



## Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review

Bente Edvardsen<sup>1\*</sup>, Eianne Sirnæs Egge<sup>1</sup> and Daniel Vaulot<sup>2</sup>

\* Corresponding author: bente.edvardsen@ibv.uio.no

<sup>1</sup> University of Oslo, Department of Biosciences, P.O. Box 1066 Blindern, 0316 Oslo, Norway

<sup>2</sup> Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7144, Station Biologique, Place Georges Teissier, 29680 Roscoff, France

With 2 figures and 2 tables in the text and an electronic supplement

**Abstract:** Microalgae of the division Haptophyta are a major component of the marine nanoplankton and present in all seas. They are important primary producers and grazers of picoplankton in the ocean, and their metabolic products can have an impact on global climate. Due to their small size, species are difficult to distinguish by microscopy, and knowledge on their diversity and distribution is incomplete. Environmental sequencing studies have revealed a high marine protist diversity. We review the current knowledge on diversity and distribution of haptophytes revealed by Sanger sequencing of clone libraries (environmental sequencing) and by high-throughput sequencing of amplicons (metabarcoding). We also discuss the methodology used. Finally, we provide a curated haptophyte reference 18S rRNA-gene database for future taxonomic assignment of environmental sequences and metabarcoding reads.

**Keywords:** clone libraries; diversity; distribution; ecology; environmental sequencing; Haptophyta; high-throughput sequencing; novel lineages; metabarcoding; reference database

### Introduction

Haptophytes are mostly single-celled nanophytoplankton (usually 2–30 µm) and have a worldwide distribution. They are one of the major groups of primary producers in the ocean, together with cyanobacteria, diatoms, dinoflagellates and prasinophytes. They have also been found to be important grazers on picoplankton in the open ocean (e.g. Frias-Lopez et al. 2009; Unrein et al. 2013). Some haptophytes, such as members of the genera *Emiliania*, *Gephyrocapsa*, *Phaeocystis*, *Chrysochromulina* and *Prymnesium*, form extensive blooms that may affect the global carbon balance and possibly climate forcing, or cause fish kills with ecological and economical impact (Jordan & Chamberlain 1997; Edvardsen & Imai 2006). However, their small cell size means that they are difficult to study. Species identification often requires electron microscopy and a high degree of taxonomic expertise. This is especially true for non-calcifying haptophytes, whereas coccolithophorids have calcified scales (coccoliths), which are easier to preserve and observe under the optical and scanning electron microscopes. Therefore knowledge on haptophyte diversity, distribution and abundances at the species level remains fragmentary. The effect of environmental factors and climate change on

populations of this important phytoplankton group is also to a large extent unclear. However, the use of molecular methods, such as environmental sequencing and metabarcoding has revealed large haptophyte diversity in the ocean that was previously unsuspected. By environmental sequencing we hereafter mean PCR amplification of marker genes (amplicons) in natural plankton samples followed by clone library construction and Sanger sequencing of specific clones. In the metabarcoding approach, the clone library construction is omitted and sequencing of amplicons is performed by high-throughput sequencing (HTS, also called next generation sequencing or NGS). Metabarcoding with the 454 Life Sciences technology is also called pyrosequencing. In contrast, metagenomics refers to whole-genome sequencing of environmental samples.

In genetic surveys of plankton communities, different size fractions of the community are often separated by filtration in order to avoid the dominance by bigger organisms with larger genomes and high number of gene copies per genome (Zhu et al. 2005). These studies have revealed high, unexpected haptophyte diversity within the pico- (0.2–2 µm) and nano-plankton (2–20 µm) size ranges. Most of these sequences do not have a close similarity to cultured species (Unrein et al. 2013; Egge et al. 2015a). They may represent

species that have been formerly described, but not yet cultured or for which DNA sequences are not yet available, or correspond to novel species and lineages at taxonomic levels from genus to class. Only a few described haptophyte species are known to be of picoplankton size (Vault et al. 2008; Worden & Not 2008), and most of the pico-haptophyte diversity is expected to represent novel taxa, or alternatively a picoplanktonic life cycle stage of a known, larger haptophyte.

In this paper, we review current knowledge of diversity and distribution of haptophytes obtained from environmental sequencing and metabarcoding studies, and give an overview of these studies and discuss the methodology used. We also propose standardised names for major clades without cultured representatives based on a global haptophyte 18S rRNA gene phylogeny. We have produced a curated haptophyte 18S rRNA-gene reference database comprising 971 sequences, mostly > 800 bp and retrieved from the Protist Ribosomal Reference database (PR<sup>2</sup>, Guillou et al. 2013), for taxonomic assignment of environmental sequences and HTS reads. We envisage that this will facilitate and promote future molecular diversity surveys of haptophytes and improve our understanding of their distribution patterns and ecological role.

## Morphology and taxonomy of Haptophyta

Division Haptophyta comprises at present ca. 312 species, 80 genera, 6 orders and 2 classes (see below, Jordan et al. 2004; Edvardsen and Medlin 2007; Edvardsen et al. unpubl.). It is a monophyletic group not closely related to any other organisms. It has been proposed to belong to the supergroup Hacrobia together with cryptophytes and some heterotrophic groups (Okamoto et al. 2009; Cavalier-Smith et al. 2015). Most species are marine, but 12 described species have been recorded from freshwater (Preisig 2002). Most haptophytes are unicellular, planktonic flagellates, but coccoid, amoeboid, palmelloid, filamentous, colonial and benthic forms also occur (Hibberd 1980; Fig. 1). Most are phototrophic, but some are heterotrophic and many are known to be mixotrophic (e.g. Jones et al. 1993, 1994; Unrein et al. 2013). In most species, at least one stage in the life cycle has two smooth flagella and between them a third appendage, the haptonema, a unique organelle that can be long and curling, to short and stiff or even vestigial. It can be used to attach to a surface or for food handling (Kawachi et al. 1991). Cells are typically covered by one to several layers of organic scales, and coccolithophorids also have calcified scales called coccoliths. Holococcolithophores are covered by simple holococcoliths that may be produced on the external cell surface, whereas heterococcolithophores have elaborate heterococcoliths produced in intracellular compartments. A third group of calcified scales are the nannoliths. Species identification within Haptophyta

is largely based on scale morphology (Leadbeater 1994). Haptophytes generally possess 1–2 yellow-brown chloroplasts, with three thylakoids in the lamellae and no girdle lamella. Haptophytes may have a heteromorphic haploid-diploid life cycle, where both the haploid and diploid stages are capable of vegetative cell division. The holococcolithophorid species (about 50) are now thought to be the haploid stage in a haplo-diploid life cycle with a diploid heterococcolithophorid, and combination cells of these also occur (Geisen et al. 2004; Houdan et al. 2004; Billard & Inouye 2004). The complete sexual life cycle has been observed in only a few species and knowledge on haptophyte life cycles is still fragmentary. The evolution of Haptophyta was recently reviewed by (Liu et al. 2010), and that of the coccolithophorids by de Vargas et al. (2007). The molecular phylogeny of Haptophyta and morphology of major clades were reviewed by Sáez et al. (2004) and Edvardsen and Medlin (2007). A haptophyte 18S rRNA gene phylogeny was presented recently (Egge et al. 2015a, 2015b) including 76 cultured species, and environmental sequences representing most (all but HAP1) major clades without a cultured representative. Here we have reconstructed a global phylogeny based on all 971 sequences in the Haptophyta database showing all major clades (Fig. 2). Fig. S1 shows the resulting tree without collapsed clades.

## Molecular markers for phylogeny and diversity

The most commonly used gene marker to infer phylogeny within Haptophyta is the nuclear-encoded SSU (18S) rRNA (e.g. Medlin et al. 1997; Simon et al. 1997; Edvardsen et al. 2000; Edvardsen et al. 2011; Bendif et al. 2011; Bendif et al. 2013). It has both conserved regions enabling a reliable alignment and phylogenetic comparison at higher taxonomic levels, and variable regions enabling differentiation down to species level. However, some closely related species have identical sequences such as *Emiliania huxleyi* and *Gephyrocapsa oceanica*, or *Prymnesium pinnenarii* and *P. simplex* (Edvardsen et al. 2011; Bendif et al. 2011). Like for most protist groups, the 18S rRNA is the gene for which most reference sequences of haptophyte species are available (Guillou et al. 2012; Table 1). The phylogeny inferred from the 18S rRNA gene is generally congruent with the taxonomy based on morphology (Edvardsen et al. 2000). Partial 28S rRNA gene sequences embracing c. 770–1000 bp of the D1 and D2 regions have hypervariable positions that usually have to be removed for a reliable alignment, resulting in few remaining informative sites and a partly unresolved phylogeny (e.g. Edvardsen et al. 2011). The ribosomal internal transcribed spacers ITS1 and ITS2 have been used to segregate populations within a species and closely related species (Medlin et al. 2000 and references therein; Bendif et al. 2014), but the high variation in length

