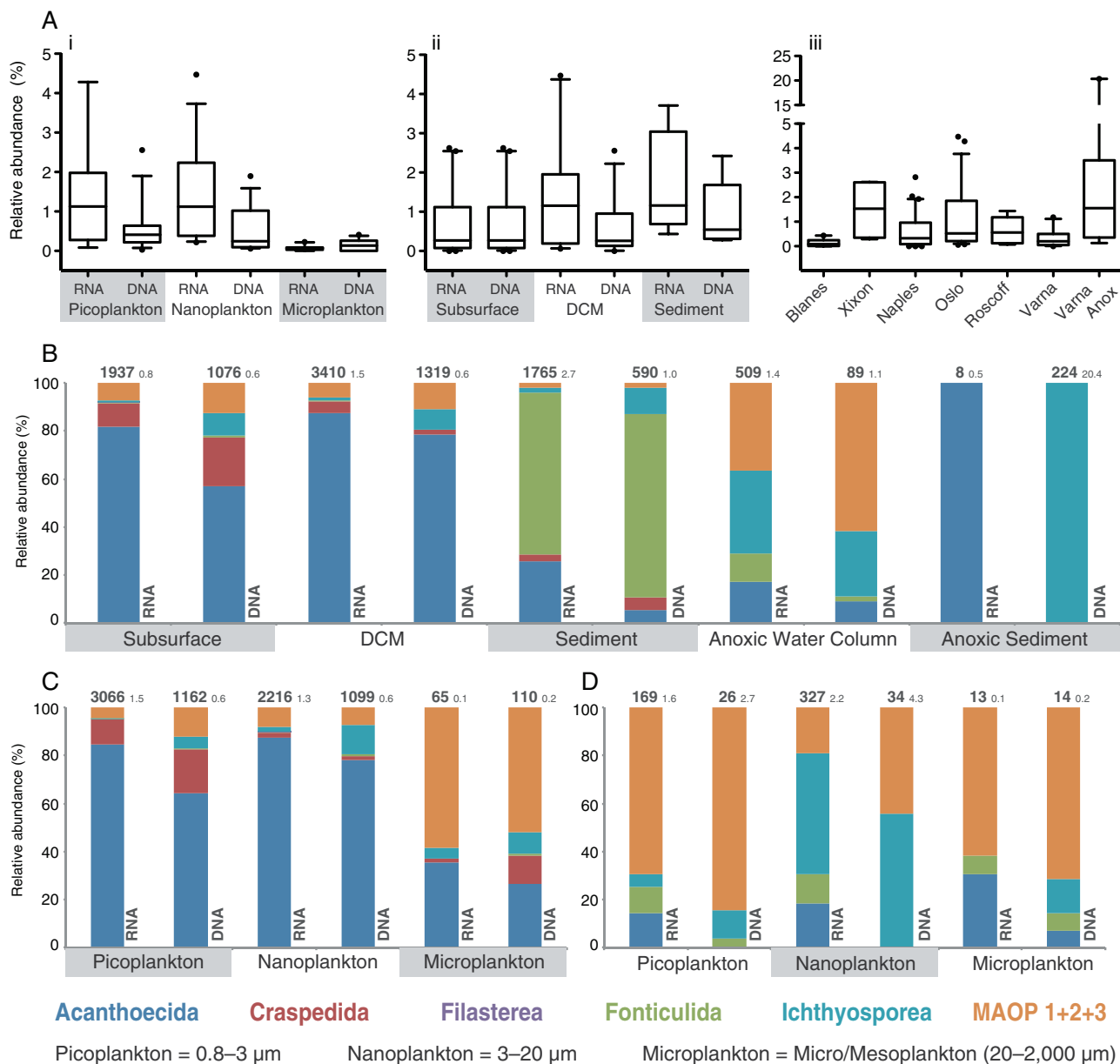


Fig. 2. A. Relative abundance of the unicellular opisthokonts among all the eukaryotes (excluding the metazoans) (A.i) at different size fractions from the oxic water column samples; (A.ii) at different depths; and (A.iii) in the different sampling sites. B. Relative abundance of the different opisthokonts groups at different depths; C. At different fractions among the oxic water column; D. And at different fractions among the anoxic water column. The correspondent numbers of analysed reads and their percentage among all the eukaryotic reads (smaller font) are shown above bars.

represented by only 8% of OTUs in the case of RNA and 14% in the case of the DNA. Indeed, three OTUs in the case of the RNA and two in the case of the DNA explain up to 25% of the whole opisthokont read abundance (Fig. 4). Interestingly, most of that abundance comes from novel sequences. Only 27.6% of the reads are more than 97% similar to any 18S DNA sequence from GenBank (Table S4) and among those only 39 OTUs (1%) are present (similarity of > 99%) in our opisthokonts reference

database (del Campo and Ruiz-Trillo, 2013). This differs from observations from other groups such as MAST, where most of the recovered reads were >98% similar to GenBank sequences (Logares *et al.*, 2012). Remarkably, just one out of the 10 most abundant OTUs is more than 97% similar to any 18S DNA sequence from a cultured organism. Therefore, more than 50% of the unicellular opisthokonts in our dataset represent uncultured taxa (Table S4).

