environmental microbiology

Environmental Microbiology (2014) doi:10.1111/1462-2920.12591

Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems

Marianne Simon, Ludwig Jardillier, Philippe Deschamps, David Moreira, Gwendal Restoux, Paola Bertolino, Purificación López-García*

Unité d'Ecologie, Systématique et Evolution, CNRS UMR 8079, Université Paris-Sud, 91405 Orsay, France.

Summary

Although inland water bodies are more heterogeneous and sensitive to environmental variation than oceans, the diversity of small protists in these ecosystems is much less well known. Some molecular surveys of lakes exist, but little information is available from smaller, shallower and often ephemeral freshwater systems, despite their global distribution and ecological importance. We carried out a comparative study based on massive pyrosequencing of amplified 18S rRNA gene fragments of protists in the 0.2–5 μm size range in one brook and four shallow ponds located in the Natural Regional Park of the Chevreuse Valley, France. Our study revealed a wide diversity of small protists, with 812 stringently defined operational taxonomic units (OTUs) belonging to the recognized eukaryotic supergroups (SAR – Stramenopiles, Alveolata, Rhizaria – Archaeplastida, Excavata, Amoebozoa, Opisthokonta) and to groups of unresolved phylogenetic position (Cryptophyta, Haptophyta, Centrohelida, Katablepharida, Telonemida, Apusozoa). Some OTUs represented deep-branching lineages (Cryptomycota, Aphelida, Colpodellida, Tremulida, clade-10 Cercozoa, HAP-1 Haptophyta). We identified several lineages previously thought to be marine including, in addition to MAST-2 and MAST-12, already detected in freshwater, MAST-3 and possibly MAST-6. Protist community structures were different in the five ecosystems. These differences did not correlate with geographical distances, but seemed to be influenced by environmental parameters.

Received 13 July, 2014; accepted 5 August, 2014. *For correspondence. E-mail puri.lopez@u-psud.fr; Tel. +33 169 157 608; Fax +33 169 154 697.

Introduction

Aquatic ecosystems occupy most of the Earth surface. Oceans alone cover around 71% of that surface (Costanza, 1999). Lakes, ponds and reservoirs cover more than 3% (nearly 4.5 million km^2) of non-oceanic regions (Downing *et al*., 2006). Although comparatively smaller, these freshwater systems offer a large array and diversity of ecological niches including various trophic levels, light accessibility, temperature and oxygen concentrations. This is especially true for small and shallow lentic inland ecosystems (sizes from 0.001 to 0.1 km^2), which are widespread, varied and numerous, corresponding to more than 99% of the total number of lakes on Earth (Downing *et al*., 2006). Microbial communities dominate aquatic ecosystems and their activity has profound impact at global scales, being largely implicated in carbon fixation (Li, 1994; Jardillier *et al*., 2010) and climate regulation (Simó, 2001). Within these communities, microbial eukaryotes play key roles in nutrient cycling acting as photosynthesizers, heterotrophs (predators, parasites) or mixotrophs (Caron, 1994; Zubkov and Tarran, 2008; Jardillier *et al*., 2010; Massana, 2011).

Small eukaryotes (< 20 μm) have been known to constitute a non-negligible part of aquatic microbial communities in both freshwater and marine systems for a long time (Johnson and Sieburth, 1982; Corpe and Jensen, 1992). Based on their cell size, small protists were initially classified in nanoeukaryotes (cells between 2 and 20 μm in diameter) and picoeukaryotes (cells ≤ 2 μm). In the last two centuries, many small eukaryotic species have been described, including phototrophs such as prasinophytes and other chlorophytes (Knight-Jones and Walne, 1951; Stockner, 1988; Guillou *et al*., 1999), which suggested a potentially significant role in primary production (Johnson and Sieburth, 1982), but also heterotrophs (Fenchel, 1982; Patterson and Larsen, 1991). Indeed, the ecological importance of heterotrophic nanoflagellates as predators has long been acknowledged (Fenchel, 1982; Wright and Coffin, 1984). However, being too small to show easily identifiable, unambiguous morphological differences, their true diversity remained largely inaccessible and their taxonomy poorly studied or oversimplified (many of these tiny protists were simply classed as *incertae sedis*).

