

ORIGINAL ARTICLE

Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA

Jon Bråte¹, Ramiro Logares², Cédric Berney³, Dan Kristofer Ree¹, Dag Klaveness¹, Kjetill S. Jakobsen^{1,4}, Kamran Shalchian-Tabrizi¹

¹Department of Biology, Microbial Evolution Research Group, University of Oslo, Oslo, Norway; ²Department of Limnology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden; ³Department of Zoology, University of Oxford, Oxford, UK and ⁴Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

Protist parasites are ecologically important, as they can have great impact on host population dynamics and functioning of entire ecosystems. Nevertheless, little is known about their prevalence in aquatic habitats. Here, we investigate the diversity and distributional patterns of the protist parasites *Perkinsus* and *Parvilucifera* (Perkinsea). Our approach included 454 pyrosequencing of the 18S rDNA gene obtained from a high-altitude lake (Lake Finsevatn, Norway) and phylogenetic analyses of all publicly available sequences related to Perkinsea. The applied PCR primers target a 450 bp region that encompass the variable V4 region of the 18S rDNA gene and have been optimized for the Titanium upgrade of the 454 technology. Nearly 5000 sequences longer than 150 bp were recovered from nearly all eukaryotic supergroups, and of those, 13 unique sequences were affiliated to Perkinsea. Thus, our new strategy for 454 amplicon sequencing was able to recover a large diversity of distantly related eukaryotes and previously unknown species of Perkinsea. In addition, we identified 40 Perkinsea sequences in GenBank generated by other recent diversity surveys. Importantly, phylogenetic analyses of these sequences identified 17 habitat-specific marine and freshwater clades (PERK 1-17). Hence, only a few successful transitions between these habitats have taken place over the entire history of Perkinsea, suggesting that the boundary between marine and fresh waters may constitute a barrier to cross-colonizations for intracellular parasites.

The ISME Journal (2010) 4, 1144–1153; doi:10.1038/ismej.2010.39; published online 15 April 2010

Subject Category: Integrated genomics and post-genomics approaches in microbial ecology

Keywords: Perkinsea; *Perkinsus*; *Parvilucifera*; pyrosequencing; freshwater; sediments

Introduction

Parasitic unicellular eukaryotes (protists) can have considerable impact on ecosystem functioning. For example, the demise of large microalgal populations may affect the entire food chain in several aquatic ecosystems (Chambouvet *et al.*, 2008). The importance of parasites has been acknowledged in ecological and evolutionary studies on macroorganisms, however their impact on protistan host populations has so far been poorly investigated. New studies are suggesting the existence of a large diversity of protist parasites that infect other unicellular eukaryotes (Chambouvet *et al.*, 2008; Lefèvre *et al.*, 2008; Lepère *et al.*, 2008). Therefore, a major question is how these parasitic protists are

distributed in spatiotemporal scales and whether they follow the distributions of their unicellular hosts.

The distribution of free-living microbes seems to be correlated with environmental salinity shifts. Recent studies indicate that saline and fresh waters contain phylogenetically distinct taxa, implying that cross-colonizations between these environments have been rare during the evolution of most eukaryotic and prokaryotic groups (reviewed in Logares *et al.*, 2009). As parasites usually need to be adapted to both the host and the extracellular environment, their spatiotemporal distribution patterns may therefore differ from that of free-living protists. In particular, parasites that spend most of their life cycle within the host's cell may be more prone to cross the saline–freshwater boundary, as their exposure to the extracellular aquatic environment would be minimal.

One of the most diverse groups of parasitic protists is Alveolata, in which Apicomplexa (for example, *Plasmodium*, *Cryptosporidium* and

Correspondence: K Shalchian-Tabrizi, Department of Biology, University of Oslo, Blindernveien 31, Oslo 0316, Norway.

E-mail: kamran@bio.uio.no

Received 3 December 2009; revised 12 February 2010; accepted 22 February 2010; published online 15 April 2010

Toxoplasma) and several lineages related to dinoflagellates (for example, Ellobiopsids, *Perkinsus* and *Parvilucifera*) seem to be entirely parasitic. In addition, a highly abundant and widespread group uncovered from environmental 18S rDNA libraries, the so-called marine alveolates (MA), appear to be largely parasitic (see Guillou *et al.*, 2008). To date, the *Perkinsus* and *Parvilucifera* (members of Perkinsea) and MA have only been identified in marine environments (Noren *et al.*, 1999; Villalba *et al.*, 2004; Groisillier *et al.*, 2006; Guillou *et al.*, 2008; Leander and Hoppenrath, 2008). However, microscopy observations, as well as reports of deeply diverging alveolate 18S rDNA sequences, point to the existence of freshwater Perkinsea species (Brugerolle, 2002, 2003; Green *et al.*, 2003; Richards *et al.*, 2005; Davis *et al.*, 2007; Lefèvre *et al.*, 2008; Lepère *et al.*, 2008).

Even though numerous surveys of the eukaryotic diversity have been undertaken in both marine and fresh waters over the last decade (López-García and Moreira, 2008), very few studies have focused on deeply diverging dinoflagellates other than the MA. Here, we investigate the diversity and distribution of another dinoflagellate group, Perkinsea, by searching publicly available sequence databases and by 454 pyrosequencing of 18S rDNA obtained from a high-mountain freshwater lake (Lake Finsevatn, Norway). We present a protocol for eukaryote-wide PCR amplification of the variable V4 region, optimized for Titanium upgrade of the GS FLX sequencing technology. Our results indicate a large and previously unknown diversity of Perkinsea, and provide rigorous evidence for the existence of species closely related to both *Perkinsus* and *Parvilucifera* in freshwater. In addition, our trees contained 17 new habitat-specific marine and freshwater clades (PERK 1-17), suggesting that cross-colonizations of marine and fresh waters have only taken place at a few occasions over the entire history of Perkinsea.

Materials and methods

Sample collection, DNA isolation and PCR amplification

Sediment samples were collected from Lake Finsevatn (60°36' N—7°30' E) during March 2009.