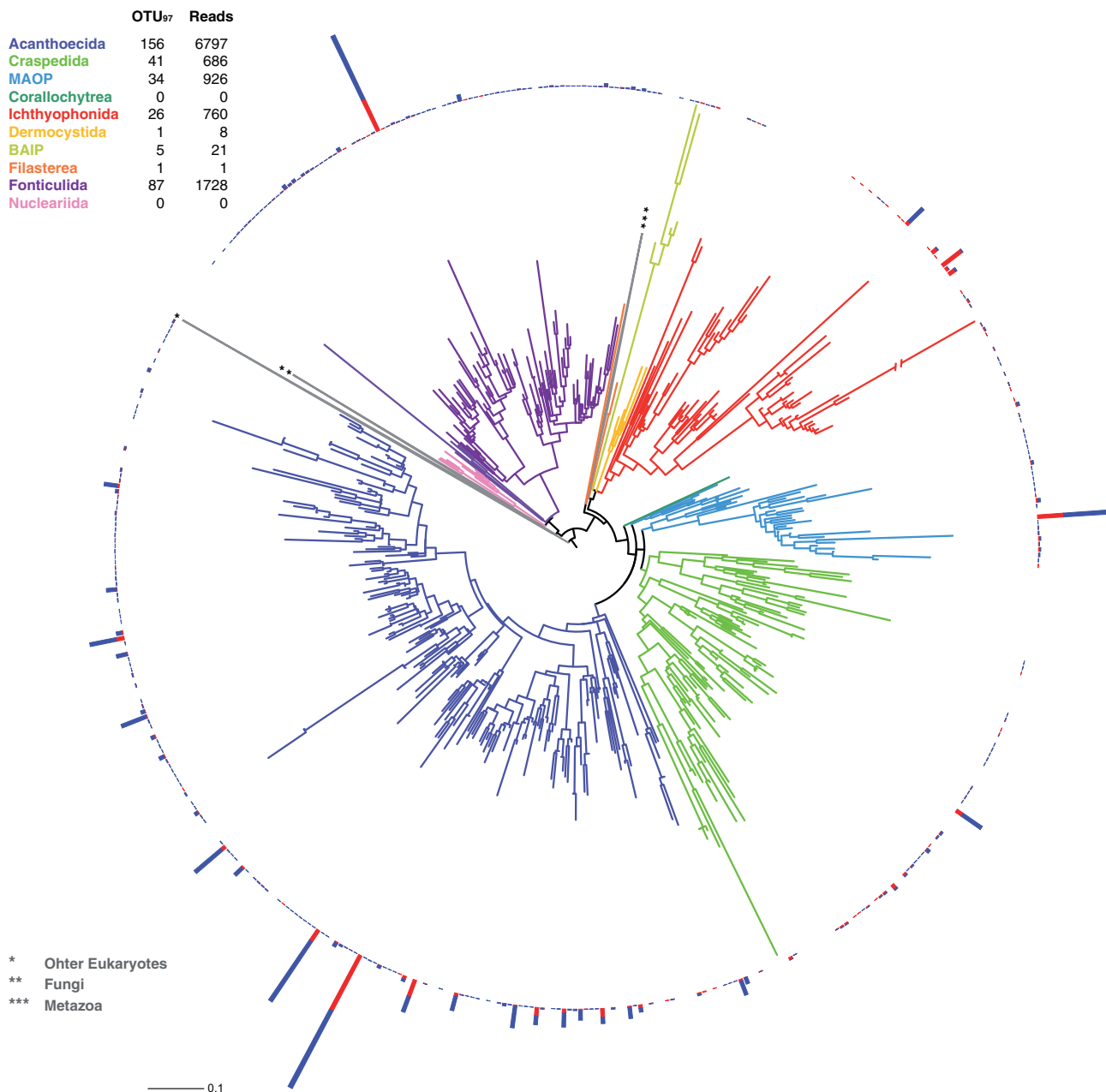


Fig. 3. Best maximum likelihood phylogenetic tree inferred from the 18S rDNA gene. Taxa include our reference database (OTU₉₇) (del Campo and Ruiz-Trillo, 2013) and BioMarKs V4 region sequences (OTU₉₇). The histograms represent the number of reads, (DNA reads as blue, RNA as red). The areas with no histogram bars correspond to reference sequences not obtained in this study. OTUs are colour coded according to taxonomic assignment as depicted in the figure.

Discussion

There is a great diversity of non-choanoflagellate unicellular opisthokonts in the marine environment

The choanoflagellates have always been considered to have an important role in the marine food chain (Arndt *et al.*, 2000; King, 2005). Traditionally, they have been detected by direct optical microscopy observation of marine samples (Fenchel, 1982; Arndt *et al.*, 2000).

Although they may not be as important as sometimes emphasized (Arndt *et al.*, 2000), they certainly have a significant role as bacterial consumers and as silica producers (in the case of the loricated forms). Therefore, it is clear that choanoflagellates have influence in both carbon and silica cycles.

Our data suggests that, besides the choanoflagellates, there are other unicellular opisthokonts lineages playing ecological roles in marine environments. Overall in our

