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

Protist

ORIGINAL PAPER

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

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

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

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

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

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

Introduction

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

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

doi:10.1016/j.protis.2010.10.003

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

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

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

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

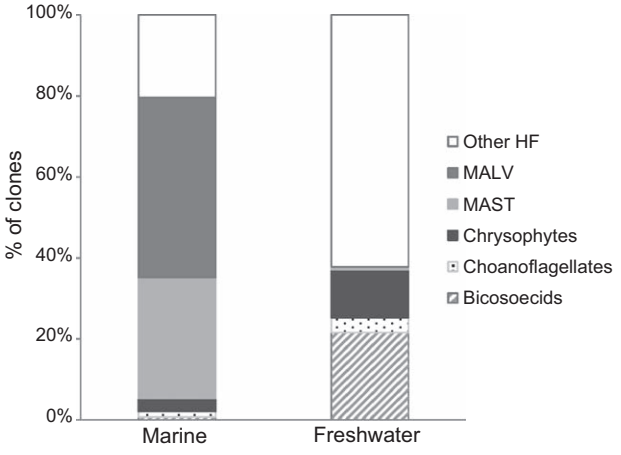


Figure 1. Relative clonal abundance of different taxonomic groups putatively forming the heterotrophic flagellate assemblages in marine and freshwater systems (data from 82 clone libraries of 18S rDNA genes; see Supplementary Table S3).

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

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

Results

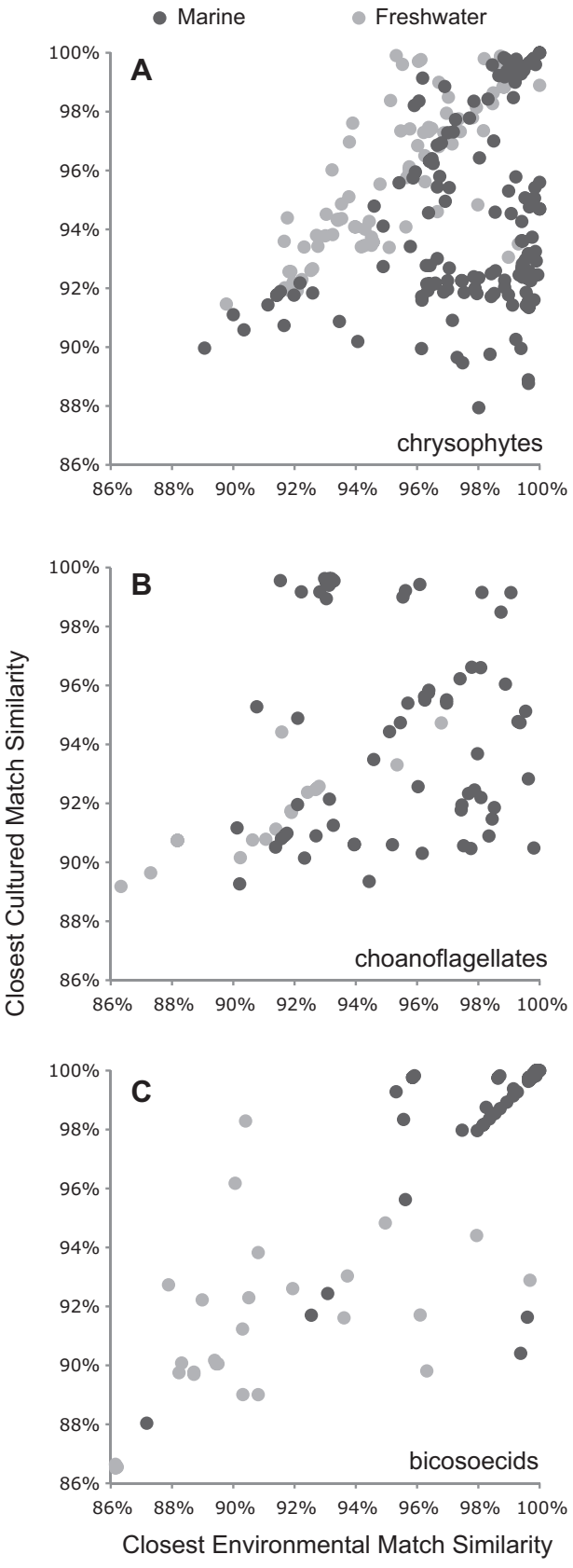
To obtain an exhaustive description of the phylogenetic diversity of chrysophytes, choanoflagellates and bicosoecids, we screened GenBank and our unpublished libraries to retrieve all sequences from these groups obtained in marine and freshwater molecular surveys. The dataset inspected included 292 environmental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