Lake Finsevatn is a high-mountain oligo- to mesotrophic lake situated at 1215 m above the sea level in an Arctic climate region. The samples were collected using a simple gravity corer at 18 m depth and the DNA was isolated from filtered cells using PowerSoil DNA isolation kit (MoBio, Carlsbad, CA, USA) following the manufacturers instructions. A PCR strategy aiming at amplifying the broadest possible eukaryotic diversity was designed for amplification and 454 pyrosequencing of the variable V4 region of the 18S rDNA gene. The length of the produced amplicon is about 450 bases and therefore suitable for the experimental conditions of the emulsion PCR. The amplification was carried out in two steps. The first by combining the universal forward primer 3Ndf (Cavalier-Smith *et al.*, 2009) with the reverse primers V4_euk_R1 and V4_euk_R2 in two separate reactions. Adaptor A and B, as well as multiplexing tag (MID) were added to these amplicons in a second PCR by using template from the first PCR and composite primers (Table 1). The amplicons were pooled and cleaned on a Wizard SV column (Promega, Madison, WI, USA) before emulsion PCR. All amplifications were carried out in an Eppendorf Mastercycler ep (Eppendorf, Hamburg, Germany). Each amplification reaction (25 µl total) contained 7–50 ng of template DNA, 1 × DreamTaq Buffer with 2.5 mM Mg²⁺ (Fermentas, Burlington, Canada), 200 µM dNTPs, 0.2 µM of each primer and 0.6 U of DreamTaq DNA Polymerase (Fermentas). The PCR program for the first amplification round was: 94 °C for 2 min, followed by 34 cycles of 30 s at 94 °C, 30 s at 59 °C, 60 s at 72 °C with a final extension at 72 °C for 7 min. The second round of amplification was as follows: 94 °C for 2 min, followed by 15 cycles of 30 s at 94 °C, 30 s at 60 °C and 1 min at 72 °C, then 20 cycles of 30 s at 94 °C, 30 s at 65 °C and 1 min at 72 °C with a final extension for 7 min at 72 °C.

Pyrosequencing and removal of low-quality sequences

Amplicon pyrosequencing of the PCR products was carried out on a GS FLX Titanium machine (454 Life Sciences, Branford, CT, USA). We assessed sequence quality by using several criteria; reads that had degenerate bases, overall low-quality score and were shorter than 150 bp were removed. In addition, we

Table 1 Primer sequences for PCR and pyrosequencing used in this study

Primer name	Sequence 5'–3'	Length (bp)
3Ndf	GGCAAGTCTGGTGCCAG	17
V4_euk_R1	GACTACGACGGTATCT(AG)ATC(AG)TCTTCG	27
V4_euk_R2	ACGGTATCT(AG)ATC(AG)TCTTCG	20
454_V4_euk_F1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTCTCTAGGCAAGTCTGGTGCCAG	57
454_V4_euk_R1	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGACTACGACGGTATCT(AG)ATC(AG)TCTTCG	57
454_V4_euk_R2	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGACTACGACGGTATCT(AG)ATC(AG)TCTTCG	50

Composite primers are shown with pyrosequencing adaptors A and B (bold) and multiplex identifiers (italics).

removed identical reads (the longest reads were kept). All remaining sequences were BLASTed against the NCBI nr database (Altschul *et al.*, 1997). Sequences related to the alveolates were then selected and used in the phylogenetic analyses. Each of these alveolate sequences was carefully checked by manually inspecting the Phred quality scores to avoid inclusion of variable sites caused by sequencing artefacts. Indels in relation to polyhomomer regions usually had low Phred score, and such columns in the alignments were ignored in the phylogenetic analyses. In addition, apomorphic nucleotide characters may be generated by errors in the PCR and sequencing processes, producing an artificially high number of unique reads. Hence, for the phylogenetic analyses, we only approved Perkinsea sequences with variable characters that were shared by at least two independent sequences, of which the longest were used.

Data mining and identification of Perkinsea-related sequences

We identified publicly available environmental 18S rDNA sequences potentially related to Perkinsea by examining the results from several published surveys (Table S2). Sequences reported as either *Perkinsus* or likely to be related to *Perkinsus*, were used as queries in BLAST unrestricted and restricted (that is, using 'uncultured' and 'freshwater' as query limitations) searches against the NCBI nr database (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>; Altschul *et al.*, 1997). BLAST searches were also carried out with *Perkinsus* sp. and *Parvilucifera* sp. as query sequences. To confirm that all new and downloaded sequences were evolutionarily related to Perkinsea, we analyzed them together with a broad sampling of eukaryotes (see Berney *et al.*, 2004). The alignment (AL1) consisted of 564 taxa and 1123 unambiguously aligned characters and was generated with MacClade v4.07 (Maddison and Maddison, 2000).

Alveolate and Perkinsea alignments

On the basis of the inferred global eukaryote tree from AL1, another alignment containing only alveolate taxa and the identified environmental sequences related to Perkinsea was constructed (AL2) to increase the number of unambiguously aligned characters (AL2 consisted of 130 taxa and 1443 nucleotide positions). The sequences were aligned with the MAFFT program online version 6.240 (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>) (Kato *et al.*, 2005), using the algorithm E-INS-I for an iterative refinement of the alignment that accounts for conserved motifs embedded among nonalignable regions. The alignment was manually edited and ambiguously aligned characters were excluded using MacClade v4.07 (Maddison and Maddison, 2000). The tree produced from the

alignment was used to identify and remove chimeric and essentially identical sequences (to reduce the alignment size), and subsequently the analyses were repeated (Figure 1). Three chimeric sequences were identified by visual inspection according to Berney *et al.* (2004). Two were removed (EU162624, EU162626) and the third (AY919735) was kept, but the chimeric section was removed (541 bp from the 3'-end). In addition, to further investigate the phylogenetic relationships within the Perkinsea, an alignment (AL3; 59 taxa and 1747 characters) containing only Perkinsea sequences was constructed based on AL 2. Furthermore, as all sequences generated in this study were covering the V4 region of the 18S rDNA gene, we assessed the impact of the short sequences on the tree topology by removing them in a separate phylogenetic analysis of the AL3 alignment (resulting alignment: 46 taxa and 1747 characters).