In the past 15 years, the use of molecular methods based on 18S rRNA gene analysis has largely overcome that problem, allowing studying the phylogenetic diversity and distribution of small protists at unprecedented scales. Most such studies largely explored oceanic systems, including surface waters (Diez *et al*., 2001; Moon-van der Staay *et al*., 2001), the deep sea (López-García *et al*., 2001) and the coastal regions (Massana *et al*., 2004a; Romari and Vaulot, 2004; Cheung *et al*., 2010). They revealed a huge protist diversity encompassing members of all recognized eukaryotic supergroups (SAR – Stramenopiles, Alveolata, Rhizaria – Archaeplastida, Excavata, Amoebozoa and Opisthokonta) as well as lineages of uncertain position in the eukaryotic tree, such as Haptophyta, Cryptophyta or Picozoa (López-García and Moreira, 2008; Massana, 2011; Seenivasan *et al*., 2013; Moreira and López-García, 2014). Molecular surveys also allowed the discovery of lineages previously unknown in spite of their abundance, such as new clades affiliated to alveolates (Groups I and II), which turned out to be members of the classical Syndiniales (Groisillier *et al*., 2006; Harada *et al*., 2007; Guillou *et al*., 2008) and stramenopiles (the MAST clades, for marine stramenopiles) (Diez *et al*., 2001; López-García *et al*., 2001; Moon-van der Staay *et al*., 2001; Stoeck *et al*., 2003; Massana *et al*., 2004b). Thought for a long time to be exclusively marine, the MAST groups remain poorly known. In a few cases a correspondence has been found between particular organisms and their environmental sequences (e.g. the colonial protist *Solenicola setigera* and the MAST-3; Gómez *et al*., 2011). However, many other MAST lineages have not yet been linked to any cultured or described organisms (Massana *et al*., 2014), even though fluorescent in situ hybridization labelling has provided some hints on their morphology, life style and ecology (Massana *et al*., 2002; 2006).

The molecular exploration of very small protists in freshwater began slightly later than in the oceans and also revealed a large diversity (Lefranc and Thénot, 2005; Richards, 2005; Šlapeta *et al*., 2005). In lakes, alveolates (especially ciliates and Perkinsozoa, by contrast to the Syndiniales-Duboscquellids-dinoflagellate dominance in marine systems), stramenopiles, cryptophytes and fungi were found to be abundant, but protists belonging to the Archaeplastida, Rhizaria and Cercozoa or groups of uncertain affiliation were also diverse (Lefranc and Thénot, 2005; Šlapeta *et al*., 2005; Lepère *et al*., 2008; Zhao *et al*., 2011; Mangot *et al*., 2012). In some cases, a significant part of the detected diversity was composed only of environmental sequences without known close relatives in databases (Lefèvre *et al*., 2008; Mangot *et al*., 2012; Triadó-Margarit and Casamayor, 2012). Some groups were observed in both oceans and freshwater systems whereas others seem, according to current knowledge, to be specific to marine (e.g. Syndiniales; Guillou *et al*., 2008) or freshwater (e.g. HAP-1 lineage of haptophytes; Šlapeta *et al*., 2005; Shalchian-Tabrizi *et al*., 2011) environments. Even within lineages present both in marine and freshwater systems, 18S rRNA sequences obtained from freshwater samples often form phylogenetic clades distinct from those of oceanic sequences, suggesting a limited number of, mostly ancient, freshwater colonization events followed by radiations (Logares *et al*., 2009). That scarcity of marine-freshwater transitions has been observed for many small eukaryote groups such as the perkinsids (Bråte *et al*., 2010), haptophytes (Shalchian-Tabrizi *et al*., 2011; Simon *et al*., 2013) or MAST lineages (Massana *et al*., 2014), and could be explained by the difficulty of crossing the salinity barrier (Logares *et al*., 2009). However, freshwater aquatic systems remain largely understudied and massive high-throughput sequencing techniques have been applied to very few among them. Therefore, it might be possible that failure to detect some of these lineages is due to undersampling. Furthermore, the vast majority of protist molecular diversity studies in freshwater have been conducted in lakes (e.g. Richards, 2005, Lepère and Boucher, 2006; Lepère *et al*., 2008). However, smaller systems like ponds or brooks have been overlooked, despite their abundance, distribution and ecological importance. Yet, previous analyses suggest that they host several lineages undetected in marine environments or lakes (Šlapeta *et al*., 2005; Simon *et al*., 2013). In addition, because they are shallow and small, this kind of freshwater systems may display very different local physico-chemical conditions and, hence, they constitute ideal models to test whether local environmental selection is more influential than geographic distance in determining community composition, which remains controversial (Green *et al*., 2004; Martiny *et al*., 2006). A recent study on lakes suggest that geographic distance might be important to explain protist biogeography (Lepère *et al*., 2013), but lakes have much more buffering capacity than small shallow systems.

In this work, we have investigated the diversity of small eukaryotes (essentially the 0.2–5 μm cell size fraction) in a set of shallow freshwater environments by amplification of 18S rRNA gene fragments and direct high-throughput 454-pyrosequencing. We selected one brook and four ponds located in the same geographic area (the Natural Regional Park of the Chevreuse Valley, France) but differing in size, depth and physico-chemical conditions. The objectives of our work were threefold: (i) describing protist diversity at unprecedented depth in this kind of habitats, (ii) checking whether increasing sequence depth leads to the discovery in freshwater systems of eukaryotes previously thought to be exclusively marine and (iii) testing

which parameters (geographical distance versus physico-chemical characteristics) determine eukaryotic microbial community structure.