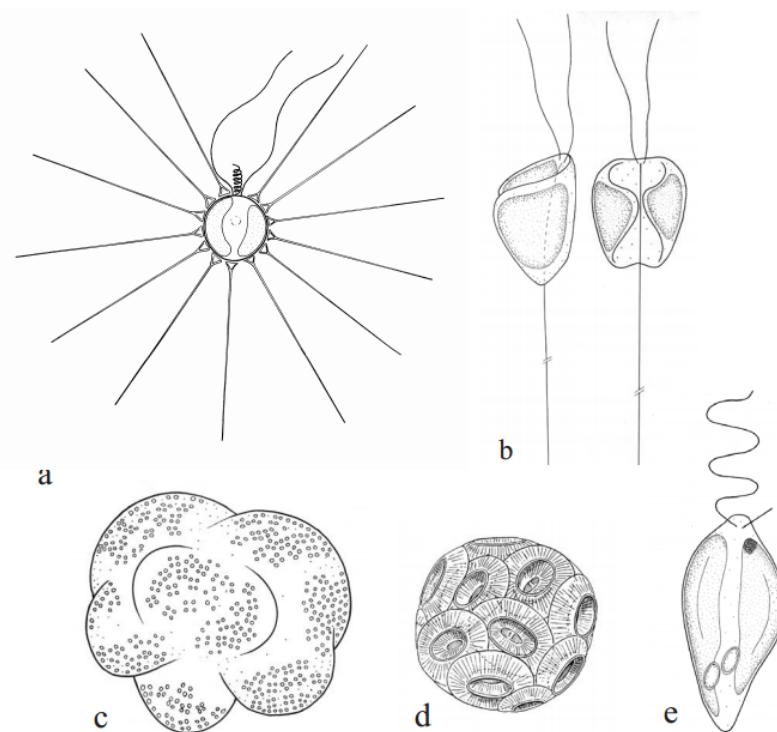


Fig. 1. Haptophytes: (a) *Haptolina hirta*, (b) *Chrysochromulina alifera*, (c) *Phaeocystis pouchetii*, (d) *Coccolithus pelagicus*, and (e) *Pavlova gyrans*. (From Throndsen, J., Hasle, G.R. & Tangen K., Phytoplankton of Norwegian Coastal Waters, Almater Forlag AS, Oslo, 2007, with permission.)

and nucleotide sequence makes it difficult to align reliably (Medlin et al. 2000). Some haptophyte phylogenies have also been inferred from the plastid-encoded 16S rRNA (e.g. Edvardsen et al. 2011) or *rbcL* genes (e.g. Fujiwara et al. 2001; Yoshida et al. 2006), but these genes suffer from the same limitation as the nuclear 18S rRNA gene in terms of resolution. The mitochondrial encoded *cox1-3* genes have in general a higher variation than coding rRNA regions and have been used as barcode to delineate sub-clades of *Emiliania huxleyi* and *Gephyrocapsa* spp. (Bendif et al. 2014). Concatenated phylogenies combining two or more genes in the same alignment may improve the robustness and/or resolution of the resulting phylogenetic tree (Edvardsen et al. 2011; Bendif et al. 2013, 2014). Some of the genes useful for phylogenetic inference of haptophytes have also been used as molecular markers (barcodes) for identification and detection of taxa in environmental samples, especially the ribosomal RNA genes coding for the 18, 28 and 16S sub-units. Cultured haptophyte species for which both the 18S and partial 28S (D1-D2 region) sequences have been determined generally show higher interspecific variation in 28S than 18S (Table S5 in Liu et al. 2009; Egge et al. unpubl.). The 28S may thus reveal higher diversity of haptophytes than 18S rRNA, and could constitute a powerful barcode in environmental sequencing. However, the number of reference sequences from

described and cultured species is lower than for 18S, which makes taxonomic assignment more difficult (Table 1).

## Reference sequences

For reconstructing molecular phylogenies and for molecular identification, reliable reference sequences are crucial. Of the 312 described and currently valid haptophyte species (Jordan et al. 2004; Edvardsen et al. unpubl.), 96 (31%) have been cultured and their sequences determined for 18S compared to only 76 (25%) for 28S (Table 1). Within Pavloales, 85% (11 of 13 spp.) have 18S and 28S sequences determined, whereas within Prymnesiales, more reference sequences are available for 18S than 28S (38 spp. (48%) and 28 spp. (35%), respectively). Within Calcihaptophycidae, the number of available reference species is 41 for 18S and 32 for 28S. The plastid 16S rRNA gene has been sequenced for 52 haptophyte species and data are available from the PhytoREF database (<http://phytoref.org>, Decelle et al. 2015).

## Haptophyte 18S rRNA gene database

We have built a Haptophyta 18S rRNA gene reference database (Table S1) with curated and updated taxonomy,

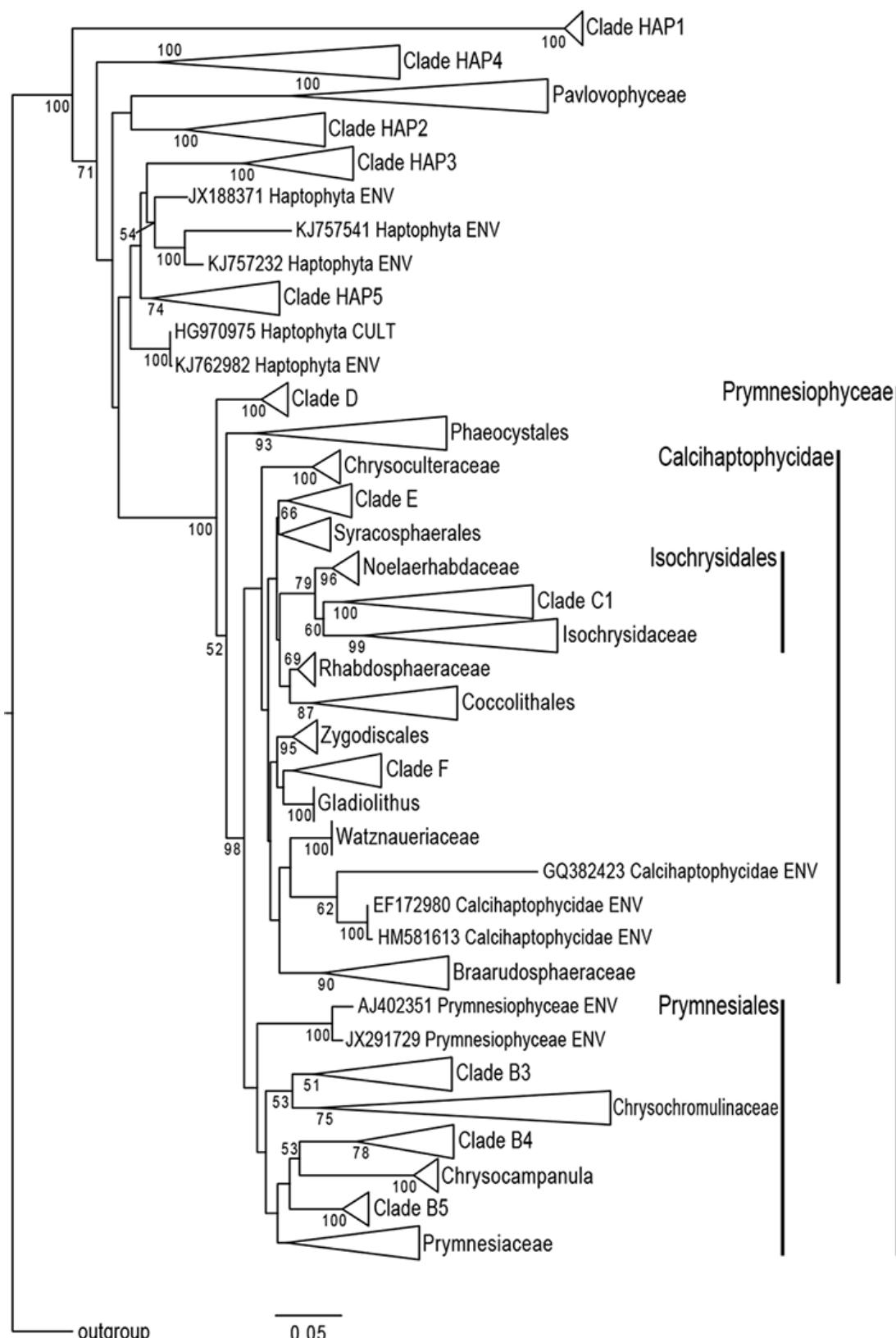


Fig. 2. Maximum-likelihood tree (RAxML using model GTRCAT, see File S1 for explanations) based on all haptophyte 18S rRNA gene sequences (971) from Table S1, collapsed to show all major clades. Five Hacrobia sequences (AF534709, AJ564771, JX988758, KJ762967, AF508268) were used as outgroup, and were pruned from the tree. Support values at the nodes represent bootstrap-values from 100 replications, and scale bar number represent substitutions/site.

comprising all sequences assigned to Haptophyta from the PR<sup>2</sup> database (Guillou et al. 2012) corresponding to the GenBank release v. 203 (October 2014). These sequences are longer than 800 bp. We have also included 27 sequences shorter than 800 bp that were phylogenetically placed and used as references by Egge et al. (2015a). Thirty-two chimeras have been identified and removed (see Table S1). The database contains 971 sequences: 451 from cultured strains and 520 from marine and freshwater environmental samples. We have determined the taxonomy of all sequences by phylogenetic analyses (MAFFT alignment and maximum likelihood (ML) analyses). Sequence metadata (sequence length, strain or clone code, name of organism in GenBank, locality, year and author of submission) were downloaded from NCBI and added to the database. Finally, the taxonomy was updated according to the latest taxonomic revisions (Edvardsen et al. 2000, 2011; Jordan et al. 2004; Bendif et al. 2011, 2013; Andersen et al. 2014, 2015b). A detailed description of the construction of the database is provided in File S1. The database is available both as an Excel-file including all information mentioned (Table S1) and as files for use with Qiime or mothur for taxonomic assignation of high-throughput reads (Files S2, S3). In addition, we also provide an alignment (File S4) for all sequences in the database with addition of five taxa as outgroup (*Chilomonas paramecium*, *Chlamydaster sternis*, *Chroomonas mesostigmatica*, *Leucocryptos marina*, *Picomonas judraskeda* and *Telonema subtile*), and a complete RAxML tree (Fig. S1).

We compiled information from studies including haptophyte environmental sequences (papers containing more than 3 sequences) focusing on 18S rRNA gene clone librar-

ies (Tables 2A, B). Table S2 compiles some of the commonly used PCR primers in these studies.