Phylogenetic analyses

Maximum likelihood analyses of all alignments were carried out using the program RaxML v.7.0.4 (Stamatakis, 2006). The general time reversible (GTR) model with parameters accounting for γ -distributed rate variation across sites (G) and invariable sites (I) was used in all analyses. Likelihood scores from 10 heuristic tree searches from random starting trees were generated, and the topology with the highest estimated likelihood was selected. Bootstrapping was carried out with 100 pseudoreplicates with the same evolutionary model as in the initial search, but with one heuristic search per replicate.

Bayesian phylogenies were reconstructed from alignments 2 and 3 using MrBayes v.3.0 (Ronquist and Huelsenbeck, 2003). In addition to the GTR+G+I model, the covarion parameters (that is, GTR+G+I+COV) was used to accommodate for different substitution rates across sequences. The four MCMC (Markov chain Monte Carlo) chains included three cold and one heated and they lasted for 4 000 000 generations. Two independent inferences, each from random starting trees, were performed. Posterior probabilities (pps) and mean marginal likelihood values of the trees were calculated after the burn-in phase, which was determined from the marginal likelihood scores of the initially sampled trees. The average split frequencies of the two runs was below 0.01, indicating convergence.

Removal of fast evolving taxa and characters with missing data

As *Parvilucifera* seem to be fast evolving, and hence may be prone to long-branch attraction artefacts, AL3 was analyzed both with and without *Parvilucifera* sequences, as well as associated long-branched taxa in the tree (EU162625 and

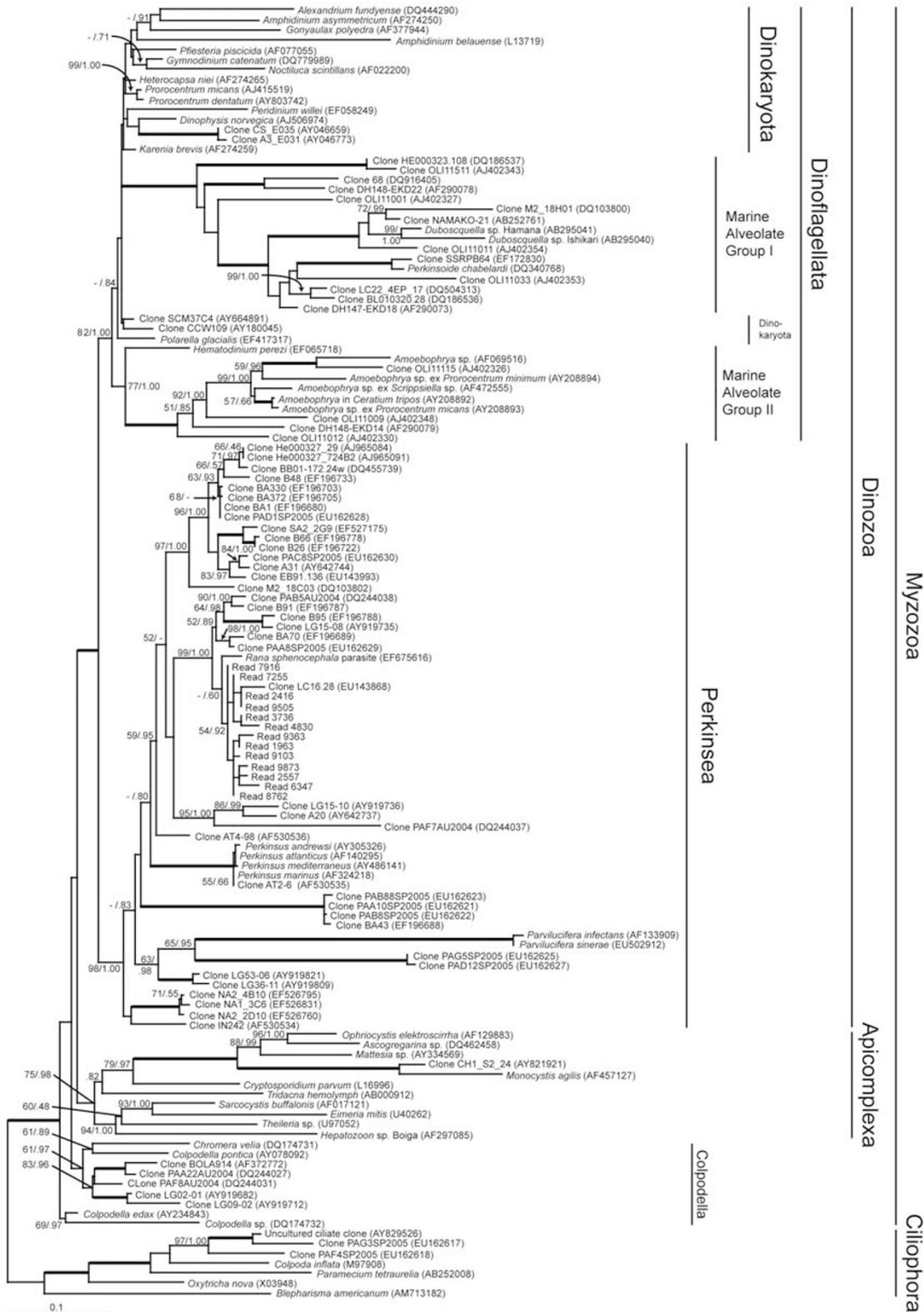


Figure 1 Bayesian phylogeny of alveolate 18S rDNA sequences reconstructed from an alignment consisting of 130 taxa 1443 characters with support values from maximum likelihood (ML) (GTR + G + I) and Bayesian (GTR + G + I + COV) inferences on the internal branches (ML bootstrap support (%)/Bayesian posterior probability values (pp)). Thick lines indicate support values of 1.00% and 1.00 pp. Values below 0.75 pp and 50% are not shown except for some backbone nodes.