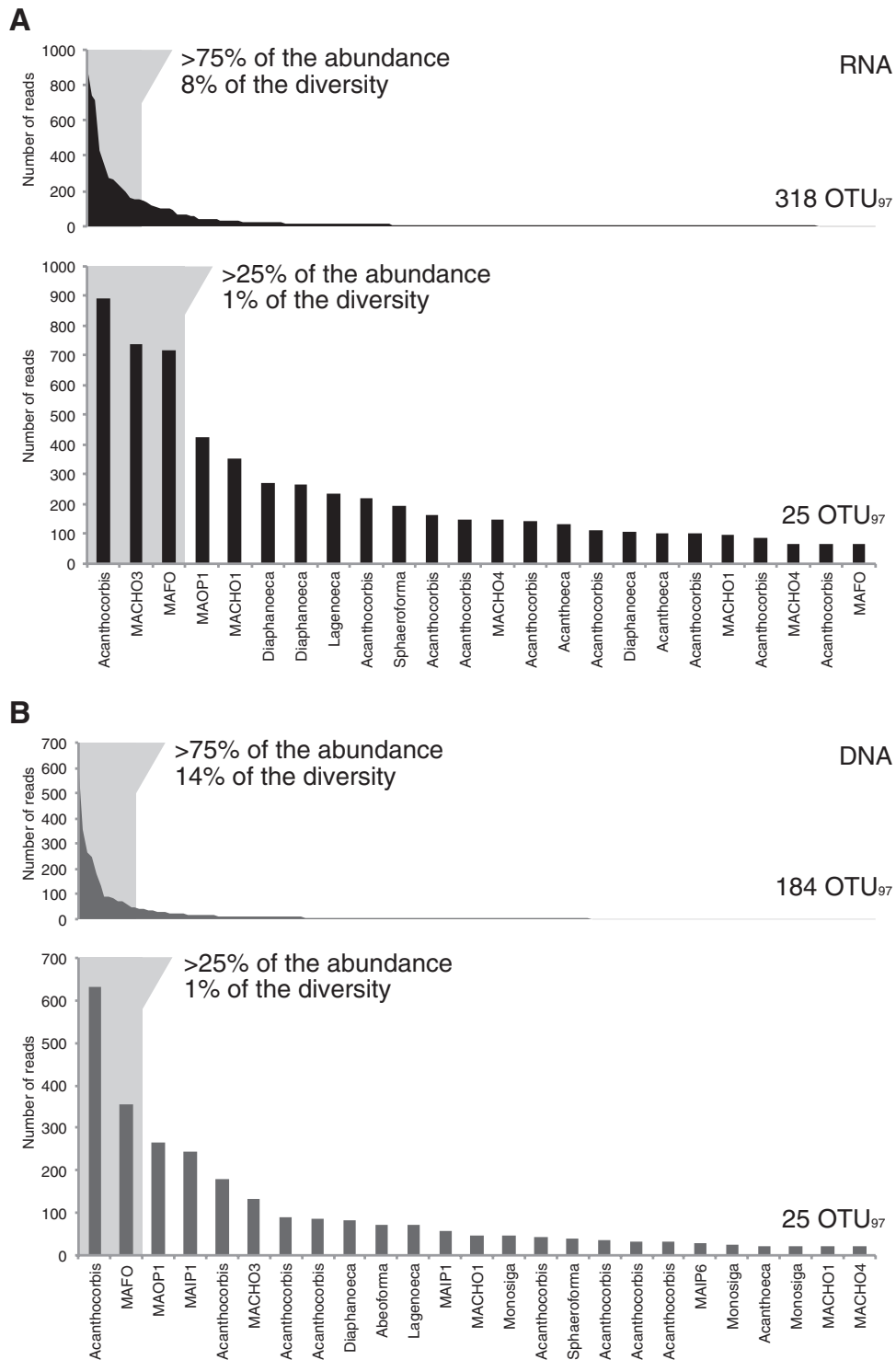


Fig. 4. Unicellular opisthokonts diversity distribution for (A) RNA and (B) DNA. The 25 most abundant OTU₉₇ are also shown for each case.

dataset, these other lineages account for 44% of unicellular opisthokonts diversity (Table S3). Importantly, this analysis has increased the number of sampled unicellular opisthokonts by 352 OTUs₉₇, thus dramatically expanding

the diversity of the Opisthokonta phylogenetic tree across six out of the 10 the major groups described in Fig. 1. We must highlight that part of the diversity at a lower phylogenetic level may potentially be the result of

artefacts introduced during the amplification process (Hadziavdic *et al.*, 2014). However, we believe that most of the recovered diversity represents real biological diversity considering the different filtering methods that were applied. It is worth mentioning that we could not have described this newly sampled diversity without the use of phylogenetic methods combined with a reliable, curated database of opisthokont sequences (del Campo and Ruiz-Trillo, 2013). Thus, our results highlights the importance of high-throughput data to understand the diversity of eukaryotes (Bittner *et al.*, 2013; Chambouvet *et al.*, 2014; Massana *et al.*, 2014) and demonstrates that the best approach currently available is to use phylogenetic analyses combined with well-curated databases as references for diversity tag annotation.

The emergence of a novel clade of Holomycota, the Marine Fonticulida

One of the main findings of this work has been the discovery of MAFO, a highly diverse sister group to *F. alba*, representing 24% of the unicellular opisthokonts OTU₉₇ in our study (Table S1). These data demonstrate that this group is abundant and ribosomally active in the oxic sediments of all the BioMarKs sampling sites. All previously described nucleariids and *F. alba* are freshwater organisms. As such, the recovery of this group could be explained by the effect of sediment transportation from the mainland or the rivers into the coastal environments sampled. However, the only three GenBank sequences that cluster within the MAFO group came also from marine surveys (one from an unpublished study and two from Edgcomb *et al.*, 2011). The emergence of MAFO in our dataset is probably a result of a more extensive sediment sampling performed here as compared with previous studies (del Campo and Massana, 2011). Therefore, we believe that MAFO is indeed a marine group representing the first non-fungal marine holomycota clade.

Based on our findings, there is not much that can be said about this new group besides that it appears to reside in oxic sediments (thus, not fractionated by size), and is likely phylogenetic affiliated to *F. alba*. According to its phylogenetic position it may potentially be an amoebal morphotype organism. Its presence in the RNA fraction indicates that it is not only abundant, but also ribosomally active. Therefore, our data show a second, previously undescribed clade among the Fonticulida, besides *F. alba*, which represents the only known species from this part of the eukaryotic tree of life (Brown *et al.*, 2009).

The ecological roles of acanthoecid choanoflagellates

Our data show that the Acanthorcorbis groups including Acanthoecida are by far the most abundant unicellular

opisthokonts in our studied environments. This is not surprising given that the acanthoecids are one of the best known choanoflagellates groups in marine and brackish waters (Arndt *et al.*, 2000). The growth rates for acanthoecids, which are bacterial feeders, seem to vary between 0.03 h⁻¹ and 0.1 h⁻¹ and their K_{max} is close to 2.5 × 10⁶ bacteria ml⁻¹ (Eccleston-Parry and Leadbeater, 1994). This places the acanthoecids in terms of feeding rate somewhere between the most common heterotrophic flagellates in culture, such as *Paraphysomans imperforata* or *Bodo sp.* (Eccleston-Parry and Leadbeater, 1994) and some of the most abundant oceanic bacterial consuming protists such as MAST 1, 2 and 4 groups (Massana *et al.*, 2006) or *Minorisa minuta* (del Campo *et al.*, 2013).