Relative Clonal Abundance in Environmental Surveys

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

Novelty of Environmental Sequences

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



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

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

Phylogenetic Trees and New Clades

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

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

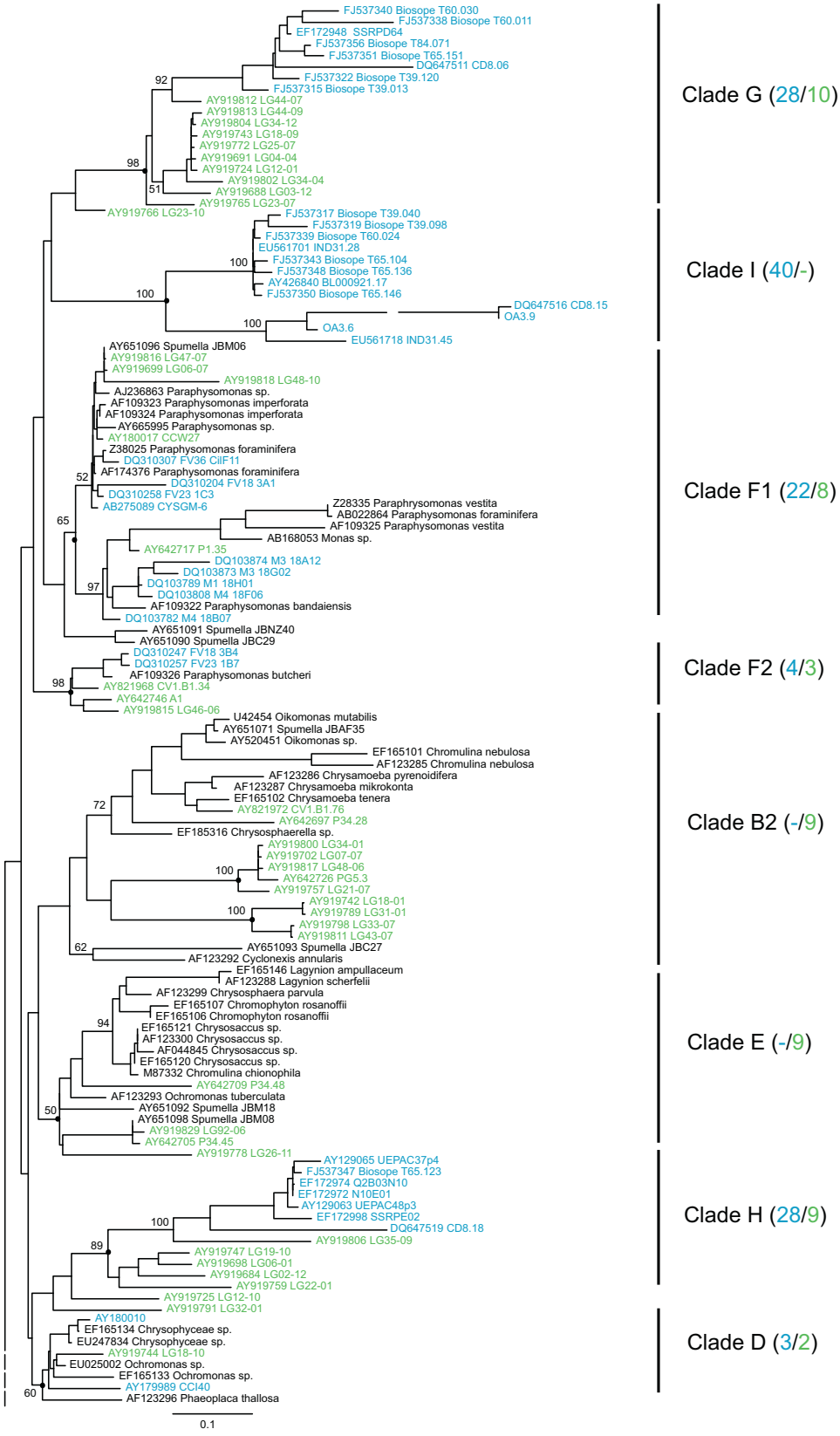

FJ537340 Biospe T60.030
FJ537338 Biospe T60.011
EF172948 SSRPD64
FJ537356 Biospe T84.071
FJ537351 Biospe T65.151
DQ647511 CD8.06
FJ537322 Biospe T39.120
FJ537315 Biospe T39.013
AY919812 LG44-07
AY919813 LG44-09
AY919804 LG34-12
AY919743 LG18-09
AY919772 LG25-07
AY919691 LG04-04
AY919724 LG12-01
AY919802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

Figure 3.