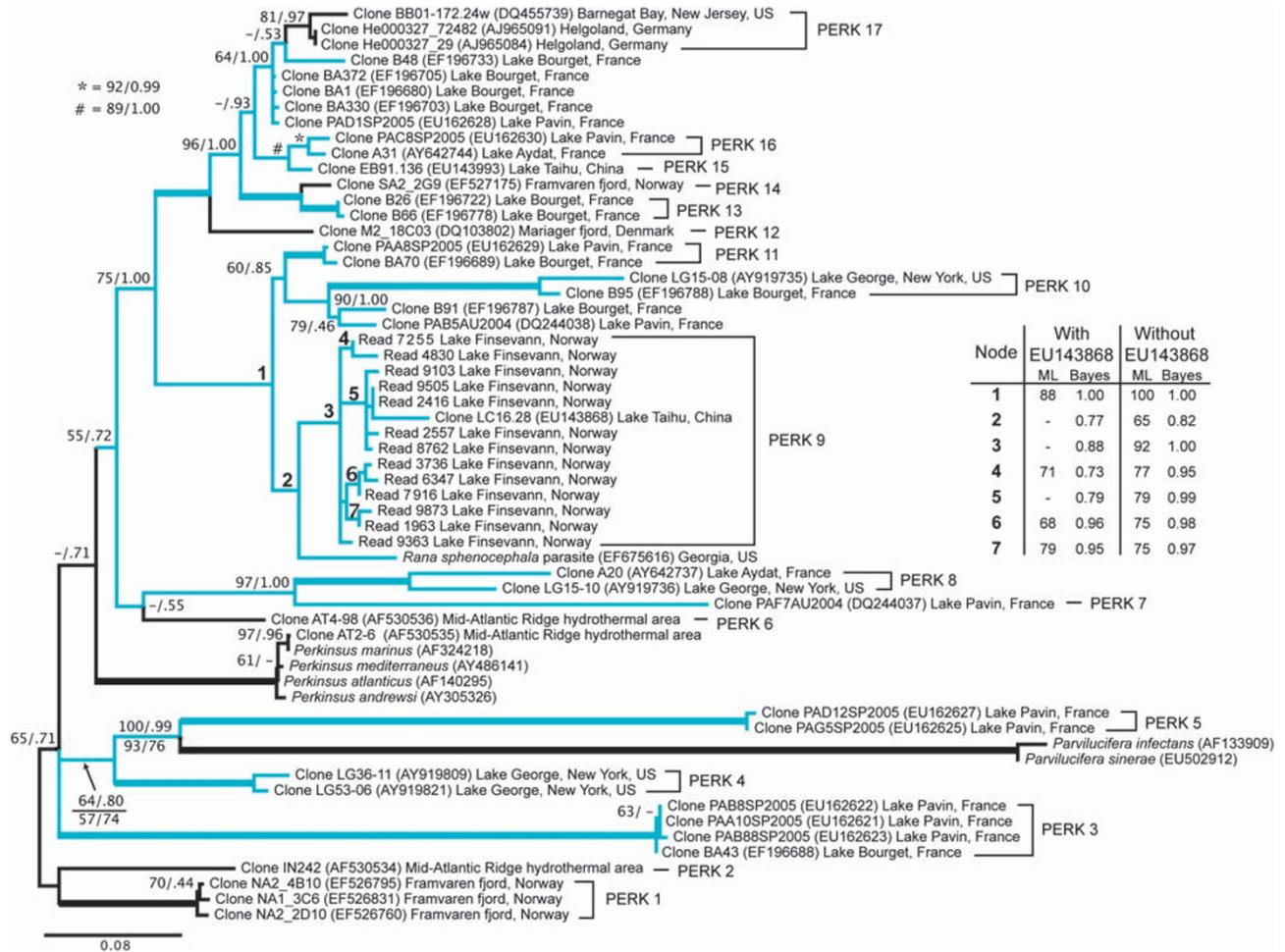


Figure 2 Bayesian phylogeny of *Perkinsea* 18S rDNA sequences inferred from an alignment consisting of 59 taxa and 1747 characters. See Figure 1 legend for description of analyses and support values. Blue lines indicate freshwater lineages and black lines indicate marine lineages. ML and Bayesian analyses were also performed without the sequence EU143868. Because of limited space, support values of the affected nodes are shown in a separate table with the corresponding nodes indicated with numbers at the nodes. The values at the nodes of *Parvilucifera* sp. and PERK 4 and 5 are showing the analyses with the fastest evolving sites removed: The values below the nodes represent ML analyses with category 8 and category 8 + 7 removed, respectively.

EU162627). To further test the impact of homoplasy and saturation, fast evolving sites were identified and removed using the AIR program package Kumar *et al.*, 2009). The site rates were calculated using the HKY85 substitution model and the tree topology obtained in the maximum likelihood analysis of alignment 3 described above. The site rates were divided into eight categories, with category 8 being the fastest evolving, and alignments were made by removing category 8 and 8 + 7. Removing category 8 left 1574 characters and removing category 8 + 7 left 1380 characters. All the phylogenetic analyses were performed on the freely available Biportal at the University of Oslo (<http://www.biportal.uio.no>).

To assess the impact of missing data in the environmental sequences, we deleted terminal regions of AL3 that contained missing nucleotides in several of the environmental sequences. As a consequence, some taxa were left with insufficient number of characters: EU143868, EU143993, EF196688 and EF196680. These were removed,

resulting in an alignment consisting of 55 taxa and 435 characters.

Testing statistically the phylogenetic separation between marine and freshwater lineages

The statistical test of the separation between marine and freshwater lineages in the inferred *Perkinsea* clade (as presented in Figure 2) was carried out using the UniFrac program (Lozupone and Knight, 2005; <http://bmf2.colorado.edu/unifrac/index.psp>). The UniFrac significance test included 100 permutations, and the *P*-value was corrected for multiple comparisons using the Bonferroni's correction.