## Diversity revealed by environmental clone libraries

The first paper to report on haptophyte diversity explored by environmental sequencing used universal eukaryotic 18S rRNA gene primers to amplify picoplankton (< 2–3 µm) from filtered sea water samples in combination with clone library construction (Moon-van der Staay et al. 2000, 2001). They found 17 haptophyte OTUs (Operational Taxonomic Units, an estimate for species) from the pico-plankton fraction of the equatorial Pacific Ocean, of which 14 represented taxa without a cultured and sequenced representative. In subsequent years, picoeukaryote diversity was investigated using environmental sequencing of the 18S rRNA gene with universal primers in the Sargasso Sea (Not et al. 2007a), Pacific Ocean near Hawaii (Frias-Lopez et al. 2009), Mediterranean Sea (Not et al. 2009), and SE Pacific Ocean (Shi et al. 2009). These studies uncovered high picoeukaryote diversity and numerous haptophyte OTUs were found representing members of known genera such as *Chrysochromulina*, *Emiliania/Gephyrocapsa* and *Phaeocystis*, and many uncultured taxa or novel lineages within Prymnesiophyceae. Edgcomb et al. (2011) investigated the protist community composition in an oxic-anoxic transition zone in the Caribbean Sea using massive sequencing of clone libraries with several primer pairs. Their study revealed a high diversity of protists (from pico- to micro-plankton), representing over 6,000 species,

Table 1. Number of species described and with available nuclear 18S, 28S and plastid 16S rRNA gene sequence available in the various taxa of Haptophyta.

Taxa	Described species	Descr. spp. 18S rDNA	% with 18S	Descr. spp. 28S rDNA	% with 28S	Descr. spp. 16S rDNA	% with 16S
Pavlovales	13	11	85	11	85	5	38
Coccolithales	26	22	85	19	73	8	31
Isochrysidales	17	9	53	8	47	7	41
Phaeocystales	10	6	60	5	50	6	60
Prymnesiales: Chrysochromulinaceae	47	11	23	8	17	9	19
Prymnesiales: Prymnesiaceae	32	27	84	20	62	14	44
Syracospaerales	58	2	3	3	5	1	2
cf. Syracuseales, incertae sedis	36	0	0	0	0	0	0
Zygodiscales	10	2	20	3	30	1	9
Holococcolithophorids	48	2	4	0	0	0	0
Nannolith-bearing, incertae sedis	13	2	15	0	0	1	10
Chrysoculteraceae	1	1	100	0	0	0	0
Watznaueriaceae	1	1	100	0	0	0	0
SUM	312	96	31	76	25	52	17

**Table 2A.** Studies including haptophyte environmental sequences ( $n > 3$ ) in our database and information on locality, size fraction, gene, and number of haptophyte OTUs (similarity level varies). M = marine, B = brackish, F = freshwater, FCM = flow cytometry, nd = no data.

Reference	Locality	M/B/F	Size fraction ( $\mu\text{m}$ )	Gene	# OTUs
Moon-van der Staay et al. 2000	Equatorial Pacific Ocean	M	< 3	18S	13
Moon-van der Staay et al. 2001	Equatorial Pacific Ocean	M	< 3	18S	4
Not et al. 2007	Sargasso Sea, Atlantic Ocean	M	< 2	18S	5
Frias-Lopez et al. 2009	Hawaii, Pacific Ocean	M	grazers on picocyanobact.	18S	19
Not et al. 2009	Mediterranean Sea, France	M	< 3	18S	10
Shi et al. 2009	SE Pacific Ocean	M	< 3	18S	20
Behnke et al. 2010	Framvaren fjord, Norway, oxic-anoxic	M	tot	18S	6
Cuvelier et al. 2010	Florida Strait; Sargasso Sea	M	FCM-sorted pico	18S	35
Marie et al. 2010	English Channel, Roscoff, France	M	FCM-sorted pico	18S	12
Masquelier unpubl.	Lake, Luxembourg	F	> 0.65	18S	4
Edgcomb et al. 2011	Caribbean Sea, Venezuela, oxic-anoxic	M	Total, > 0.45	18S	19
Fujimoto et al. unpubl.	Gunma Lake, Japan	F	nd	18S	6
Lépère et al. 2011	SE Pacific Ocean	M	FCM-sorted pico	18S	24
Balzano et al. 2012	Beaufort Sea, Arctic, Canada	M	< 3	18S	10
Domaizon et al. 2012	Thau lagoon, Mediterranean, France	M	Total	18S	9
Fujimoto et al. unpubl.	Gunma Lake, Japan	F	nd	18S	8
Koid et al. 2012	NW Pacific Ocean	M	20–3, 3–0.8, 0.8–0.1	18S	12
Orsi et al. 2012	Vancouver, Saanich Inlet, N Pacific, Canada	M	2.7–0.2	18S	59
Terrado et al. 2011, 2013	Arctic region, Baffin Bay, Canada	M	0.2–3, 3–50	18S	13
Thompson et al. 2012	ALOHA station, N Pacific Ocean	M	FCM sorted pico	18S	7
Weber et al. 2012	Baltic Sea, Germany	B	200–3, 3–0.2	18S	4
Dasilva et al. 2013	N Atlantic, AZMP station, Canada	M	0.2–3	18S	9
Simon et al. 2013	Med. Sea, France; Marmara Sea, Turkey; N Atlantic	F+M	0.22–5, 5–30, 0.22–60	18S	105
Lie et al. 2014	SPOT station N Pacific; Gulf Stream, E Pacific Rise; Arctic	M	> 0.7–80 (GF/F)	18S	18
Wu et al. 2014	S China Sea, W Pacific Ocean	M	< 3	18S	30

but only 19 OTUs (Tables 2A, S1, 0.3% of the OTUs) and 0.5% of the sequences were assigned to Haptophyta. In contrast, Orsi et al. (2012) studied picoeukaryote diversity in oxic-anoxic waters off the west coast of Canada by clone libraries and universal 18S rRNA gene primers and recovered 59 haptophyte OTUs (9%), most assigned to the genus *Chrysochromulina* (Table S1) despite the fact that the forward primer used (Euk515F) has one mismatch to all sequences from cultures in our alignment (an inserted A). Not et al. (2009) suggested that analysis of RNA instead of DNA may provide a better image of the protist community since DNA may not correspond to active organisms. It may also reduce the bias due to the differences in DNA copy number per cell among taxonomic groups (Not et al. 2009). Another sampling approach is to sort by flow cytometry specific photosynthetic eukaryote populations based on their size and chlorophyll signal. This strategy allows better recovery of haptophyte sequences, including sequences not observed in filtered samples (Shi et al. 2009; Marie et al. 2010; Cuvelier et al. 2010). The low recovery of haptophyte sequences from environmental 18S rRNA clone libraries can also be overcome by the use of haptophyte-specific or -biased 18S and 28S primers (Liu et al. 2009; Shalchian-Tabrizi et al. 2011; Bittner et al. 2013; Egge et al. 2015a, b; Simon et al. 2013; Young et al. 2014; Table S2). For example, using haptophyte-specific 28S (D1-D2 domain) rRNA gene primers, Liu et al. (2009) revealed an extreme diversity of pico-planktonic (0.2–3 µm) non-calcifying haptophytes in 5 samples from Arctic subpolar and subtropical (South Indian Ocean) oceanic waters (Table 2B). From c. 1000 clones, they recovered 674 OTUs (at 100% similarity). Phylogenetic analyses showed that

the majority of the sequences clustered within the genus *Chrysochromulina*. Phaeocystales, Prymnesiaceae and Calcihaptophycidae were also represented among the novel pico-haptophyte sequences. As mentioned above, to date, more haptophyte species have been sequenced for the 18S compared to the 28S rRNA gene (Table 1), in particular for Prymnesiales. Thus it is possible that some of the novel sequences in this study represent described species without available 28S sequence.

Targeting the plastid-encoded 16S rRNA gene instead of the nuclear rRNA genes allows biased amplification of the phytoplankton (Fuller et al. 2006; McDonald et al. 2007; Shi et al. 2011), filtering out heterotrophic protists that may dominate in marine pico- and nanoplankton metabarcoding data sets (e.g. de Vargas et al. 2015). These studies revealed a much higher contribution of haptophytes to photosynthetic sequences than estimated from 18S rRNA studies with universal primers. For example Fuller et al. (2006) found at two stations in the Arabian Sea 17 and 37% of the plastid 16S rRNA clones belonging to Prymnesiophyceae. However, this contribution strongly depends on the primers used as shown on sorted samples from the SE Pacific Ocean (Shi et al. 2011).

## Diversity revealed by metabarcoding

Environmental sequencing of clone libraries usually underestimates the diversity compared to metabarcoding with HTS (e.g. Egge et al. 2013), where a large sequencing depth can be obtained enabling detection of rare taxa. Some HTS technologies suffer from high error rates that demand

**Table 2B.** Selected studies including haptophyte environmental sequences ( $n > 3$ ) not in our database and information on locality, size fraction, gene, and number of haptophyte OTUs. See Tables 2A for abbreviations. HTS = high throughput sequencing.