Moreover, the Acanthoecida have been considered to be sedentary and attached to a substrate or passively suspended (Carr *et al.*, 2008). The presence of acanthoecids in surface and deep chlorophyll maximum waters at the smallest fractions seems to suggest that acanthoecids may be passively suspended. The complex lorica that surrounds the acanthoecids cells is a good method to avoid sinking to the bottom of the ocean by allowing the cells to become transported by the water currents (Leakey *et al.*, 2002; Leadbeater *et al.*, 2009). This way, the acanthoecids may remain suspended in the surface and DCM where their preys are more abundant and where the dissolved Si is available for lorica synthesis (Tréguer and De La Rocha, 2013). Moreover, the lorica may also be a good protection against predation (Tong, 1997a). Also the possible disadvantage of not being active swimming predators can be compensated by their efficient strategy of filtering water as a way of catching bacteria (Arndt *et al.*, 2000). Given the diversity unraveled in this study, one cannot discard the possibility that there may also be active swimming species of Acanthoecida.

Another aspect to be considered is the morphology of the retrieved acanthoecids. The described lorica for the known acanthoecids is bigger in diameter than 5 µm. Thus, given this size, it is somehow surprising that in our study the acanthoecids are dominant in the picoplankton fractions which preferentially samples cells < 3 µm. There are two possible explanations: (i) the lorica breaks or retracts during the filtration process, allowing the cell to cross the filter pores, or (ii) there is a pool of unknown cell diversity corresponding to choanoflagellate forms smaller than cells currently available in culture or previously described.

Moreover, the ecological data retrieved in our study can be useful to understand evolutionary processes among the opisthokonts. For example, it is known that a sphingolipid produced by the eubacteria *Algoriphagus machipongonensis* triggers the colony formation in the

craspedid choanoflagellate *Salpingoeca rosetta* (Alegado *et al.*, 2012). It has indeed been suggested that a bacterivore organism could have been at the base of the origin of the metazoans and that eukaryote–bacterial interactions may have played a role in animal origins (McFall-Ngai *et al.*, 2013). Thus, in order to understand the origin of multicellularity, it is important to establish not only the evolutionary context that allowed animals to emerge but also the ecological framework to explain why and where this process evolved.

The poorly known ichthyosporeans and their putative role in marine environments

The ichthyosporeans (or Mesomycetozoa) have been described as a group of parasites (Ragan *et al.*, 1996; Mendoza *et al.*, 2002). However, their ecology remains fairly unknown, and according to Glockling *et al.*, 2013 the life cycle of most of these taxa is probably incomplete. For example, *Sphaeroforma arctica* and *Creolimax fragrantissima* within the *Sphaeroforma* group (Marshall *et al.*, 2008) and *Abeoforma whisleri* and *Pirum gemmata* that constitute the *Abeoforma* group (Marshall and Berbee, 2010) are known to live associated with invertebrates, but their role as parasites has not yet been proven. The presence and abundance of ichthyosporeans in environmental surveys suggest that they may indeed have some free-living stage or there may exist some currently undescribed members that are free living (del Campo and Ruiz-Trillo, 2013). Alternatively, the reads may arise from parasites, whose host cell has been destroyed or sampled during the process of filtering at the sample collection stage. According to our data, both the *Sphaeroforma* and *Abeoforma* groups are present in the smallest fractions of the water column of both oxic and anoxic samples and MAIP1 is dominant in anoxic DNA sediments. Interestingly, and in contrast to the *Sphaeroforma*, *Abeoforma* and MAIP groups, the number of reads found for those taxa that are known to be parasites such as the *Dermocystida* (Ragan *et al.*, 1996) or *Anurofeca* (Baker *et al.*, 1999) are minimal. Therefore, we believe this supports the hypothesis that some of the ichthyosporeans sequences sampled may be derived from free-living stages.

The case of MAIP1 is, however, different than the *Sphaeroforma* and *Abeoforma* groups. MAIP dominates the sediment fraction at the anoxic environment (20% of all the eukaryotic reads; Fig. 2), which has not been fractionated by size. Moreover, MAIP is only present in the DNA template sampled while absent in equivalent RNA samples indicating that the DNA detected probably arises from dead or dormant cells. Thus, in this case, it is possible that MAIP sequences are being retrieved from inside their host.

The uncultured Marine Opisthokonts, Filasterea and Corallochytreia diversity

Our data show that the MAOP clade (mainly MAOP1) is present in all the depths and fractions and are the dominant unicellular opisthokonts at the oxic microplankton fraction and the anoxic water column in both the DNA and RNA derived diversity tag sequence samples (Fig. 2). In previous studies, MAOP1 sequences were found in both oxic and micro-oxic environments (Romari and Vaultot, 2004; Cheung *et al.*, 2008; Not *et al.*, 2008; Amacher *et al.*, 2009; Edgcomb *et al.*, 2011). So the most likely explanations are that the MAOP group has either a wide size range among its members or has a cell cycle that include different cells sizes and a putative ability to live in micro-oxic and anoxic environments or being a parasite inside larger cells.