Results and discussion

New freshwater *Perkinsea* and improved phylogeny of enigmatic dinozoan lineages

The pyrosequencing of 18S rDNA V4 amplicons generated from Lake Finsevatn generated about

10 000 reads with an average length of 237 bp. Trimming of low-quality reads and reads shorter than 150 bp left 4769 sequences for further analyses. Of these, 53 reads were categorized as highly similar to Perkinsea on the basis of BLAST searches against the NCBI nr database, resulting in 13 unique sequences after quality assessment. In addition to the new sequences, we identified 40 different 18S rDNA sequences of putative Perkinsea origin in publicly available databases, and thus, our study includes one of the largest sets of environmental sequences potentially related to *Perkinsus* and *Parvilucifera* that has been compiled to date.

Owing to the increased taxon selection in this branch of the alveolate diversity and by analyzing the sequences with the covarion evolutionary model, we obtained a high support for the clustering of *Perkinsus* and *Parvilucifera* together in the class Perkinsea (Cavalier-Smith, 2004) (Figure 1). The grouping of Perkinsea together with the MA and the dinoflagellates into the subphylum dinozoa received maximum support (100% maximum likelihood bootstrap support and 1.00 Bayesian pp), with the dinoflagellates being closest to the MA (82%, 1.00 pp). The branching order of these groups is highly congruent with recent classification of the alveolates (Cavalier-Smith, 2004), but whether the MA should be included among the dinoflagellates or as sisters to them is still unclear (see Massana *et al.*, 2008). This unresolved branching order between the dinoflagellates and the MA (Figure 1) could be due to a rapid radiation, early in the evolution of dinoflagellates (Saldarriaga *et al.*, 2004; Shalchian-Tabrizi *et al.*, 2006); multiple gene phylogenies are needed to help determine the relationships between these groups. The apicomplexans were retrieved as monophyletic (75%, 0.98 pp), whereas *Colpodella* sp. was polyphyletic and branched together with *Chromera velia* as well as freshwater environmental sequences at the base of the Myzozoa, similar to other 18S rDNA phylogenies (Leander *et al.*, 2003b; Moore *et al.*, 2008).

Perkinsea was highly supported (98%, 1.00 pp) and consisted of several environmental sequences from marine and freshwaters, as well as a sequence from an uncultured frog parasite and sequences from cultured *Perkinsus* and *Parvilucifera* species (Figure 1). A more thorough phylogenetic analysis of Perkinsea (Figure 2) showed that, in addition to the *Perkinsus* and *Parvilucifera* lineages, Perkinsea could be further subdivided into several highly supported (>95%, 1.00 pp) clades and solitary sequences, hereafter named PERK 1–17 (Figure 2). In addition, several larger assemblages were supported. *Parvilucifera* and PERK 1–5 had basal positions within Perkinsea, but their interrelationship was only weakly supported. PERK 7–17 formed a large assembly of uncultured sequences excluding *Perkinsus* and *Parvilucifera*. This assembly was almost exclusively composed of freshwater sequences and encompassed large sequence variation.

Several solitary sequences, for example, PERK 6 and 12, are excluded from other clades with high support and could belong to larger, undersampled groups. The removal of the 13 sequences generated in this study did not change the tree topology in any significant way (see Supplementary Figure 1).

The taxonomic rank of the PERK groups revealed in the 18S rDNA phylogeny is not clear. However, the branch lengths of these sequences (that is, nucleotide variation), which are considerably longer than the marine *Perkinsus* sequences, indicates that each likely represent different species and that the different subgroups and solitary sequences constitute genera or higher order taxonomic levels.

Unknown Perkinsea diversity in marine and freshwaters

The genera *Perkinsus* and *Parvilucifera* have until now been regarded as strictly marine parasites (Delgado, 1999; Noren *et al.*, 1999; Erard-Le Denn *et al.*, 2000; Villalba *et al.*, 2004; Figueroa *et al.*, 2008; Leander and Hoppenrath, 2008). However, in recent environmental surveys of freshwater lakes, the existence of taxa related to Perkinsea has been suggested (Richards *et al.*, 2005; Lefèvre *et al.*, 2008; Lepère *et al.*, 2008). Nevertheless, as the analyses did not include any sequences from the MA or statistical support estimations for the tree topologies (Richards *et al.*, 2005; Lefèvre *et al.*, 2008; Lepère *et al.*, 2008), it cannot be determined whether the sequences belong to Perkinsea or to any of the MA.

Our extensive phylogenetic analyses of Perkinsea, including virtually all related environmental 18S sequences, showed a large and previously unknown diversity of both freshwater and marine taxa (Figure 2). Most of the subgroups within Perkinsea were entirely freshwater, except PERK 1, 2, 6, 12, 14 and 17, which were marine. The sequences from the Lake Finsevatn sediments formed the subgroup PERK 9 together with a sequence from a Chinese lake. However, as this sequence had very few overlapping characters with the sequences from this study, its position in the tree given in Figure 2 was very unstable. Analyses without this sequence substantially improved the support for several of the nodes while the branching order remained essentially identical (see Figure 2 legend). PERK 9 branched off together with the only environmental sequence of known origin, a sequence from a parasite of the leopard frog, *Rana sphenoccephala* (EF675616). This fell into a large and diverse, exclusively freshwater group, including the subgroups PERK 9–11 (Figure 2). This parasite is known to cause massive killings of frogs by infecting the internal organs, especially the liver and kidneys (Green *et al.*, 2003; Davis *et al.*, 2007). There are also other reports of freshwater parasites with an outer morphology resembling *Perkinsus* (for example, a parasite of the freshwater cryptomonad *Chilomonas*

paramecium), but there are no sequences available for these species (Brugerolle, 2002, 2003). Considering the position of the *Rana* parasite among the Perkinsea sequences in the tree (Figure 2), as well as the evidence from other studies indicating that putative members of Perkinsea are infecting a wide range of species, such as mollusks, frogs, dinoflagellates and cryptomonads (Brugerolle, 2002, 2003; Green *et al.*, 2003; Villalba *et al.*, 2004; Davis *et al.*, 2007), we suspect that the entire Perkinsea is parasitic. This is very important to clarify because it would alter the present view of the role of picoeukaryotes in 'microbial loops' and the carbon cycle in aquatic systems. This view was recently presented by Lefèvre *et al.* (2008) as the 'parasite/saprotroph-dominated HF [heterotrophic flagellates] hypothesis' and should be further investigated by microscopy and fluorescence *in situ* hybridization methods.