Reference	Locality	M/B/F	Size fraction (µm)	HTS	Gene	# OTUs
McDonald et al. 2007	Gulf of Naples, Mediterranean Sea, Italy	M	< 5		16S	114
Liu et al. 2009	Sub-Arctic, S Indian Ocean	M	0.2–3		28S	674
Nersveen 2011	Southern Ocean, Atlantic	M	nano, pico		28S	34
Shalchian-Tabrizi et al. 2011	Finsevatn, Norway	F	> 1 (GF/C)	X	18S	8
Shi et al. 2011	South Pacific Ocean	M	FCM-sorted pico		16S	29
Bittner et al. 2013	Gulf of Naples, Mediterranean Sea, Italy	M	0.8–3, 3–20	X	28S	627
Kiliias et al. 2013	Arctic Atlantic Ocean, Fram strait	M	micro	X	18S	> 53
Kiliias et al. 2014	Arctic Atlantic Ocean, Fram strait	M	pico	X	18S	nd
Lie et al. 2014	SPOT station N Pacific; Gulf Stream, E Pacific Rise; Arctic	M	> 0.7–80 (GF/F)	X	18S	nd
Taylor & Cunliffe 2014	English Channel, N Atlantic	M	total	X	18S	> 260
Thiele et al. 2014	Southern Ocean, Atlantic	M	nano, pico	X	18S	nd
Egge et al. 2015a	Outer Oslofjord, Skagerrak, N Atlantic, Norway	M+B	< 3 and 3–45	X	18S	156
Young et al. 2014	Atlantic Ocean, Pacific Ocean, Mediterranean Sea	M	> 5–10		28S	> 74

thorough bioinformatics treatment. Egge et al. (2013) tested various bioinformatics pipelines on a haptophyte mock community sequenced by 454 and found that some pipelines overestimated the species richness more than 100 times clustering at 99% similarity level. The treatment that best estimated the actual species richness included initial filtering in mothur (Schloss et al. 2009) followed by denoising by AmpliconNoise and chimera check and removal by Perseus (Quince et al. 2011). As different studies use different pipelines, some derived variables such as OTU richness can be difficult to compare.

Amplification of the V4 region of the 18S rRNA gene with eukaryote-general (Stoeck et al. 2010) or haptophyte-specific primers (Egge et al. 2013) gives c. 400 bp. fragments suitable for 454 HTS (Table S2) and also more recently for Illumina. Reads can usually be taxonomically assigned to species level using a clustering level of 99–99.5% (Egge et al. 2015a). Egge et al. (2015a, 2015b) explored the haptophyte diversity in the outer Oslofjorden during two years with monthly sampling using haptophyte specific primers targeting the 18S V4 region. From c. 400,000 reads, 156 haptophyte OTUs were obtained (99.5% similarity). Most OTUs (84%) represented uncultured and/or not yet 18S-sequenced species, most of which were affiliated to Prymnesiales. Kilias et al. (2013, 2014) investigated protist composition in Atlantic Arctic waters using 454-metabarcoding of the 18S V4. As expected, reads of *Phaeocystis pouchetii* were abundant at all stations in the picoplankton fraction.

The V9 region of 18S rRNA gene (Amaral-Zettler et al. 2009) relies on a shorter fragment length of c. 130 bp, which is suitable for the Illumina platform, but appears to have lower taxonomic resolution compared to the longer V4 region for Haptophyta (Edvardsen unpubl.). Illumina sequencing is considerably less expensive on a per base pair basis than with 454 and was the chosen platform for 18S V9 metabarcoding of samples from the Tara Oceans expedition. De Vargas et al. (2015) assessed the eukaryotic diversity from 334 size-fractionated photic-zone plankton communities collected across tropical and temperate oceans during this circumglobal expedition and detected > 700 haptophyte OTUs.

Bittner et al. (2013) targeted haptophytes with haptophyte-specific 28S primers and 454 metabarcoding and compared pico- and nano-plankton size fractions at two depths from the Gulf of Naples as well as amplicons obtained from rDNA and rRNA/cDNA. Clustering at 97% similarity resulted in 627 OTUs. Only 1% of the OTUs could be assigned to a described and cultured species with available 28S sequence (at 97% similarity), and less than 12% clustered with reference sequences obtained previously from cloning and Sanger sequencing of environmental samples. The LSU-metabarcoding also revealed high diversity and relative abundance of Chrysochromulinaceae.

## Diversity and distribution within established clades

### Class Pavlovophyceae

Pavlovophyceae consists of one order, one family, 4 genera and 13 described species of which all but 2 (*Pavlova calcicola* and *Rebecca helicata*) have been cultured and the 18S rRNA gene sequence determined (Bendif et al. 2011, Tables 1, S1). At least 27 additional strains have been isolated into culture for which the 18S rRNA gene sequence has been determined. They represent several novel species that await formal description (Bendif et al. 2011, Table S1). Pavlovophytes have been described mostly from littoral, brackish or sometimes fresh waters, and may be common in near shore planktonic and benthic microalgal communities as well as ponds and lakes (Preisig 2002; Not et al. 2012). None of the available 18S environmental sequences assigned to Pavlovophyceae originate from open marine waters, confirming their near-shore or freshwater distribution. Environmental 18S sequencing has revealed only 5 OTUs that could not be assigned to a cultured and sequenced species ( $\geq 99\%$  similarity). All are more than 95% similar to a known and cultured species. Metabarcoding seasonal data from outer Oslofjorden revealed 3 OTUs without match to a cultured strain or available environmental sequence (Egge et al. 2015a). They formed a sister group to *Diacronema*, suggesting that they represent novel species of *Diacronema* or a novel genus. A fourth OTU, present in May only, was very similar (1–2 bp difference) to environmental sequences (e.g. JX680423) from a lake in France (Simon et al. 2013). The low number of OTUs without a cultured representative suggests that this group is easy to cultivate in standard nutrient rich algal medium and may be mainly photoautotrophic.

### Class Prymnesiophyceae

#### Order Phaeocystales

This order presently consists of one family and one genus, *Phaeocystis*, with 10 current species that are all marine, of which 7 are well characterised by EM and 6 for which 18S rRNA has been sequenced (Table 1, S1). Three of these, *P. antarctica*, *P. globosa* and *P. pouchetii* produce colonies as part of their life cycle and may form extensive blooms with large biogeochemical impact. *Phaeocystis cordata*, *P. jahnii* and *P. scrobiculata* are only known as solitary flagellates (Medlin & Zingone 2007) and *P. rex* as non-motile or flagellated single cells (Andersen et al. 2015a). In addition, at least 10 *Phaeocystis* sp. strains with available 18S rRNA gene sequences are included in our database (Table S1), and some may represent novel *Phaeocystis* species (e.g. Medlin & Zingone 2007). Decelle et al. (2012) found by DNA sequencing that symbionts of marine planktonic acantharians (Radiolaria) belonged to *Phaeocystis*, either to well-known, abundant and free-living species such as *P. globosa*,

*P. antarctica* and *P. cordata*, or to two novel ribotypes (Phaeo 1 and Phaeo 2, Decelle et al. 2012). Metabarcoding data from the Oslofjorden revealed an unknown diversity within this clade that may represent novel species and genera (Egge et al. 2015a). Fifteen OTUs belonged to this order, of which 4 represented known species ( $\geq 99\%$  similarity, *P. pouchetii*, *P. globosa*, *P. cordata* and *P. jahni*). Two well-supported sub-clades were identified consisting of environmental sequences only, originating from the South and North Pacific Ocean, the Florida Strait (Frias-Lopez et al. 2009; Cuvelier et al. 2010; Thompson et al. 2012), and outer Oslofjorden (Egge et al. 2015a). These, and other environmental sequences from marine open-ocean, without affinity to cultured representatives, may represent novel genera or species complexes within *Phaeocystis*.

#### Claude D

Claude D was introduced by Moon-van der Staay (2000) to embrace environmental picoplankton sequences originating from oligotrophic, equatorial Pacific waters. This well supported clade, only consisting of uncultured ribotypes, has an unstable placement and either diverges after Phaeocystales within the Prymnesiophyceae, or forms a sister clade to Phaeocystales (Fig. 2; Edvardsen et al. 2000, 2011; Moon-van der Staay et al. 2000; Egge et al. 2015a). Environmental sequences within this clade have been reported from many studies since then, in particular within the picoplankton size fraction from Equatorial, South and North Pacific Ocean, Florida strait, Caribbean Sea off Venezuela, Marmara Sea, Skagerrak, Southern Ocean suggesting a wide marine distribution (Table S1; Shi et al. 2009; Cuvelier et al. 2010; Lie et al. 2014; Edvardsen unpubl.). Claude D taxa do not appear to be strongly related to any other haptophyte taxa and may represent a novel order. In view of their early divergence, one may hypothesize that they do not have coccoliths.

#### Order Prymnesiales

Prymnesiales consists of two families, Chryschromulinaceae with one genus, *Chryschromulina*, containing 47 described species (listed by Chrétiennot-Dinet et al. 2014), and Prymnesiaceae with 6 genera and 32 described species (Edvardsen et al. 2011; Edvardsen et al. unpubl.). The taxonomy of the order has been revised recently and the phylogeny and taxonomy described in detail elsewhere (Edvardsen et al. 2011; Bendif et al. 2013). Within these families, 23% and 84%, respectively, of the described species have been cultured and the 18S rRNA gene sequence determined (Table 1). Most cultured strains have been isolated from coastal waters where Prymnesiales species may be abundant and even form blooms. Recently, metabarcoding studies have revealed high diversity within Prymnesiales without match to cultured species. In the Skagerrak, Egge et al. (2015a) found 93 OTUs affiliated to Prymnesiales (60% of all haptophyte OTUs) and only 8% of these matched cul-

tured species ( $\geq 99\%$  similarity). Some of these probably represent known, but not yet cultured species. Haptophyta OTU richness was more than twice the species richness observed by electron microscopy over the years in this well studied region (Egge et al. 2015a).