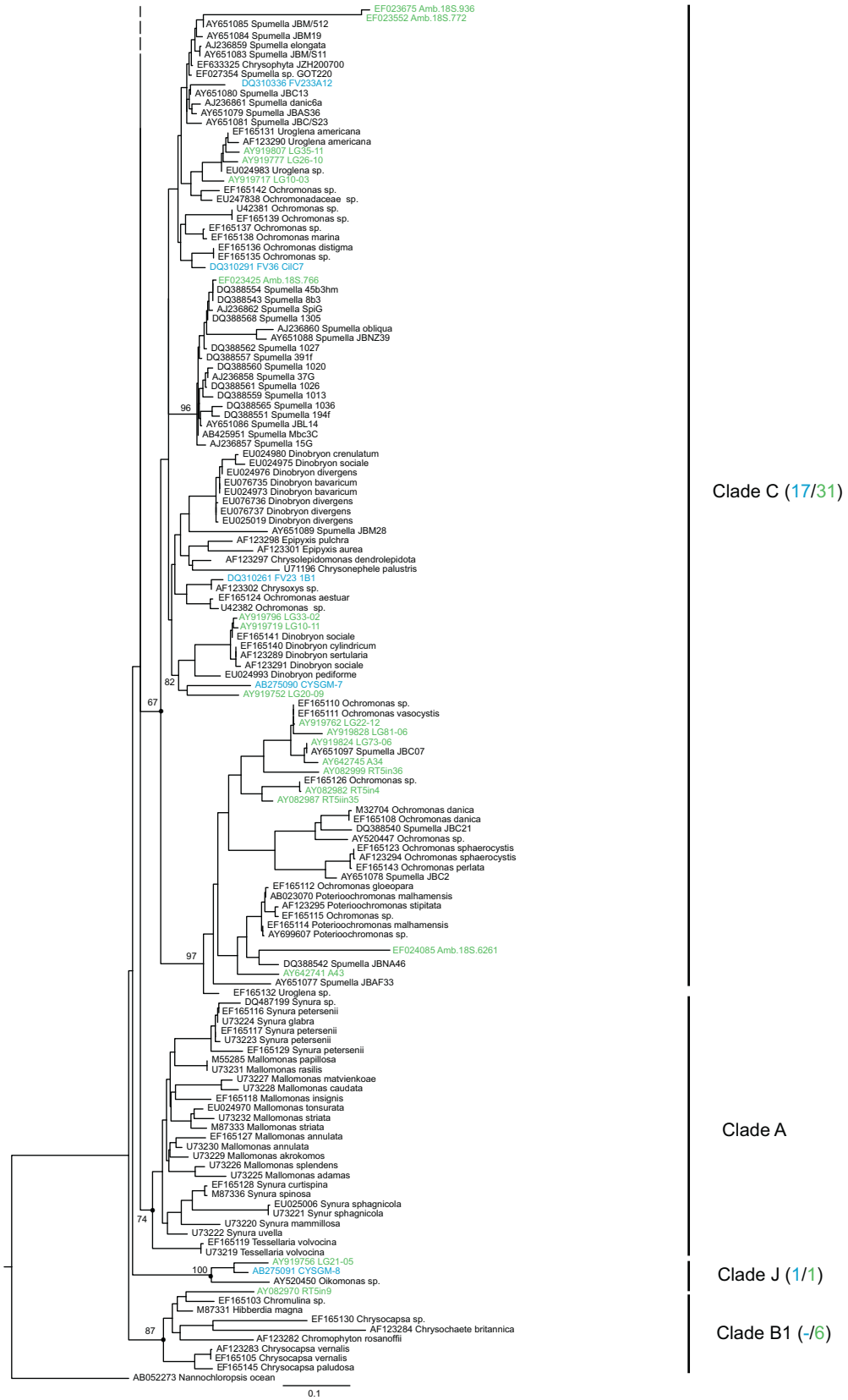


Figure 3. (*Continued*).

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

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

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

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

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