In contrast to the MA, which have a worldwide distribution in the oceans, the Perkinsea seemed to predominate in fresh waters. Hence, each of these large putative parasitic groups seems to dominate different aquatic environments. This striking difference could potentially highlight important factors determining the dispersal and diversification of parasitic protists in aquatic environments. However, the sampling from different habitats and locations has not been equally extensive. In particular, freshwater habitats have so far been poorly sampled, and thus, more freshwater diversity may be revealed in future samplings. Furthermore, most of the environmental sequences related to Perkinsea have so far been obtained with universal PCR primers that often recover only a fraction of the total diversity. We therefore consider it likely that the total diversity of Perkinsea will be shown to be substantially larger by employing Perkinsea-specific primers in future environmental surveys.

Phylogeny of enigmatic perkinsoid species

The deep phylogenetic relationships within dinozoa have been difficult to resolve, particularly because of the rapid diversification and highly uneven evolutionary rates of many of the groups (Siddall *et al.*, 2001; Kuvardina *et al.*, 2002; Leander *et al.*, 2003a, 2003b; Cavalier-Smith and Chao, 2004; Silberman *et al.*, 2004; Groisillier *et al.*, 2006; Shalchian-Tabrizi *et al.* 2006). The increased taxon sampling of Perkinsea and application of covarion evolutionary models in this study has provided support for Perkinsea as sister to the dinoflagellates and MA. Nevertheless, the relationship between the two latter groups is still unresolved (Figure 1). It has been noted that the positions of *Perkinsus* and *Parvilucifera* in 18S rDNA phylogenies may be sensitive to the taxon sampling, and despite the increased taxon sampling here, long-branch artifacts could still affect their position (Cavalier-Smith and Chao, 2004). Therefore, to test the possibility that

the position of *Parvilucifera* in our tree (Figure 2) was a result of long-branch attraction artifacts, the phylogenetic analyses were performed both with and without the two *Parvilucifera* sequences and the sequences EU162627 and EU162625. After removal of the two latter sequences, *Parvilucifera* still grouped together with AY919821 and AY919809, and the removal of the *Parvilucifera* sequences did not have any significant impact on the overall topology (results not shown). Removing fast evolving sites and missing terminal characters in the alignment did not change the position of *Parvilucifera*, although the bootstrap support was slightly altered (Figure 2).

Recently, two parasite species have been described as potential relatives of *Perkinsus*, a parasite of the southern leopard frog (*R. sphenoccephala*) and a parasite of the Atlantic sardine (*Sardina pilchardus*), *Perkinsoide chabelardi* (Gestal *et al.*, 2006; Davis *et al.*, 2007). The *R. sphenoccephala* parasite has been suggested to be a close relative of *Perkinsus* on the basis of morphology and phylogenetic analysis of the 18S rDNA gene (Davis *et al.*, 2007). Our trees confirm that the *R. sphenoccephala* parasite belong to Perkinsea (Figure 2). In contrast, *P. chabelardi*, which has been previously determined as *Perkinsus*-like (Gestal *et al.*, 2006), is placed with high statistical support within group 1 of the MAs in our work (Figure 1), suggesting that previous classifications were incorrect (Gestal *et al.*, 2006). The placing of *P. chabelardi* in MA group 1, strengthens the view that this alveolate group is entirely parasitic (see Groisillier *et al.*, 2006; Harada *et al.*, 2006; Dolven *et al.*, 2007; Guillou *et al.*, 2008).

Marine-freshwater colonizations

In our phylogenetic reconstructions, the majority of the environmental sequences within Perkinsea fell into distinct clades according to their freshwater or marine origin (Figure 2). A UniFrac analysis of the Perkinsea clade, in which the fraction of unique branch lengths to each community against the total branch lengths between communities is measured based on a phylogenetic tree (see Lozupone and Knight, 2005), showed that this separation is statistically significant (P -value < 0.01). This suggests that only a handful of marine-freshwater cross-colonization events have occurred during the evolution of Perkinsea. A few marine sequences are placed robustly within the freshwater clades, showing that recolonizations into the marine habitat have taken place as well (Figure 2). However, five sequences originated from fjords (EF526795, EF527175, EF526760, EF526831 and DQ103802) (Anke Behnke, personal communication) and could be the result of freshwater runoff. If this is indeed the case, PERK 1 would be entirely freshwater, and the large clade comprising PERK 12–17 would include only a single marine clade; hence, the number of colonization events would be reduced

considerably, and increase the possibility that Perkinsea originated in a freshwater habitat. Two marine sequences were sampled at the mid-Atlantic ridge (AF530536 and AF530534) and can hardly be the result of freshwater runoff. These two mid-Atlantic ridge sequences, together with the sequence DQ103802, have a solitary position within Perkinsea and could represent a larger, undersampled taxa hosting a greater diversity.

The limited transitions between marine and freshwater lineages are in great contrast to the wide geographic distribution of species within each type of habitat. Among the freshwater subgroups, sequences from as different locations as France, China and the United States are grouped together showing that dispersal between freshwater habitats over long distances has taken place without recolonizing the marine habitat.

Overall, our results on the putative parasites of Perkinsea show only a few successful cross-colonization events between marine and fresh waters during the evolutionary diversification of this group, implying that the biogeochemical differences between marine and fresh waters represent a strong barrier against cross-colonizations, concordant with recent results from other groups of free-living protists (Logares *et al.*, 2009).