Surprisingly, we only retrieve one read from the Filasterea and none from the Corallochytreia. We expected a higher number of reads from these groups given that representative marine organisms have been described for both groups: *Ministeria vibrans* (Tong, 1997a) and *Corallochytrium limacisporum* (Raghukumar, 1987). However, it is known that some taxa are not detected in environmental surveys because of technical issues, such as primer bias (Berney *et al.*, 2004). In this particular case, we observe that the forward primer used presented a 4–5 bp mismatch for the two available 18S *Ministeria sp.* sequences and 1 mismatch for *C. limacisporum*. Thus, these mismatches in the primer sequence may account to the lack of reads from those groups. In addition, *C. limacisporum* has so far only been detected in coral reefs, so it might be endemic from this environment and therefore not present in the environments we sampled. With the exception of Filasterea and Corallochytreia, our data cover most of the unicellular opisthokonts lineages. Moreover, it recovers novel and previously unsampled groups (see below and Table S4).

HTS unravels unknown diversity

Our results are a new example that the most abundant microorganisms in natural systems correspond mainly to species that are not recovered into sustainable culture, highlighting the existence of a culture bias in the unicellular opisthokonts (del Campo and Ruiz-Trillo, 2013; del Campo *et al.*, 2013) (Fig. 4). Given the evolutionary interest and environmental diversity of these taxa, we believe it is important to try to identify and isolate these organisms. Alternatively, single-cell sampling techniques could be used to obtain genomic and transcriptomic data from those taxa (del Campo *et al.*, 2014).

Our data distribution is, in general, similar to the dataset we had generated in our previous meta-analysis

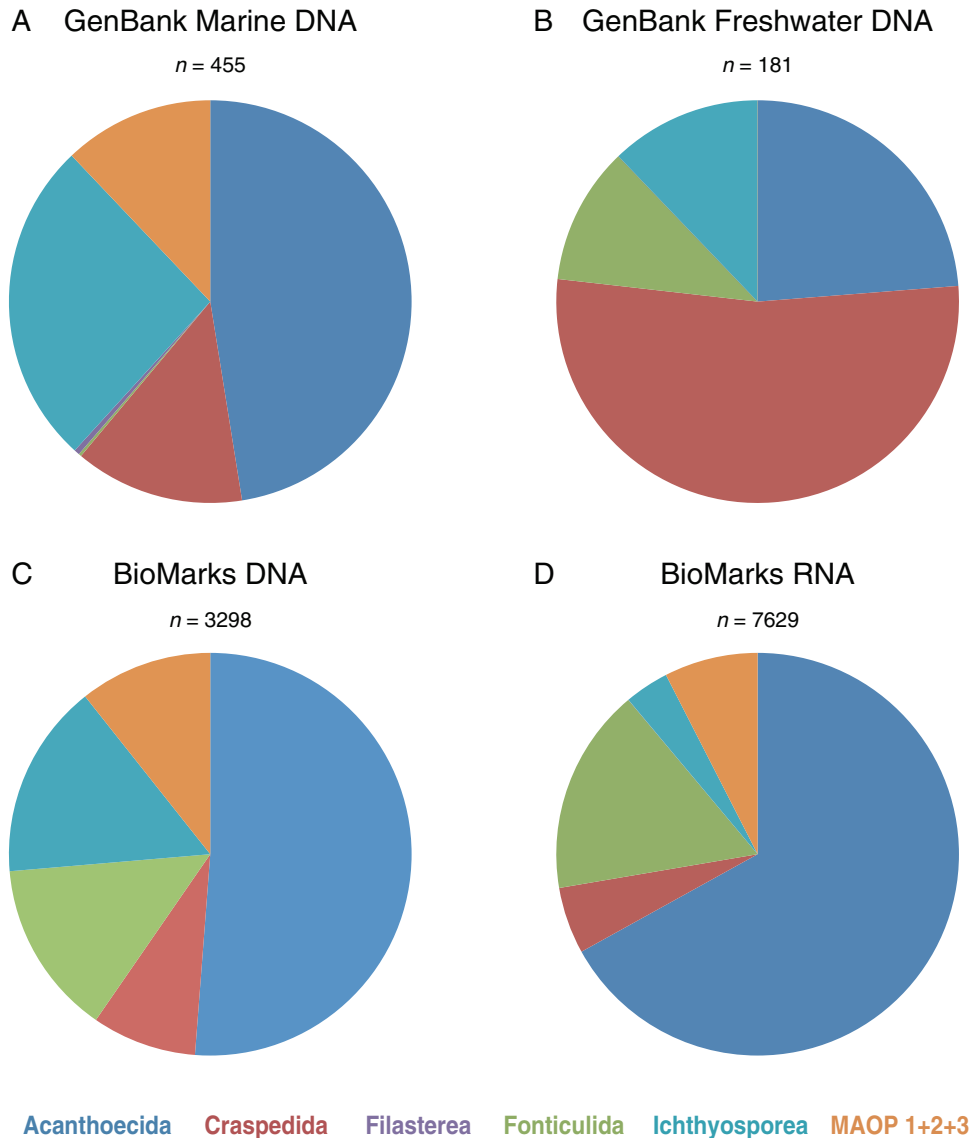


Fig. 5. Relative abundance of unicellular opisthokonts 18S rDNA reads from (A) marine environments in GenBank (B) freshwater environments in GenBank (C) BioMarks 18rDNA (D) 18S rRNA reads from BioMarks.

of marine environmental surveys (del Campo and Ruiz-Trillo, 2013), except for the appearance of the novel Holomycota group MAFO within Fonticulida (Fig. 5). This indicates that meta-analysis of environmental surveys for the different protist groups is necessary and useful from both an ecological and phylogenetically point of view. Indeed, the reference trees and reference sequence databases generated by those studies are the backbone for the present HTS studies. However, the diversity and distribution pattern retrieved from the RNA samples differs slightly to the distribution observed in previous analysis. In particular, here we observe a higher abundance and diversity of Acanthoecida choanoflagellates

as well as the non-fungal Holomycota, indicating that these two groups are both ribosomally active in their respective environment: the water column and the sediments (Not *et al.*, 2009).