Prymnesiales is usually the order that contains the largest number of OTUs in environmental sequencing studies. In our database (Table S1) 57% of the environmental haptophyte sequences were affiliated to Prymnesiales and 43% to *Chryschromulina*. Within oceanic picoplankton a large proportion of the haptophyte 18S OTUs have been assigned to *Chryschromulina* (e.g. Cuvelier et al. 2010; Orsi et al. 2012; Simon et al. 2013; Wu et al. 2014). Surprisingly, only 3 of the described *Chryschromulina* species are in the pico-size range ( $\leq 3 \mu\text{m}$ , Vaulot et al. 2008). Many have however a size ca 4–6  $\mu\text{m}$  and may be squeezed through a 3  $\mu\text{m}$  pore size filter. Many novel picoplanktonic species of *Chryschromulina* are therefore probably waiting to be described. Egge et al. (2015a) found 6 OTUs assigned to *Chryschromulina* that were only retrieved from the picoplankton size fraction and not from the nanoplankton (3–45  $\mu\text{m}$ ). One of these was assigned to *C. rotalis*, described to be 4–6  $\mu\text{m}$  in diameter (Eikrem & Throndsen 1999). This species was found to have a haplo-diploid life cycle embracing haploid cells bearing aberrant scales that were smaller than the typical diploid cells described in the original description of this species (Edvardsen & Imai 2006; Edvardsen unpubl.). The 18S environmental sequencing and metabarcoding data support previous observations by microscopy that many species of *Chryschromulina* usually co-exist (Leadbeater 1972). Similarly, studies using the LSU rRNA gene as a marker revealed a high diversity especially within Prymnesiales and Chryschromulinaceae (Liu et al. 2009; Bittner et al. 2013). OTUs assigned to Prymnesiaceae are predominately found in neritic waters such as in the Skagerrak, Marmara Sea, or Saanich Inlet (Table S1), whereas OTUs assigned to Chryschromulinaceae are found both in open oceans and neritic waters, which is in agreement with microscopical observations (Thomsen et al. 1994; Not et al. 2012). OTUs assigned to *Chryschromulina* have been found in all seas from the Arctic in the north (e.g. Balzano et al. 2012) to the Southern Ocean (Nersveen 2011). In an environmental clone library study using haptophyte specific LSU primers Nersveen (2011) detected 29 OTUs (clustered at 99.3%) assigned to Chryschromulinaceae along a transect at 15°E between 54–69°S in the Southern Ocean. In this study, *Chryschromulina simplex*, the only OTU that matched a cultured and sequenced haptophyte species, was detected at all stations. Egge et al. (2015a) found *C. simplex* to be the most frequent *Chryschromulina* species, present in all monthly samples during two years, supporting its wide geographical and temporal distribution.

A number of environmental sequences clustering with the base of Prymnesiales form well-supported clades without cultured representatives (Fig. 2). Clades B3, B4 and B5

consist of environmental sequences only, and may represent novel families and/or genera within Prymnesiales.

*Clade B3* was introduced by Simon et al. (2013) to include 7 environmental sequences from the Sea of Marmara, the Sargasso Sea and the Florida strait (Atlantic Ocean, Table S1; Not et al. 2007; Cuvelier et al. 2010). Here we place 11 additional sequences from Canada (off Vancouver, NW Pacific and Beaufort Sea, Arctic) and from an alpine lake in Norway into this clade (Table S1). Metabarcoding of samples from outer Oslofjorden, Norway detected 7 OTUs clustering with clade B3 (Egge et al. 2015a). Two of these were only found in the picoplankton size-fraction. Members of this clade may thus be found both in marine and fresh, oceanic and coastal, cold to warm, and oligotrophic to eutrophic waters, and some may belong to the picoplankton.

*Clade B4* was introduced by Egge et al. (2015a) to embrace 13 OTUs from outer Oslofjorden (N Atlantic) together with environmental sequences from Villefranche (Mediterranean Sea, France), Hawaii (Pacific Ocean) and Sargasso Sea/Florida (Atlantic Ocean). Here we also place sequences from South China Sea (E Pacific Ocean) and California (NW Pacific Ocean), and the Baltic Sea (Table S1, 2A). All B4 sequences are from marine or brackish surface waters, from coastal or open oceans.

*Clade B5* is introduced here to embrace four OTUs. Two originate from the equatorial Pacific Ocean (clones OLI51059, LOI51033), one from off California, and one from a lagoon in the Mediterranean Sea.

#### Subclass Calcihaptophycidae

The subclass Calcihaptophycidae was erected to comprise the calcifying coccolithophorids as well as non-calcifying species affiliated with these, such as members of the genus *Isochrysis* (de Vargas et al. 2007). This taxon is convenient for novel environmental sequences clustering with the coccolithophores, as it is not known whether they represent actually calcifying species or not. In Edvardsen et al. (2000) this clade was named Clade C. About 208 extant species of Calcihaptophyceae have been formally described: 26 Coccolithales, 17 Isochrysidales, 58 Syracosphaerales, 10 Zygodiscales, 36 cf. Syracosphaerales incertae sedis, 48 holococcolithophorids and 13 nannolith-bearing species (Jordan et al. 2004; Edvardsen et al. unpublished). However, each holococcolithophorid is now believed to be part of the life cycle of a heterococcolithophorid species (Houdan et al. 2004; Billard & Inouye 2004), and thus 160 species are left, of which about 120 are well-described (Geisen et al. 2004). For 41 spp. 18S rRNA gene sequences are available: 85% are available for Coccolithales, but only 3% for Syracosphaerales (Table 1). Environmental sequences in our database assigned to the coccolithophorid orders Coccolithales (10), Syracosphaerales (3), Zygodiscales (1), and families Braarudosphaeraceae (13), and Noëlaerhabdaceae in Isochrysidales (8) all originate from

marine waters (Table S1). Only two sequences originated from brackish waters (assigned to Hymenomonadaceae, cf. *Jomoniolithus* sp.). Three coccolithophorid species (*Hymenomonas roseola*, *Acanthoica schilleri*, *A. ornata*) have been recorded from freshwater (Preisig 2002), but their DNA sequences are not available. Environmental sequences assigned to Isochrysidales fall into three main clades representing the families Noëlaerhabdaceae (incl. *Emiliania* / *Gephyrocapsa*), the non-calcifying Isochrysidae, and a clade with only environmental sequences (originally named Clade EV and renamed here to Clade C1) introduced by Simon et al. (2013) (Table S1). The naked genus *Dicrateria* was transferred to Prymnesiales and now also includes members of the previous genus *Imantonia* (Bendif et al. 2013). Sequences assigned to Clade C1 of Isochrysidales all originate from freshwater lakes (Simon et al. 2013). The clade forms a sister clade to Isochrysidae in our global 18S rRNA gene tree (Fig. 2, S1) and may represent a novel family of Isochrysidales.

The calcifying family Braarudosphaeraceae presently consists of the genus *Braarudosphaera* and the species *Chrysochromulina parkae*, suggested to be a life cycle stage of *B. bigelowii* or a sibling species to *B. bigelowii* (Hagino et al. 2013). In our analysis Braarudosphaeraceae nests within Calcihaptophycidae (Fig. 2), but its placement in the haptophyte tree is uncertain and changes depending on the analysis, and may also fall within Prymnesiales (Hagino et al. 2013).

The low diversity of coccolithophorids from 18S environmental sequencing studies (listed in Table S1) compared to morphological species may be due to various biases. Some species may have identical 18S rRNA (such as *E. huxleyi* and *Gephyrocapsa oceanica*) or be joined in a common haploid-diploid life cycle. Several environmental sequencing studies included only the picoplanktonic size-fraction (Table 2). Coccolithophorids are in general larger (c. 4–40 µm) and would be removed during 3 µm-filtration (Young et al. 2014) used in many studies.

*Clade E* was introduced by Moon-van der Staay (2000) to include two environmental picoplankton sequences without cultured representatives originating from oligotrophic, equatorial Pacific Ocean. Shi et al. (2009) described one additional picoplankton OTU also from the equatorial Pacific Ocean falling in this clade. One OTU from outer Oslofjorden from the pico-size fraction was further assigned to Clade E. All environmental sequences in Clade E are thus from the pico-size fraction. This well supported clade has an uncertain placement and is either placed as sister to the coccolithophorids (e.g. Edvardsen et al. 2000), between Isochrysidales and the remaining coccolithophorid orders (Edvardsen & Medlin 2007; Egge et al. 2015a), or, as here (Fig. 2), as a sister to members of Syracosphaerales. *Chrysoculter rhomboides* was described in 2005 from culture and suggested to belong to Clade E (Nakayama et al. 2005). However, in the phylogeny of Egge et al. (2015a) it falls in a sister

clade of Clade E, and in this study (Fig. 2, S1) at the base of Calcihaptophycidae.

*Clade F* was introduced by Egge et al. (2015a) and comprises environmental sequences from the Caribbean Sea, off Vancouver in the NW Pacific Ocean, Sargasso Sea in the Atlantic Ocean, and from sea ice in the Baltic Sea (Table S1 with references), as well as 3 metabarcode OTUs from outer Oslofjorden (Skagerrak, N Atlantic). In the latter study, OTU 4 dominated completely the haptophyte community during the diatom-dominated spring bloom (Egge et al. 2015b). Clade F forms a sister group to Zygodiscales in Egge et al. (2015a) and in this study (Fig. 2).

### Major clades without cultured and described species

Phylogenetic analyses based on 18S rRNA gene sequences show that there are a number of major clades without cultured representatives. This suggests that novel taxa, ranging from class to genus level, have yet to be described.

#### Diversity of putative novel haptophyte classes, Clades HAP1 to HAP5

Sequences assigned to Clade HAP1 have been detected twice and only in fresh water sediments (Slapeta et al. 2005; Shalchian-Tabrizi et al. 2011). In these studies, phylogenetic analyses placed it between the two described haptophyte classes, and it was proposed to represent a novel class. In the present analysis Clade HAP1 falls basal to all haptophytes.

Sequences assigned to HAP2, 3, 4, and 5, have only been recorded in marine plankton samples – in diverse environments from the Arctic to the tropics. All four clades have been found both in the pico- and nano-plankton size-fractions (Egge et al. 2015a).