Test of new 18S primers and pyrosequencing

Although our main goal with pyrosequencing of the lake Finsevatn sediments was to identify Perkinsea species, we could also uncover a large diversity of other eukaryote groups. In fact, all eukaryotic supergroups except Excavata could be clearly identified in our library (for description of supergroups, see Burki *et al.*, 2008), as well as several sequences highly similar to enigmatic lineages that traditionally have not been placed in any of these supergroups, such as Apusozoa and Telonemia (Figure 3). Altogether, the data show that the protocol presented here for PCR of the variable V4 18S rDNA region and 454 pyrosequencing is sensitive enough to detect even minute cell numbers, and hence are suitable for surveys of the overall eukaryote diversity. However, it should be noted that the abundance of the different groups was variable and could indicate that some groups are preferentially amplified over others; the chromalveolates are by far the most abundant group in the library, whereas only less than 1% belong to Amoebozoa.

Recently, the V9 region, which is typically about 150 bp long, was suggested as a suitable marker for diversity surveys of eukaryotes, and methods optimized for earlier generations of the 454 pyrosequencing technology has been used with success (Amaral-Zettler *et al.*, 2009; Stoeck *et al.*, 2009). As the PCR-based methods can be biased, it is likely that both the V4 and V9 regions could recover somewhat different diversity. It is a question

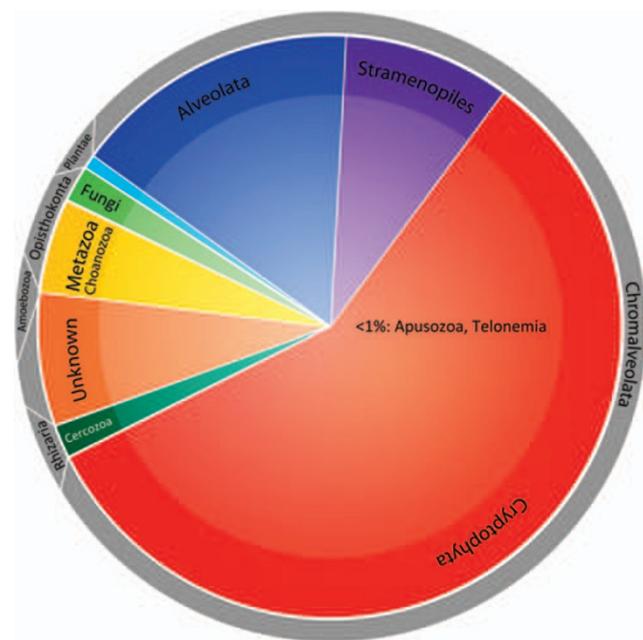


Figure 3 Pie chart showing the phylogenetic association of 4769 reads longer than 150bp, as determined by BLAST searches against the NCBI nr database. Abundance of groups: cryptophyta, 57.50%; stramenopiles, 9.26%; alveolata, 15.92%; plantae, 0.77%; cercozoa, 1.80%; amoebozoa, 0.11%; fungi, 1.99%; choanozoa, 0.09%; metazoa, 5.21%; apusozoa, 0.02%; telonemia, 0.04%; unknown 7.29%.

whether the V4 or V9 region is most suitable for aberrant 18S genes, as often found in Foraminifera and other fast evolving species. A major advantage of the longer V4 region, which can now be covered with the Titanium upgrade, is the longer sequence lengths that allow for a more reliable taxon identification and more rigorous phylogenetic analyses. The many distantly related groups revealed in the data are likely to be real because of the substantial difference between the sequences that characterize the eukaryotic supergroups; however, because the quality assessment of sequences produced by the 454 and Sanger differs, it has been shown that the 454 data can overestimate the actual diversity in the environment (Quince *et al.*, 2009).