Discussion

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

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

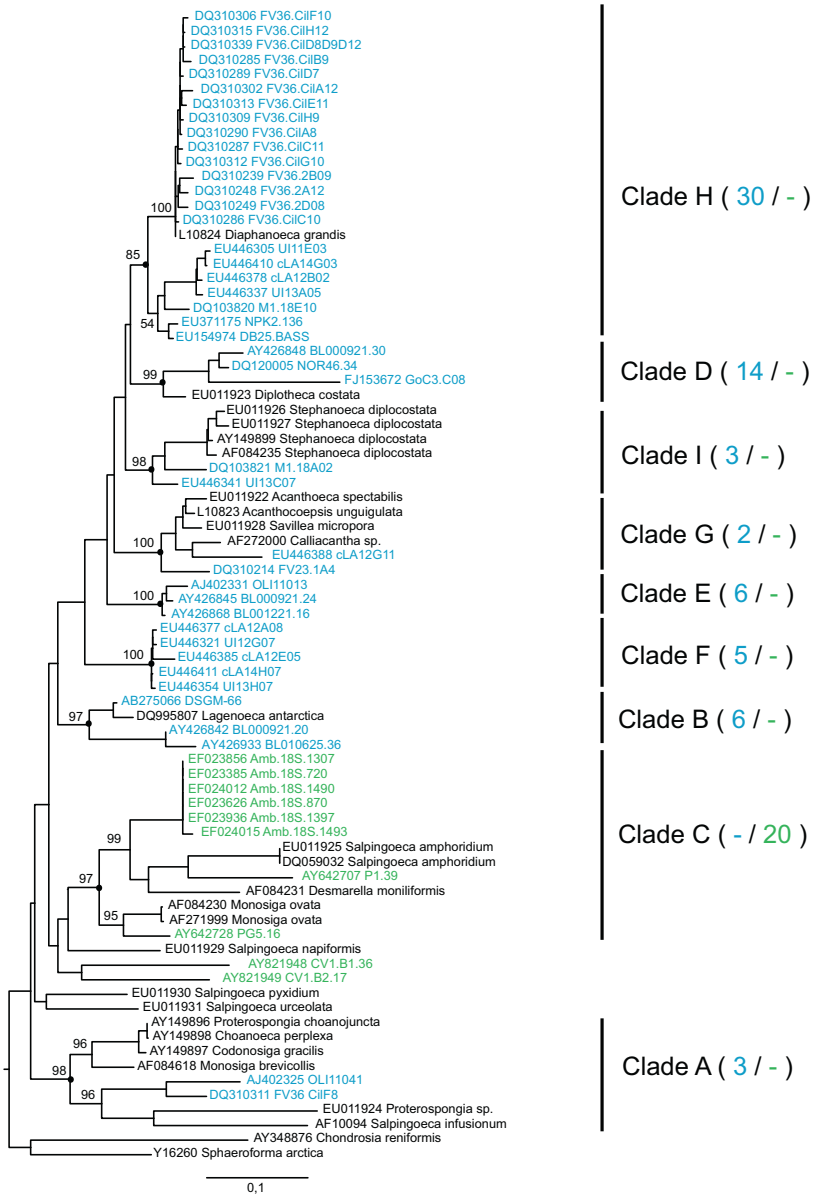


Figure 4. Maximum Likelihood phylogenetic tree of choanoflagellates constructed with 79 complete 18S rDNA sequences (1428 informative positions). Legend as in Figure 3.