Sequences assigned to Clade HAP2 were first recorded from picoplankton from the DCM of the equatorial Pacific Ocean and were suggested by the authors to represent a novel class (Shi et al. 2009). Later, environmental sequences from off Florida, USA, (Cuvelier et al. 2010) and the Caribbean Sea (Edgcomb et al. 2011) were assigned to this clade that may indeed warrant a new class when the morphology is revealed.

Clade HAP3 was introduced by Simon et al. (2013) and also suggested to represent a novel class. Sequences assigned to HAP3 have been recorded from the Marmara Sea (Simon et al. 2013), off Vancouver in the NW Pacific (Orsi et al. 2012) and from the South China Sea in the W Pacific Ocean (Wu et al. 2014), all from the pico-size fraction, and in the outer Oslofjorden (Egge et al. 2015a) from both the pico- and nano-size fractions. Egge et al. (2015a) found four OTUs of which two were present in surface water in the outer Oslofjorden most of the year whereas the two others only occurred during autumn. This clade has only c. 90% sequence similarity to representatives from the two current haptophyte classes, and its position within the hap-

tophyte phylogenetic tree is unstable (Simon et al. 2013; Egge et al. 2015a). It is either placed between the classes Pavlovophyceae and Prymnesiophyceae (Simon et al. 2013; Egge et al. 2015a; this study) or basal to all haptophytes (Simon et al. 2013), both positions without support.

Clades HAP4 and HAP5 were introduced by Egge et al. (2015a) and were suggested to represent putative novel classes. Clade HAP4 consists of seven environmental sequences from the Caribbean Sea (Edgcomb et al. 2011), from 2500 m depth in the Sargasso Sea (Countway et al. 2007), 120 m deep in the Saanich Inlet (N Pacific coast) (Orsi et al. 2012), the N Pacific Ocean (Frias-Lopez et al. 2009), in the Arctic at 500 m depth (Lie et al. 2014), as well as in the South China Sea (Wu et al. 2014; Fig. S1; Table S1). Egge et al. (2015b) found six metabarcode OTUs in outer Oslofjorden assigned to HAP4 present mainly during autumn (Sep-Nov), when the influence of the saline N Atlantic current is strongest. Clearly, members of this novel lineage have a wide geographical distribution and may live at great depths in the open ocean, suggesting heterotrophic nutrition, as well as in surface coastal waters.

Clade HAP5 was erected to embrace two environmental sequences, one from off Vancouver, N Pacific (identified as a chimera in this study and removed, Table S1), and one in the Atlantic Ocean, and three metabarcoding OTUs from Oslofjorden, Skagerrak (Egge et al. 2015a). One of the Oslofjorden OTUs was present most of the year and was the 15th most abundant OTU from this two years monthly study (Egge et al. 2015b). We have included one additional environmental sequence from S China Sea.

### Distribution patterns revealed by environmental sequencing

Haptophyte distribution patterns have initially been studied by light and electron microscopy, mainly for the well-calcified coccolithophorids or colony-forming *Phaeocystis* species (e.g. Thomsen et al. 1994; Jordan & Chamberlain 1997; Medlin & Zingone 2007). Environmental sequencing and metabarcoding studies have revealed that tiny non-calcifying members of Prymnesiales have a very wide distribution both geographically and vertically (Table S1), and may contribute to most of the haptophyte diversity and relative abundance (Liu et al. 2009; Cuvelier et al. 2010; Egge et al. 2015b). A high diversity of *Chrysotrichulina* species has been found in the picoplankton in open oceans. Data from environmental sequencing support that some species such as *Emiliania huxleyi* have a very wide geographic distribution whereas others have a more restricted distribution both in time and space (Jordan & Chamberlain 1997; Egge et al. 2015b). Among the novel lineages without a cultured representative, members of clades HAP1 and Clade C1 seem to be restricted to freshwater, and Clades HAP2 to HAP5 and D, E, F to marine waters. Members of Clade D are found in tropical,

temperate and polar waters. Some haptophyte OTUs are present all throughout the year in the temperate Skagerrak (e.g. *Chrysotrichomonas simplex* and *Phaeocystis cordata*) while others occur mainly in the late summer-late autumn, such as Coccolithales (Egge et al. 2015b). The latter is supported by microscopical surveys in the same region (Gaarder 1971).

## Conclusions and perspectives

Environmental sequencing of clone libraries has been fundamental for our knowledge of haptophyte diversity and distribution. Especially for the tiny pico-haptophytes and the non-calcifying members, that are fragile, have few easily observable morphological characters, and often require transmission electron microscopy and specific taxonomic expertise for species identification. A high diversity, expected to represent many novel species and lineages, has been revealed, and their distribution pattern begins to be deciphered. More full-length sequences, from the rRNA genes (nuclear 18S, 28S and plastid 16S) and also from more resolutive DNA regions such as the rRNA gene spacers ITS1 and ITS2 are needed, from cultured species, flow cytometry-sorted populations, or individually picked cells. More taxa need to be formally described, to be able to assign accurately a sequence to a known species and thus connect a genotype to a phenotype. Studies coupling metabarcoding and electron microscopy may also provide educated guesses as to the identity of environmental sequences that do not match reference sequences from cultures (e.g. Young et al. 2014). Using fluorescence *in situ* hybridisation (FISH), it may be possible to detect and observe under the microscope cells belonging to a taxon for which specific oligonucleotide probes can be designed (e.g. Kolodziej & Stoeck 2007). For metabarcoding, the use of more than one primer set, each with different specificity will enable to generate data for groups for which we have little information such as Clades HAP1–HAP5 that may correspond to novel classes. Long read length is necessary for a correct phylogenetic placement and usually for taxonomic resolution down to the species level. The HTS 454 Roche technology is progressively phased out and the Illumina MiSeq, presently allowing read length of 2 x 300 bp, is now replacing it for protist metabarcoding surveys. The technological developments in this field are extremely fast and new technologies are emerging such as Single Molecule, Real-Time DNA sequencing pioneered by Pacific Biosciences resulting in very long read length up to 20K bp. We suggest that plankton, DNA and/or RNA/cDNA from molecular plankton surveys should be stored in Biobanks to be able to reanalyse these with improved HTS technologies in the future. There is a need for standardised protocols for field sampling, laboratory work, HTS, and bioinformatics treatment to be able to compare results from different metabarcoding studies, especially if the data are to be used in long term monitoring. Curated taxonomic sequence data-

bases, such as the one presented here for Haptophyta, will not only contribute to this standardisation, but also enable a more detailed taxonomic assignation needed in ecological and evolutionary studies.

**Acknowledgements:** The BioMarKs consortium (ERA-Net BiodivERsA-2008-6530) is thanked for the idea of taxonomic protist reference databases, and Laure Guillou and Lucie Bittner for assistance with a previous version of the Haptophyte database. Sandra Gran Stadniczenko is thanked for running the RAxML analysis on the Abel Computing Cluster. The Research Council of Norway is acknowledged for financial support through the projects HAPTODIV (190307/S40) and MicroPolar (225956/E10). Financial support for this work was also provided by the European Union programs MicroB3 (UE-contract-287589) and MaCuMBA (FP7-KBBE-2012-6-311975).

## References

- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. (2009): A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. – PLoS One 4:e6372.
- Andersen, R.A., Bailey, J.C., Decelle, J. & Probert, I. (2015a): *Phaeocystis rex* sp. nov. (Phaeocystales, Prymnesiophyceae): a new solitary species that produces a multilayered scale cell covering. – Eur. J. Phycol. 50: 207–222.
- Andersen, R.A., Kim, J.I., Tittley, I. & Yoon, H.S. (2014): A re-investigation of *Chrysotila* (Prymnesiophyceae) using material collected from the type locality. – Phycologia 53: 463–473.
- Andersen, R.A., Kim, J.I., Tittley, I. & Yoon, H.S. (2015b): *Chrysotila dentata* comb. nov., *Chrysotila roscoffensis* comb. nov. and *Chrysocapsa wetherbeeii* sp. nov. – Phycologia 54: 321–322.
- Balzano, S., Marie, D., Gourvil, P. & Vault, D. (2012): Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. – ISME J. 6: 1480–1498.
- Behnke, A., Barger, K.J., Bunge, J. & Stoeck, T. (2010): Spatio-temporal variations in protistan communities along an O<sub>2</sub> /H<sub>2</sub>S gradient in the anoxic Framvaren Fjord (Norway). – FEMS Microbiol. Ecol. 72: 89–102.
- Bendif, E.M., Probert, I., Carmichael, M., Romac, S., Hagino, K. & de Vargas, C. (2014): Genetic delineation between and within the widespread coccolithophore morpho-species *Emiliania huxleyi* and *Gephyrocapsa oceanica* (Haptophyta). – J. Phycol. 50: 140–148.
- Bendif, E.M., Probert, I., Hervé, A., Billard, C., Goux, D., Lelong, C., Cadoret, J.P. et al. (2011): Integrative taxonomy of the Pavlovophyceae (Haptophyta): a reassessment. – Protist 162: 738–761.
- Bendif, E.M., Probert, I., Schroeder, D.C. & Vargas, C. (2013): On the description of *Tisochrysis lutea* gen. nov. sp. nov. and *Isochrysis nuda* sp. nov. in the Isochrysidales, and the transfer of *Dicrateria* to the Prymnesiales (Haptophyta). – J. Appl. Phycol. 25: 1763–1776.
- Billard, C. & Inouye, D.W. (2004): What is new in coccolithophore biology? – In: Thierstein, H.R. & Young, J.R. (Eds.),