Moreover, these initial HTS studies are the first step to broadly recognize potentially important ecological actors from the different taxonomic groups. In this analysis, we have been able to identify the more relevant unicellular opisthokont groups at coastal environments, which are likely to play an important role as bacterial predators or as parasites. The next step will be to target these taxa for genomic analyses in order to increase our understanding of their evolutionary and ecological role.

Material and methods

Opisthokonts sequences derived from 454-BioMarkS dataset

During the BioMarkS project (biomarks.eu), samples were collected in six European coastal sites at different times of the year, at different water column depths (surface and DCM) and sampling three size fractions (pico-, nano- and micro/mesoplankton) and sediment samples (Fig. S1, Table S1). Extracts of DNA and RNA were obtained from these samples, the later reverse transcribed to complementary DNA (cDNA; check Logares *et al.*, 2012 for details), and analyses of protist diversity were constructed by 454 pyrosequencing of eukaryotic specific V4 18S rDNA PCR amplicons using the following primers: TAReuk454FWD1 (5'-CCAGCA(G/C)C(C/T)GCGGTAAT TCC-3') and TAReukREV3 (5'-ACTTTCGTTCTTGAT (C/T)(A/G)A-3') (Stoeck *et al.*, 2010). Pyroreads were inspected to remove short reads, reads with low quality and chimeras as described in Logares and colleagues (2012) (Table S2). An OTU table (reads per sample) was conducted using the USEARCH (Edgar, 2010) tool at a clustering threshold of 97% similarity (OTU₉₇). Analysed was eDNA (DNA in background environmental water; Bohmann *et al.* 2014) in order to determine its possible influence on the dataset and just 39 reads were retrieved. Thus, eDNA has no impact on the final results. OTUs were taxonomically classified using several eukaryotic reference datasets (Guillou *et al.*, 2013; Pernice *et al.*, 2013; Yilmaz *et al.*, 2014) including an unicellular opisthokont specific reference database (del Campo and Ruiz-Trillo, 2013), and primarily assigned to a group when they had an e-value below 10⁻¹⁰⁰ against a reference sequence. Metazoan pyrotags were removed from the OTU table before calculating the contributions of different groups. Phylogenetic assignment of the opisthokonts OTUs was then further validated using phylogeny.

Phylogenetic analyses

An alignment was constructed with 352 unique unicellular opisthokonts V4 reads and 345 18 rDNA reference sequences sampled from GenBank nr Database. Sequences were aligned using MAFFT auto mode (-auto) (Katoh *et al.*, 2002) using a set of representative eukaryotic taxa as an out-group. Alignments were checked using SEAVIEW 4 (Gouy *et al.*, 2010), and highly variable regions of the alignment were removed using TRIMAL 1.2 (-gt 0.3 -st 0.001) (Capella-Gutiérrez *et al.*, 2009). A maximum likelihood (ML) phylogenetic tree with complete, and partial V4 region sequences were constructed with the FASTML method of RAXML 7.2.6 (Stamatakis, 2006) assuming the GTR-CAT-I substitution model. The resulting ML tree came out of 2000 independ-

ent tree searches (starting from distinct randomized maximum parsimony trees). In order to assess tree uncertainty, a non-parametric bootstrap analysis was performed with 2000 replicates. Tree figures were edited with FIGTREE v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and iTOL (Letunic and Bork, 2011).

Diversity and distribution patterns in unicellular opisthokonts using 454

A total of 10 927 'unicellular opisthokonts' (i.e. the non-fungi, non-metazoan opisthokonts, including choanoflagellates, filastereans, ichthyosporeans, nucleariids, *F. alba* and Chorallochytreia) reads retrieved from different sites, depths and fractions were used to define the diversity and distribution of these organisms across the European coastal sites sampled by the BioMarkS consortium. For each sample, DNA and cDNA (RNA derived) reads were processed and analysed to obtain information from both presence and ribosomal activity. The unicellular opisthokonts OTU table obtained (Information S1, Information S2) was processed for community analysis using QIIME (Caporaso *et al.*, 2010).

Acknowledgements

The authors are grateful to Stéphane Audic for bioinformatic support. This work was supported by an Institució Catalana de Recerca i Estudis Avançats contract, a European Research Council Starting Grant (ERC-2007-StG-206883), and a grant (BFU-2011-23434) from Ministerio de Economía y Competitividad (MINECO) to I.R.-T. D.M. was supported by a fellowship (BES-2010-031014 at the University of Vigo) from MINECO. The work is part of EU ERA-Net program BiodivERsA, under the project BioMarkS (Biodiversity of Marine euKaryotes).

References

- Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., *et al.* (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**: 429–514.
- Alegado, R.A., Brown, L.W., Cao, S., Dermenjian, R.K., Zuzow, R., Fairclough, S.R., *et al.* (2012) A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *Elife* **1**: e00013.
- Amacher, J., Neuer, S., Anderson, I.J., and Massana, R. (2009) Molecular approach to determine contributions of the protist community to particle flux. *Deep Sea Res Part I Oceanogr Res Pap* **56**: 2206–2215.
- Arndt, H., Dietrich, D., Auer, B., Clevén, E.J., Gräfenhan, T., Weitere, M., and Mylnikov, A.P. (2000) Functional diversity of heterotrophic flagellates in aquatic ecosystems. In *The Flagellates: Unity, Diversity and Evolution*. Leadbeater, B.S.C., and Green, J.C. (eds). London, UK: CRC Press, pp. 240–268.
- Baker, G.C., Beebee, T.J., and Ragan, M.A. (1999) *Prototheca richardsi*, a pathogen of anuran larvae, is