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

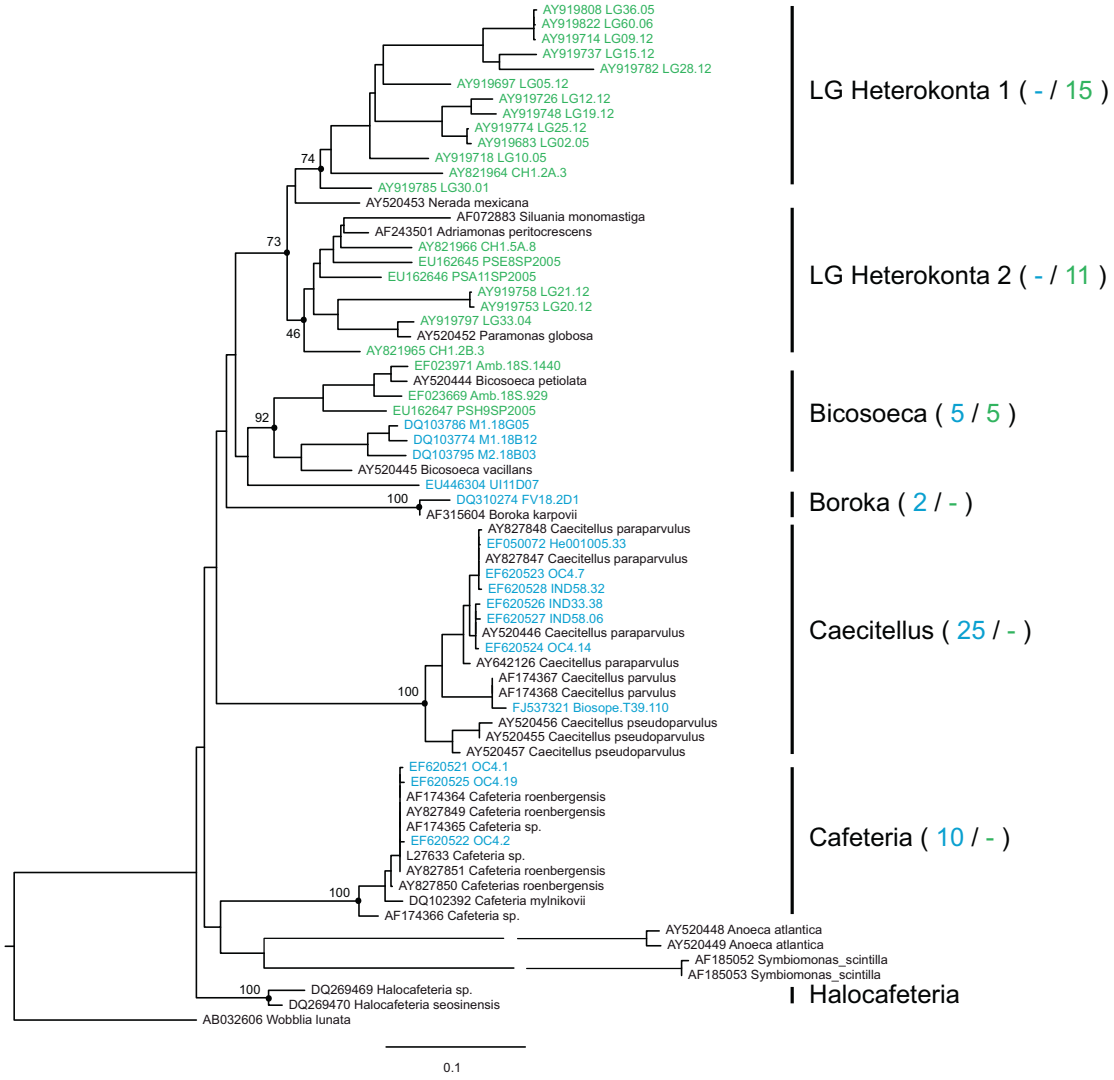


Figure 5. Maximum Likelihood phylogenetic tree of bicosoecids constructed with 66 complete 18S rDNA sequences (1485 informative positions). Legend as in Figure 3.