Acknowledgements

We thank the Bioportal team for implementing bioinformatics applications on the Bioportal (<http://www.bioportal.uio.no>). This work has been supported by the University of Oslo grants to KST, PhD fellowship to JB and travel funds to RL. The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) via Uppsala Microbiomics Centre (UMC; <http://www.microbiomics.se>) has given financial support to RL.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W *et al.* (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 parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4**: e6372.
- Behnke A, Bunge J, Barger K, Breiner H-W, Alla V, Stoeck T. (2006). Microeukaryote community patterns along an O₂/H₂S gradient in a supersulfidic anoxic fjord (Framvaren, Norway). *Appl Environ Microbiol* **72**: 3626–3636.
- Berney C, Fahrni J, Pawlowski J. (2004). How many novel eukaryotic ‘kingdoms’? Pitfalls and limitations of environmental DNA surveys. *BMC Biol* **2**: 13.
- Brugerolle G. (2002). *Cryptophagus subtilis*: a new parasite of cryptophytes affiliated with the Perkinsozoa lineage. *Eur J Protistol* **37**: 379–390.
- Brugerolle G. (2003). Apicomplexan parasite *Cryptophagus* renamed *Rastrimonas* gen. nov. *Eur J Protistol* **39**: 101–101.
- Burki F, Shalchian-Tabrizi K, Pawlowski J. (2008). Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biol Lett* **4**: 366–369.
- Cavalier-Smith T. (2004). Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton. In: Hirt RP, Horner DS (eds). *Organelles Genomes and Eukaryotic Phylogeny*. CRC Press: New York, pp 75–108.
- Cavalier-Smith T, Chao EE. (2004). Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). *Eur J Protistol* **40**: 185–212.
- Cavalier-Smith T, Lewis R, Chao EE, Oates B, Bass D. (2009). *Helkesimastix marina* n. sp. (Cercozoa: Sainouroidea superfam. n.) a Gliding zooflagellate of novel ultrastructure and unusual ciliary behaviour. *Protist* **160**: 452–479.
- Chambouvet A, Morin P, Marie D, Guillou L. (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* **322**: 1254–1257.
- Davis AK, Yabsley MJ, Keel MK, Maerz JC. (2007). Discovery of a novel alveolate pathogen affecting southern leopard frogs in Georgia: description of the disease and host effects. *Ecohealth* **4**: 310–317.
- Delgado M. (1999). A new ‘diablillo parasite’ in the toxic dinoflagellate *Alexandrium catenella* as a possibility to control harmful algal blooms. *Harmful Algae News* **19**: 1–3.
- Dolven JK, Lindqvist C, Albert VA, Bjorklund KR, Yuasa T, Takahashi O *et al.* (2007). Molecular diversity of alveolates associated with neritic North Atlantic radiolarians. *Protist* **158**: 65–76.
- Erard-Le Denn E, ChrÉtienne-Dinet MJ, Probert I. (2000). First report of parasitism on the toxic dinoflagellate *Alexandrium minutum* Halim. *Estuar Coast Shelf Sci* **50**: 109–113.
- Figueroa RI, Garcés E, Massana R, Camp J. (2008). Description, host-specificity, and strain selectivity of the dinoflagellate parasite *Parvilucifera sinerae* sp. nov. (Perkinsozoa). *Protist* **159**: 563–578.
- Gestal C, Novoa B, Posada D, Figueras A, Azevedo C. (2006). *Perkinsoide chabelardi* n. gen., a protozoan parasite with an intermediate evolutionary position: possible cause of the decrease of sardine fisheries? *Environ Microbiol* **8**: 1105–1114.
- Green DE, Feldman SH, Wimsatt J. (2003). Emergence of a Perkinsus-like agent in anuran liver during die-offs of local populations: PCR detection and phylogenetic characterization. *Proc Am Assoc Zoo Vet*, 120–121.
- Groissillier A, Massana R, Valentin K, Vaulot D, Guillou L. (2006). Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquat Microb Ecol* **42**: 277–291.
- Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R *et al.* (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ Microbiol* **10**: 3349–3365.
- Harada A, Ohtsuka S, Horiguchi T. (2006). Species of the parasitic genus *duboscquella* are members of the enigmatic marine alveolate group I. *Protist* **158**: 337–347.
- Katoh K, Kuma K-I, Toh H, Miyata T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucl Acids Res* **33**: 511–518.
- Kumar S, Skjaeveland A, Orr R, Enger P, Ruden T, Mevik B-H *et al.* (2009). AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics* **10**: 357.
- Kuvarina ON, Leander BS, Aleshin VV, Myl’nikov AP, Keeling PJ, Simdyanov TG. (2002). The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans. *J Eukaryot Microbiol* **49**: 498–504.
- Leander BS, Clopton RE, Keeling PJ. (2003a). Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and beta-tubulin. *Int J Syst Evol Microbiol* **53**: 345–354.
- Leander BS, Kuvarina ON, Aleshin VV, Myl’nikov AP, Keeling PJ. (2003b). Molecular phylogeny and surface morphology of *Colpodella edax* (Alveolata): insights into the phagotrophic ancestry of apicomplexans. *J Eukaryot Microbiol* **50**: 334–340.
- Leander BS, Hoppenrath M. (2008). Ultrastructure of a novel tube-forming, intracellular parasite of dinoflagellates: *Parvilucifera prorocentri* sp. nov. (Alveolata, Myzozoa). *Eur J Protistol* **44**: 55–70.
- Lefèvre E, Roussel B, Amblard C, Sime-Ngando T. (2008). The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS ONE* **3**: e2324.
- Lepère C, Domaizon I, Debroas D. (2008). Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Appl Environ Microbiol* **74**: 2940–2949.
- Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K. (2009). Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* **17**: 414–422.
- López-García P, Moreira D. (2008). Tracking microbial biodiversity through molecular and genomic ecology. *Res Microbiol* **159**: 67–73.
- Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

- Maddison D, Maddison W. (2000). *MacClade 4: Analysis of Phylogeny and Character Evolution*, 4 edn. Sinauer Associates: Sunderland, MA.
- Massana R, Karniol B, Pommier T, Bodaker I, Oded B. (2008). Metagenomic retrieval of a ribosomal DNA repeat array from an uncultured marine alveolate. *Environ Microbiol* **10**: 1335–1343.
- Moore RB, Obornik M, Janouskovec J, Chrudimsky T, Vancova M, Green DH et al. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**: 959–963.
- Noren F, Moestrup O, Rehnstam-Holm AS. (1999). *Parvilucifera infectans* Noren et Moestrup gen. et sp nov (Perkinsozoa phylum nov.): a parasitic flagellate capable of killing toxic microalgae. *Eur J Protistol* **35**: 233–254.
- Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM et al. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**: 639–641.
- Richards TA, Vepritskiy AA, Gouliamova DE, Nierzwicki-Bauer SA. (2005). The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. *Environ Microbiol* **7**: 1413–1425.
- Ronquist F, Huelsenbeck JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Saldarriaga JF, Max' Taylor FJR, Cavalier-Smith T, Menden-Deuer S, Keeling PJP. (2004). Molecular data and the evolutionary history of dinoflagellates. *Eur J Protistol* **40**: 85–111.
- Shalchian-Tabrizi K, Minge MA, Cavalier-Smith T, Nedrek-lepp JM, Klaveness D, Jakobsen KS (2006). Combined Heat Shock Protein 90 and Ribosomal RNA Sequence Phylogeny Supports Multiple Replacements of Dinoflagellate Plastids. *J Eukaryot Microbiol* **53**: 217–224.
- Siddall ME, Reece KS, Nerad TA, Bureson EM. (2001). Molecular determination of the phylogenetic position of a species in the genus colpodella (Alveolata). *Am Mus Novit* **3314**: 1–12.
- Silberman JD, Collins AG, Gershwin LA, Johnson PJ, Roger AJ. (2004). Ellobiopsids of the genus *Thalassomyces* are alveolates. *J Eukaryot Microbiol* **51**: 246–252.
- 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 M, Chistoserdov A et al. (2009). Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biology* **7**: 72.
- Villalba A, Reece KS, Ordás MC, Casas SM, Figueras A. (2004). Perkinsosis in molluscs: a review. *Aquat Living Resour* **17**: 411–432.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)