- Coccolithophores: from molecular processes to global impact. Springer, Berlin, pp. 1–29.
- Bittner, L., Gobet, A., Audic, S., Romac, S., Egge, E.S., Santini, S., Ogata, H. et al. (2013): Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. – Mol. Ecol. 22: 87–101.
- Burki, F., Okamoto, N., Pombert, J.-F. & Keeling, P.J. (2012): The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. – Proc. R. Soc. B Biol. Sci. 279: 2246–2254.
- Cavalier-Smith, T., Chao E.E., & Lewis, R. (2015): Multiple origins of Heliozoa from flagellate ancestors: New cryptist subphylum Corbihelia, superclass Corbistoma, and monophyly of Haptista, Cryptista, Hacrobia and Chromista. – Mol. Phylogen. Evol. 93: 331–362.
- Chrétiennot-Dinet, M.-J., Desreumaux, N. & Vignes-Lebbe, R. (2014): An interactive key to the *Chryschromulina* species (Haptophyta) described in the literature. – PhytoKeys 34: 47–60.
- Countway, P.D., Gast, R.J., Dennett, M.R., Savai, P., Rose, J.M. & Caron, D.A. (2007): Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). – Environ. Microbiol. 9: 1219–1232.
- Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messié, M., Tringe, S.G., Woyke, T. et al. (2010): Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. – Proc. Natl. Acad. Sci. U. S. A. 107: 14679–14684.
- Dasilva, C.R., Li, W.K.W. & Lovejoy, C. (2014): Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean. – J. Plankton Res. 36: 344–363.
- de Vargas, C., Aubry, M.-P., Probert, I. & Young, J. (2007): Origin and evolution of Coccolithophores: from coastal hunters to oceanic farmers. – In: Falkowski, P. & Knoll, A.H. (Eds.), Evolution of Primary Producers in the Sea, pp. 251–285.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E. et al. (2015): Eukaryotic plankton diversity in the sunlit ocean. – Science 348: 1261605.
- Decelle, J., Probert, I., Bittner, L., Desderves, Y., Colin, S., de Vargas, C., Galí, M. et al. (2012): An original mode of symbiosis in open ocean plankton. – Proc. Natl. Acad. Sci. U. S. A. 109: 18000–18005.
- Decelle, J., Romac, S., Stern, R.F., Bendif, E.M., Zingone, A., Audic, S., Guiry, M.D. et al. (2015): PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. – Mol. Ecol. Resour. 15: 1435–1445.
- Domaizon, I., Lepère, C., Debroas, D., Bouvy, M., Ghiglione, J., Jacquet, S., Bettarel, Y. et al. (2012): Short-term responses of unicellular planktonic eukaryotes to increases in temperature and UVB radiation. – BMC Microbiol. 12: 202.
- Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., Holder, M. et al. (2011): Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. – ISME J. 5: 1344–1356.
- Edvardsen, B., Eikrem, W., Green, J., Andersen, R.A., Moon-van der Staay, S.Y. & Medlin, L.K. (2000): Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. – Phycologia 39: 19–35.
- Edvardsen, B., Eikrem, W., Throndsen, J., Sáez, A.G., Probert, I. & Medlin, L.K. (2011): Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta). – Eur. J. Phycol. 46: 202–228.
- Edvardsen, B. & Imai, I. (2006): The ecology of harmful flagellates within Prymnesiophyceae and Raphidophyceae. – In: Graneli, E. & Turner, J.T. (Eds.), Ecology of Harmful Algae. Springer, Berlin, pp. 67–79.
- Edvardsen, B. & Medlin, L. (2007): Molecular systematics of the Haptophyta. – In: Lewis, J. & Brodie, J. (Eds.), Unravelling the algae – the past, present and future of algal molecular systematics. The Systematics Association. Taylor and Francis, London, New York, pp. 183–196.
- Egge, E.S., Bittner, L., Andersen, T., Audic, S., de Vargas, C. & Edvardsen, B. (2013): 454 Pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. – PLoS One 8: e74371.
- Egge, E.S., Eikrem, W. & Edvardsen, B. (2015a): Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. – J. Eukaryot. Microbiol. 62: 121–140.
- Egge, E.S., Johannessen, T.V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., Sandaa, R.-A. et al. (2015b): Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. – Mol. Ecol. 24: 3026–3042.
- Eikrem, W. & Throndsen, J. (1999): The morphology of *Chryschromulina rotalis* sp. nov. (Prymnesiophyceae, Haptophyta), isolated from the Skagerrak. – Sarsia 84: 445–449.
- Frias-Lopez, J., Thompson, A., Waldbauer, J. & Chisholm, S.W. (2009): Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. – Environ. Microbiol. 11: 512–525.
- Fujiwara, S., Tsuzuki, M., Kawachi, M., Minaka, N. & Inouye, I. (2001): Molecular phylogeny of the Haptophyta based on the rbc L gene and sequence variation in the spacer region of the RUBISCO operon. – J. Phycol. 129: 121–129.
- Fuller, N.J., Campbell, C., Allen, D.J., Pitt, F.D., Le Gall, F., Vaultor, D. & Scanlan, D.J. (2006): Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using a PCR biased towards marine algal plastids. – Aquat. Microb. Ecol. 43: 79–93.
- Gaarder, K.R. (1971): Comments on the distribution of coccolithophorids in the oceans. – In: Funnell, B. & Riedel, W. (Eds.), The Micropalaeontology of Oceans. Cambridge University Press, Cambridge, pp. 97–103.
- Geisen, M., Young, J.R., Probert, I., Sáez, A.G., Baumann, K.H., Sprengel, C., Böllmann, J. et al. (2004): Species level variation in coccolithophores. – In: Thierstein, H.R. & Young, J.R. (Eds.), Coccolithophores – From Molecular Processes to Global Impact. Springer, Berlin, pp. 327–366.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C. et al. (2012): The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. – Nucleic Acids Res. 41(DI): D597–D604.
- Houdan, A., Billard, C., Marie, D., Not, F., Sáez, A.G., Young, J.R. & Probert, I. (2004): Holococcolithophore-heterococcolithophore (Haptophyta) life cycles: flow cytometric analysis of relative ploidy levels. – Syst. Biodivers. 1: 453–465.
- Jones, H.L.J., Leadbeater, B.S.C. & Green, J.C. (1993): Mixotrophy in marine species of *Chryschromulina* (Prymnesiophyceae): ingestion and digestion of a small green flagellate. – J. Mar. Biol. Assoc. United Kingdom 73: 283–296.