- related to a clade of protistan parasites near the animal-fungal divergence. *Microbiology* **145**: 1777–1784.
- Baldauf, S.L., and Palmer, J.D. (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci* **90**: 11558–11562.
- Berney, C., Fahrni, J.F., and Pawlowski, J. (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol* **2**: 13.
- Bik, H.M., Sung, W., De Ley, P., Baldwin, J.G., Sharma, J., Rocha-Olivares, A., and Thomas, W.K. (2012) Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Mol Ecol* **21**: 1048–1059.
- Bittner, L., Gobet, A., Audic, S., Romac, S., Egge, E.S., Santini, S., *et al.* (2013) Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. *Mol Ecol* **22**: 87–101.
- Bohmann, K., Evans, A., Gilbert, M.T., Carvalho, G.R., Creer, S., Knapp, M., *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol* **29**: 358–367.
- Brown, M.W., Spiegel, F.W., and Silberman, J.D. (2009) Phylogeny of the 'forgotten' cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Mol Biol Evol* **26**: 2699–2709.
- del Campo, J., and Massana, R. (2011) Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist* **162**: 435–448.
- del Campo, J., and Ruiz-Trillo, I. (2013) Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Mol Biol Evol* **30**: 802–805.
- del Campo, J., Balagué, V., Forn, I., Lekunberri, I., and Massana, R. (2013) Culturing bias in marine heterotrophic flagellates analyzed through seawater enrichment incubations. *Microb Ecol* **66**: 489–499.
- del Campo, J., Not, F., Forn, I., Sieracki, M.E., and Massana, R. (2013) Taming the smallest predators of the oceans. *ISME J* **7**: 351–358.
- del Campo, J., Sieracki, M.E., Molestina, R.E., Keeling, P.J., Massana, R., and Ruiz-Trillo, I. (2014) The others: our biased perspective of eukaryotic genomes. *Trends Ecol Evol* **29**: 252–259.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Carr, M., Leadbeater, B.S.C., Hassan, R., Nelson, M., and Baldauf, S.L. (2008) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci* **105**: 16641–16646.
- Cavalier-Smith, T. (1987) The origin of fungi and pseudofungi. In *Evolutionary Biology of the Fungi*. Rayner, A.D.M., Brasier, C.M., and Moore, D. (eds). Cambridge, UK: Cambridge University Press, pp. 339–353.
- Cavalier-Smith, T., and Chao, E.E.-Y. (2003) Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *J Mol Evol* **56**: 540–563.
- Chambouvet, A., Berney, C., Romac, S., Audic, S., Maguire, F., De Vargas, C., and Richards, T.A. (2014) Diverse molecular signatures for ribosomally 'active' Perkinsea in marine sediments. *BMC Microbiol* **14**: 110.
- Cheung, M.K., Chu, K.H., Li, C.P., Kwan, H.S., and Wong, C.K. (2008) Genetic diversity of picoeukaryotes in a semi-enclosed harbour in the subtropical western Pacific Ocean. *Aquat Microb Ecol* **53**: 295–305.
- Eccleston-Parry, J.D., and Leadbeater, B.S.C. (1994) A comparison of the growth kinetics of six marine heterotrophic nanoflagellates fed with one bacterial species. *Mar Ecol Prog Ser* **105**: 167–177.
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Edgcomb, V., Orsi, W., Bunge, J.A., Jeon, S., Christen, R., Leslin, C., *et al.* (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* **5**: 1344–1356.
- Fairclough, S.R., Dayel, M.J., and King, N. (2010) Multicellular development in a choanoflagellate. *Curr Biol* **20**: R875–R876.
- Fenchel, T. (1982) Ecology of heterotrophic microflagellates. I. Some important forms and their functional morphology. *Mar Ecol Prog Ser* **9**: 211–223.
- Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P., *et al.* (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat Commun* **1**: 98.
- Glockling, S.L., Marshall, W.L., and Gleason, F.H. (2013) Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). *Fungal Ecol.* **6**: 237–247.
- Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221–224.
- Gozlan, R.E., Marshall, W.L., Lilje, O., Jessop, C.N., Gleason, F.H., and Andreou, D. (2014) Current ecological understanding of fungal-like pathogens of fish: what lies beneath?. *Front Microbiol* **5**: 1–16.
- 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–D604.
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., and Troedsson, C. (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE* **9**: e87624.
- Jeuck, A., Arndt, H., and Nitsche, F. (2014) Extended phylogeny of the Craspedida (Choanomonada). *Eur J Protistol* **50**: 430–443.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.
- King, N. (2005) Choanoflagellates. *Curr Biol* **15**: 113–114.
- Leadbeater, B.S.C., Yu, Q., Kent, J., and Stekel, D.J. (2009) Three-dimensional images of choanoflagellate loricae. *Proc R Soc B Biol Sci* **276**: 3–11.