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

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

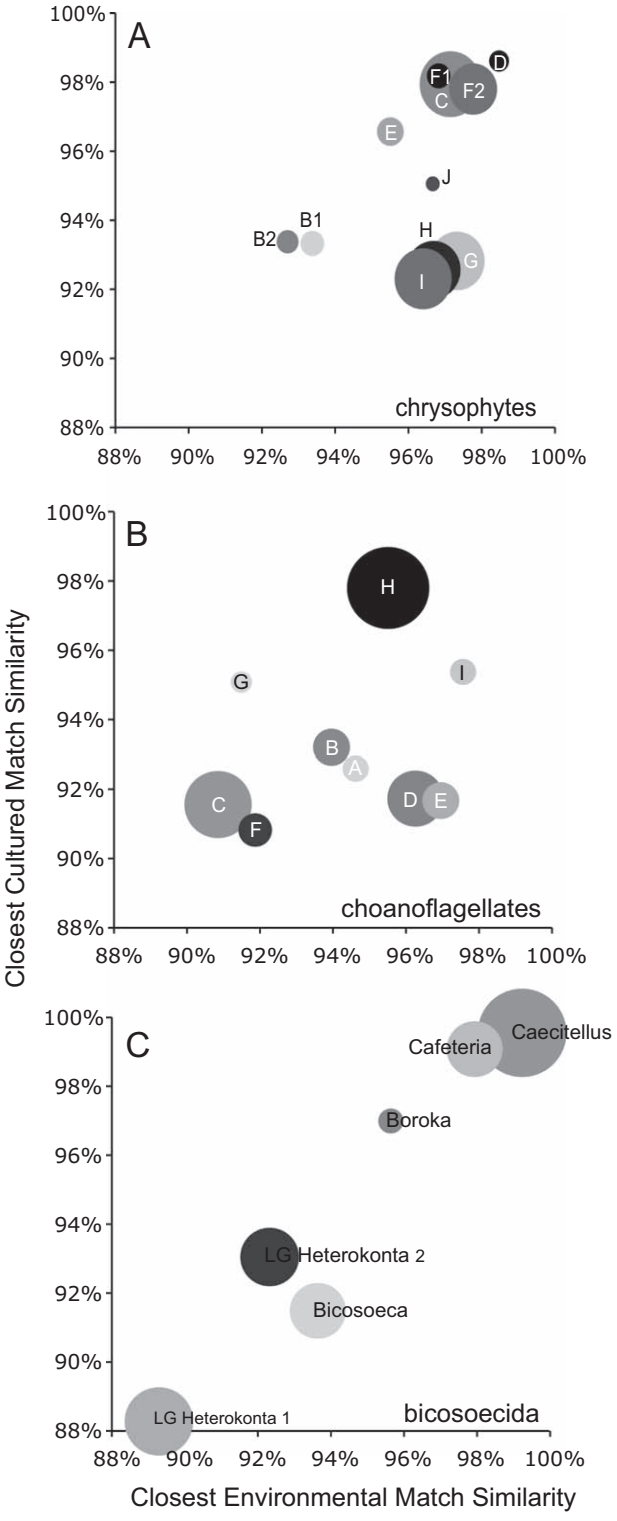


Figure 6. Novelty pattern derived from each described clade within chrysophytes (**A**), choanoflagellates (**B**) and bicosoecids (**C**). Dots representing the novelty of the clades (average similarity against CEM and CCM for all environmental sequences within the clade) have a size proportional to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

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

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

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

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

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

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

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

Methods

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

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

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

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

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

Acknowledgements

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

Appendix A. Supplementary data

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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