- Jordan, R., Cros, L. & Young, J.R. (2004): A revised classification scheme for living haptophytes. – *Micropaleontology* 50: 55–79.
- Jordan, R.W. & Chamberlain, A.H.L. (1997): Biodiversity among haptophyte algae. – *Biodivers. Conserv.* 6: 131–152.
- Kawachi, M., Inouye, I., Maeda, O. & Chihara, M. (1991): The haptonema as a food-capturing device: observations on *Chryschromulina hirta* (Prymnesiophyceae). – *Phycologia* 30: 563–573.
- Kiliias, E., Wolf, C., Nöthig, E.-M., Peeken, I. & Metfies, K. (2013): Protist distribution in the Western Fram Strait in summer 2010 based on 454-pyrosequencing of 18S rDNA. – *J. Phycol.* 49: 996–1010.
- Kiliias, E.S., Nöthig, E.-M., Wolf, C. & Metfies, K. (2014): Picoeukaryote plankton composition off West Spitsbergen at the entrance to the Arctic Ocean. – *J. Eukaryot. Microbiol.* 61: 569–579.
- Koid, A., Nelson, W.C., Mraz, A. & Heidelberg, K.B. (2012): Comparative analysis of eukaryotic marine microbial assemblages from 18S rRNA gene and gene transcript clone libraries by using different methods of extraction. – *Appl. Environ. Microbiol.* 78: 3958–3965.
- Kolodziej, K. & Stoeck, T. (2007): Cellular identification of a novel uncultured marine stramenopile (MAST-12 Clade) small-subunit rRNA gene sequence from a norwegian estuary by use of fluorescence in situ hybridization-scanning electron microscopy. – *Appl. Environ. Microbiol.* 73: 2718–2726.
- Leadbeater, B.S.C. (1972): Fine structural observations in six new species of *Chryschromulina* (Haptophyceae) from Norway with preliminary observations on scale production in *C. microcylindra* sp. nov. – *Sarsia* 49: 65–80.
- Leadbeater, B.S.C. (1994): Cell coverings. – *The Haptophyta algae* 51: 23–46.
- Lepère, C., Vaulot, D. & Scanlan, D.J. (2009): Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. – *Environ. Microbiol.* 11: 3105–3117.
- Lie, A.A.Y., Liu, Z., Hu, S.K., Jones, A.C., Kim, D.Y., Countway, P.D., Amaral-Zettler, L.A. et al. (2014): Investigating microbial eukaryotic diversity from a global census: Insights from a comparison of pyrotag and full-length sequences of 18S rRNA genes. – *Appl. Environ. Microbiol.* 80: 4363–4373.
- Liu, H., Aris-Brosou, S., Probert, I. & de Vargas, C. (2010): A time line of the environmental genetics of the Haptophytes. – *Mol. Biol. Evol.* 27: 161–176.
- Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F. et al. (2009): Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. – *Proc. Natl. Acad. Sci. U. S. A.* 106: 12803–12808.
- Marie, D., Shi, X.L., Rigaut-Jalabert, F. & Vaulot, D. (2010): Use of flow cytometric sorting to better assess the diversity of small photosynthetic eukaryotes in the English Channel. – *FEMS Microbiol. Ecol.* 72: 165–178.
- McDonald, S.M., Sarno, D., Scanlan, D.J. & Zingone, A. (2007): Genetic diversity of eukaryotic ultraphytoplankton in the Gulf of Naples during an annual cycle. – *Aquat. Microb. Ecol.* 50: 75–89.
- Medlin, L. & Zingone, A. (2007): A taxonomic review of the genus *Phaeocystis*. – *Biogeochemistry* 83: 3–18.
- Medlin, L.K., Kooistra, W.H.C.F., Potter, D., Saunders, G. & Andersen, R.A. (1997): Phylogenetic relationships of the “golden algae” (haptophytes, heterokont chromophytes) and their plastids. – In: Bhattacharya, D. (Ed.), *Origins of Algae and Their Plastids*. Springer-Verlag, Vienna, pp. 187–219.
- Medlin, L.K., Lange, M., Edvardsen, B. & Larsen, A. (2000): Cosmopolitan haptophyte flagellates and their genetic links. – *Syst. Assoc. Spec. Vol. Ser.* 59: 288–308.
- Moon-van der Staay, S.Y., De Wachter, R. & Vaulot, D. (2001): Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. – *Nature* 409: 607–610.
- Moon-van der Staay, S.Y., van der Staay, G.W.M., Guillou, L., Vaulot, D., Claustre, H. & Medlin, L.K. (2000): Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. – *Limnol. Oceanogr.* 45: 98–109.
- Nakayama, T., Yoshida, M. & Noël, M. (2005): Ultrastructure and phylogenetic position of *Chrysoculter rhomboideus* gen. et sp. nov. (Prymnesiophyceae), a new flagellate haptophyte from Japanese coastal waters. – *Phycologia* 44: 369–383.
- Nersveen, L. (2011): Diversitet av nano- og pikoplankton i den atlantiske delen av Sørishavet, undersøkt med molekylærbiologiske metoder og bruk av numeriske likhets- og artsrikhetsanalyser. MSc. thesis, University of Oslo, 78 pp.
- Not, F., del Campo, J., Balagué, V., de Vargas, C. & Massana, R. (2009): New insights into the diversity of marine picoeukaryotes. – *PLoS One* 4: e7143.
- Not, F., Gausling, R., Azam, F., Heidelberg, J.F. & Worden, A.Z. (2007a): Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. – *Environ. Microbiol.* 9: 1233–1252.
- Not, F., Siano, R., Kooistra, W.H.C.F.H.C.F., Simon, N., Vaulot, D. & Probert, I. (2012): Diversity and ecology of eukaryotic marine phytoplankton. – In: Piganeau, G. (Ed.), *Genomic Insights Gained into the Diversity, Biology and Evolution of Microbial Photosynthetic Eukaryotes*. Elsevier, Amsterdam, pp. 1–53.
- Not, F., Zapata, M., Pazos, Y., Campaña, E., Doval, M. & Rodríguez, F. (2007b): Size-fractionated phytoplankton diversity in the NW Iberian coast: a combination of microscopic, pigment and molecular analyses. – *Aquat. Microb. Ecol.* 49: 255–265.
- Okamoto, N., Chantangsi, C., Horák, A., Leander, B.S. & Keeling, P.J. (2009): Molecular phylogeny and description of the novel Katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. – *PLoS One* 4: e7080.
- Orsi, W., Song, Y.C., Hallam, S. & Edgcomb, V. (2012): Effect of oxygen minimum zone formation on communities of marine protists. – *ISME J.* 6: 1586–1601.
- Preisig, H. (2002): Phylum Haptophyta (Prymnesiophyta). – In: John, D., Whitton, B. & Brook, A. (Eds.), *The Freshwater Algal Flora of the British Isles*. Cambridge University Press, Cambridge, pp. 211–213.
- Quince, C., Lanzen, A., Davenport, R. & Turnbaugh, P. (2011): Removing noise from pyrosequenced amplicons. – *BMC Bioinformatics* 12: 38.
- Sáez, A.G., Probert, I., Yound, J.R., Edvardsen, B., Eikrem, W. & Medlin, L.K. (2004): A review of the phylogeny of the Haptophyta. – In: Thierstein, H.R. & Young, J.R. (Eds.), *Coccolithophores – From Molecular Processes to Global Impact*. Springer, Berlin, pp. 251–269.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A. et al. (2009): Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. – *Appl. Environ. Microbiol.* 75: 7537–7541.

- Shalchian-Tabrizi, K., Reier-Røberg, K., Ree, D.K., Klaveness, D. & Bråte, J. (2011): Marine-freshwater colonizations of haptophytes inferred from phylogeny of environmental 18S rDNA sequences. – *J. Eukaryot. Microbiol.* 58: 315–318.
- Shi, X.L., Lepèze, C., Scanlan, D.J. & Vaulot, D. (2011): Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. – *PLoS One* 6: e18979.
- Shi, X.L., Marie, D., Jardillier, L., Scanlan, D.J. & Vaulot, D. (2009): Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. – *PLoS One* 4: e7657.
- Simon, M., López-García, P., Moreira, D. & Jardillier, L. (2013): New haptophyte lineages and multiple independent colonizations of freshwater ecosystems. – *Environ. Microbiol. Rep.* 5: 322–332.
- Simon, N., Brenner, J., Edvardsen, B. & Medlin, L. (1997): The identification of *Chrysochromulina* and *Prymnesium* species (Haptophyta, Prymnesiophyceae) using fluorescent or chemiluminescent oligonucleotide probes: a means for improving studies on toxic algae. – *Eur. J. Phycol.* 32: 393–401.
- Slapeta, J., Moreira, D. & López-García, P. (2005): The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. – *Proc. Biol. Sci.* 272: 2073–2081.
- Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.W. & 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.
- Taylor, J.D. & Cunliffe, M. (2014): High-throughput sequencing reveals neustonic and planktonic microbial eukaryote diversity in coastal waters. – *J. Phycol.* 50: 960–965.
- Terrado, R., Scarella, K., Thaler, M., Vincent, W.F. & Lovejoy, C. (2013): Small phytoplankton in Arctic seas: vulnerability to climate change. – *Biodiversity* 14: 2–18.
- Thiele, S., Wolf, C., Schulz, I.K., Assmy, P., Metfies, K. & Fuchs, B.M. (2014): Stable composition of the nano- and picoplankton community during the Ocean iron fertilization experiment LOHALEX. – *PLoS One* 9: e113244.
- Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaulot, D., Kuypers, M.M.M. et al. (2012): Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. – *Science* 337: 1546–50.
- Thomsen, H.A., Buck, K.R. & Chavez, F.P. (1994): Haptophytes as components of marine phytoplankton. – In: Green, J.C. & Leadbeater, B.S.C. (Eds.), *The Haptophyte Algae*. Clarendon Press, Oxford, pp. 187–208.
- Unrein, F., Gasol, J.M., Not, F., Forn, I. & Massana, R. (2013): Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. – *ISME J.* 8: 164–176.
- Vaulot, D., Eikrem, W., Viprey, M. & Moreau, H. (2008): The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. – *FEMS Microbiol. Rev.* 32: 795–820.
- Weber, F., del Campo, J., Wylezich, C., Massana, R. & Jürgens, K. (2012): Unveiling Trophic Functions of Uncultured Protist Taxa by Incubation Experiments in the Brackish Baltic Sea. – *PLoS One* 7: e41970.
- Worden, A.Z. & Not, F. (2008): Ecology and diversity of picoeukaryotes. – In: Kirchman, D.L. (Ed.), *Microbial Ecology of the Ocean*. 2nd ed. Wiley, pp. 159–205.
- Wu, W., Huang, B., Liao, Y. & Sun, P. (2014): Picoeukaryotic diversity and distribution in the subtropical-tropical South China Sea. – *FEMS Microbiol. Ecol.* 89: 563–579.
- Yoshida, M., Noël, M.-H., Nakayama, T., Naganuma, T. & Inouye, I. (2006): A haptophyte bearing siliceous scales: ultrastructure and phylogenetic position of *Hyalolithus neolepis* gen. et sp. nov. (Prymnesiophyceae, Haptophyta). – *Protist* 157: 213–34.
- Young, J.R., Liu, H., Probert, I., Aris-Brosou, S. & Vargas, C. De (2014): Morphospecies versus phyllospecies concepts for evaluating phytoplankton diversity: the case of the cocolithophores. – *Cryptogam. Algol.* 35: 353–77.
- Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. (2005): Mapping of picoeukaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. – *FEMS Microbiol. Ecol.* 52: 79–92.

Manuscript received: 10 October 2015

Accepted: 14 March 2016

Handling editor: Hervé Moreau

The pdf version (Adobe JavaScript must be enabled) of this paper includes an electronic supplement:

**Fig. S1.** Maximum-likelihood tree (RAxML v. 8.026, GTRCAT) based on 971 haptophyte 18S rRNA gene sequences from Table S1: 451 are from cultured strains and 520 from environmental clone libraries. Bootstrap values are marked at the nodes. The colours indicate taxonomic groups.

**Table S1.** Haptophyte 18S rRNA gene sequence reference database including 971 sequences: 451 from cultures and 520 from environmental clone libraries. See File S1 for description of its construction. Chimeras (32) identified by uchime in mothur and by manual blast are listed on sheet 2.

**Table S2.** Commonly used PCR primers for environmental sequencing and metabarcoding of haptophytes.

**File S1.** Description of the construction of the Haptophyta 18S rRNA gene reference sequence database, alignment and final RAxML phylogenetic tree.

**File S2.** Fasta file with accession numbers and 18S rRNA gene sequence of 971 haptophyte reference sequences for taxonomic assignment using mothur or Qiime.

**File S3.** Taxonomy file with systematic placement of 971 haptophyte 18S rRNA gene reference sequences for taxonomic assignment using mothur or Qiime.

**File S4.** Curated haptophyte 18S rRNA gene alignment for Figs 2 and S1.

Please download the electronic supplement and rename the file extension to .zip (For security reasons Adobe does not allow to embed .exe, .zip, .rar etc. files).

All supplementary material are available at figshare: <https://dx.doi.org/10.6084/m9.figshare.2759983.v1>