- Leakey, R.J.G., Leadbeater, B.S.C., Mitchell, E., McCready, S.M.M., and Murray, A.W.A. (2002) Abundance and biomass of choanoflagellates and other nanoflagellates in waters of contrasting temperature to the north-west of South Georgia in the Southern Ocean. *Eur J Protistol* **38**: 333–350.
- Letunic, I., and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475–W478.
- Levin, T.C., and King, N. (2013) Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. *Curr Biol* **23**: 2176–2180.
- Liu, Y., Steenkamp, E.T., Brinkmann, H., Forget, L., Philippe, H., and Lang, B.F. (2009) Phylogenomic analyses predict sister group relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol Biol* **9**: 272.
- Logares, R., Audic, S., Santini, S., Pernice, M.C., de Vargas, C., and Massana, R. (2012) Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J* **6**: 1823–1833.
- McDonald, D., Clemente, J.C., Kuczynski, J., Rideout, J.R., Stombaugh, J., Wendel, D., et al. (2012) The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* **1**: 7.
- McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci* **110**: 3229–3936.
- Marshall, W.L., and Berbee, M.L. (2010) Facing unknowns: living cultures (*Pirum gemmata* gen. nov., sp. nov., and *Abeoforma whisleri*, gen. nov., sp. nov.) from Invertebrate Digestive Tracts Represent an Undescribed Clade within the Unicellular Opisthokont Lineage Ichthyosporea (Mesomycetozoa). *Protist* **162**: 33–57.
- Marshall, W.L., Celio, G., McLaughlin, D.J., and Berbee, M.L. (2008) Multiple isolations of a culturable, motile Ichthyosporean (Mesomycetozoa, Opisthokonta), *Creolimax fragrantissima* n. gen., n. sp., from marine invertebrate digestive tracts. *Protist* **159**: 415–433.
- Massana, R. (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65**: 91–110.
- Massana, R., Guillou, L., Terrado, R., Forn, I., and Pedrós-Alió, C. (2006) Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol* **45**: 171–180.
- Massana, R., del Campo, J., Sieracki, M.E., Audic, S., and Logares, R. (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J* **8**: 854–866.
- Mendoza, L., Taylor, J.W., and Ajello, L. (2002) The class mesomycetozoa: a heterogeneous group of microorganisms at the animal-fungal boundary. *Annu Rev Microbiol* **56**: 315–344.
- Not, F., Latasa, M., Sucharek, R., Viprey, M., Karleskind, P., Balagué, V., et al. (2008) Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep Sea Res Part I Oceanogr Res Pap* **55**: 1456–1473.
- Not, F., del Campo, J., Balagué, V., De Vargas, C., and Massana, R. (2009) New insights into the diversity of marine picoeukaryotes. *PLoS ONE* **4**: 7.
- Paps, J., and Ruiz-Trillo, I. (2010) *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd.
- Paps, J., Medina-Chacón, L.A., Marshall, W.L., Suga, H., and Ruiz-Trillo, I. (2013) Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. *Protist* **164**: 2–12.
- Pedrós-Alió, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.
- Pernice, M.C., Logares, R., Guillou, L., and Massana, R. (2013) General patterns of diversity in major marine microeukaryote lineages. *PLoS ONE* **8**: e57170.
- Ragan, M.A., Goggin, C.L., Cawthorn, R.J., Cerenius, L., Jamieson, A., V, Plourde, S.M., et al. (1996) A novel clade of protistan parasites near the animal-fungal divergence. *Proc Natl Acad Sci* **93**: 11907–11912.
- Raghukumar, S. (1987) Occurrence of the traustochytrid *Corallochytrium limacisporum* gen. et sp. nov. in the coral reef lagoons of the Lakshadweep Islands in the Arabian Sea. *Bot Mar* **30**: 83–89.
- Richards, T.A., Jones, M.D.M., Leonard, G., and Bass, D. (2012) Marine fungi: their ecology and molecular diversity. *Ann Rev Mar Sci* **4**: 495–522.
- Romari, K., and Vaulot, D. (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49**: 784–798.
- Ruiz-Trillo, I., Inagaki, Y., Davis, L.L.A., Landfald, B., Roger, A.J., and Sperstad, S. (2004) *Capsaspora owczarzakii* is an independent opisthokont lineage. *Curr Biol* **14**: R946–R947.
- Ruiz-Trillo, I., Roger, A.J., Burger, G., Gray, M.W., and Lang, B.F. (2008) A phylogenomic investigation into the origin of metazoa. *Mol Biol Evol* **25**: 664–672.
- Sebé-Pedrós, A., Irimia, M., del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., et al. (2013) Regulated aggregative multicellularity in a close unicellular relative of metazoa. *Elife* **2**: e01287.
- Shalchian-Tabrizi, K., Minge, M.A., Espelund, M., Orr, R.J.S., Ruden, T., Jakobsen, K.S., and Cavalier-Smith, T. (2008) Multigene phylogeny of choanozoa and the origin of animals. *PLoS ONE* **3**: 2098.
- Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Steenkamp, E.T., Wright, J., and Baldauf, S.L. (2006) The protistan origins of animals and fungi. *Mol Biol Evol* **23**: 93–106.
- Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., and Richards, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19**: 21–31.
- Suga, H., and Ruiz-Trillo, I. (2013) Development of ichthyosporeans sheds light on the origin of metazoan multicellularity. *Dev Biol* **377**: 284–292.

- Tong, S.M. (1997a) Heterotrophic flagellates and other protists from Southampton Water, U.K. *Ophelia* **47**: 71–131.
- Tong, S.M. (1997b) Heterotrophic flagellates from the water column in Shark Bay, Western Australia. *Mar Biol* **128**: 517–536.
- Torruella, G., Derelle, R., Paps, J., Lang, B.F., Roger, A.J., Shalchian-Tabrizi, K., and Ruiz-Trillo, I. (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single copy protein domains. *Mol Biol Evol* **29**: 531–544.
- Tréguer, P.J., and De La Rocha, C.L. (2013) The world ocean silica cycle. *Ann Rev Mar Sci* **5**: 477–501.
- Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., *et al.* (2014) The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res* **42**: D643–D648.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. BioMarKs sampling map and dataset characteristics summary.

Fig. S2. 2000 trees ML 18S phylogeny including our reference database (OTU₉₇) in black and BioMarKs V4 region sequences (OTU₉₇) in red. All major clades are defined.

Table S1. Description of the sampling sites.

Table S2. Summary of samples including the total number of reads after quality control.

Table S3. Number of reads and number of OTU₉₇ for every taxonomic group found. Percentage of total reads number for the different groups (among the eukaryotes and among the unicellular opisthokonts). Percentage of OTU₉₇ for the different groups among the unicellular opisthokonts.

Table S4. OTU₉₇ taxonomic classification, the closest GenBank match for every OTU₉₇, similarity and size of the aligned fragment and e-value. For the most abundant OTU₉₇ the BLAST information for the closest cultured match is also included.

Information S1. BioMarKs unicellular opisthokonts OTU₉₇ table in biome format (McDonald *et al.*, 2012).

Information S2. BioMarKs samples map file to develop the community analysis using QIIME (Caporaso *et al.*, 2010).