Results

To study protist diversity in five small and shallow freshwater ecosystems, we selected one brook and four shallow ponds in the Natural Regional Park characterized by different physico-chemical parameters and trophic status (Table 1; Fig. S1). The Mare Gabard is a small pond located in the middle of the forest (Fig. S1). It smelled strongly of H_2S when sediments were stirred and had high phosphate concentrations (0.15 mg l^{-1}) but the lowest pH, conductivity and TDS (total dissolved solids) values. Saint Robert pond is the most anthropized system selected because of its location in a hamlet (Fig. S1). It contained very high ammonia (1.22 mg l^{-1}) and, to a lesser extent, also nitrite concentrations (Table 1, Fig. 1), likely influenced by the permanent presence of a sizeable duck population. La Claye pond is part of a complex of several ponds on ancient peat bog substrate, with a dense population of eagle ferns in the surroundings. Lying on acidic and organic-rich soils, it was characterized by high concentrations of dissolved organic carbon (DOC; 36.3 mg l^{-1}). The Etang des Vallées is a nearly 1.5 m deep pond and, contrary to other sampled systems, was supersaturated in oxygen (116.8%; Table 1). It also had high temperature at the time of sampling (13.1°C), and high concentrations of nitrate (5.1 mg l^{-1}) as compared with the other systems. The Ru Sainte Anne is the brook and had the highest amounts of total dissolved solids (746 mg l^{-1}) possibly due to sediment input in a very shallow (~ 10 cm) system of running waters (Table 1). Principal component analysis (PCA) of the different physico-chemical parameters measured showed that, despite their close proximity (distances between these systems varied from 2 to 9 km; Fig. S1), the ponds and brook were clearly distinct from each other (Fig. 1). This made of this set of shallow aquatic ecosystems a good model to study whether geographic proximity is more influential than physico-chemical parameters in determining community composition.

Overall protist community composition

The composition of protist communities was estimated for the five selected shallow freshwater systems based on 454-pyrosequenced 18S rDNA fragments amplified from DNA of plankton of the 0.2–5 μm cell diameter fraction. In addition, protist diversity in the size fraction 5–30 μm was studied for the largest sampled ecosystem, the Etang des Vallées. Replicate samples were included in all cases. We described protist diversity in these systems from a total of 146 549 quality-filtered reads using highly stringent criteria to avoid sequence artefacts and chimeras. The reads from the 12 different samples (the 0.2–5 μm cell fraction of five systems with replicas plus the two replicas of the 5–30 μm for the Etang des Vallées) were treated together in order to define operational taxonomic units (OTUs) that could be fully compared among samples. All those sequences grouped in 812 OTUs defined using a cut-off value of 98% sequence identity. From these, 768 OTUs (128 661 reads) were detected at least in the 0.2–5 μm size fraction, the remaining 44 OTUs corresponding to protists found only in the larger 5–30 μm at the Etang de Vallées (Table 1). The different OTUs were then assigned to known taxonomic groups based on sequence similarity, which revealed a wide phylogenetic diversity of protists in general and of small eukaryotes in particular. Our stringently defined OTUs affiliated to the major recognized eukaryotic supergroups SAR (Stramenopiles, Alveolata, Rhizaria), Archaeplastida, Excavata, Amoebozoa and Opisthokonta (López-García and Moreira, 2008; Adl *et al.*, 2012) and to several lineages of unresolved phylogenetic position such as Cryptophyta, Haptophyta and Apusozoa (Fig. 2).

To evaluate the reliability of the community composition determined for each ecosystem at this stage, we compared the pairwise Bray–Curtis distances between the two replicates for each ecosystem. These were very small (0.26 on average, min. 0.12, max. 0.36) as compared with distances between libraries from distinct ecosystems (0.84 on average, min. 0.44, max. 0.99). Replicates grouped together in both non-metric multidimensional scaling (NMDS; Fig. S2) and Correspondence (CoA) analyses (Fig. 3), revealing their high similarity in OTU composition.

The composition of protist communities from the small size fraction samples (0.2–5 μm) differed greatly between ecosystems (Fig. 2). First, richness and diversity indexes were highly variable (Table 1). Samples from Ru Sainte Anne and Etang des Vallées were the richest and the most diverse, whereas La Claye was highly dominated by few abundant OTUs (evenness = 0.34–0.27 in replicates). La Claye and Saint Robert appeared relatively close in NMDS plots (Fig. S2) and clustered close together in CoA plots (Fig. 3). These two systems displayed relatively similar physico-chemical parameters, notably oxygen content and conductivity, and low diversity as compared with the other ecosystems (Table 1). The similarity between La Claye and St Robert community composition was likely influenced by the fact that they were largely dominated by the same cryptophyte OTU (Fig. 2; see below).

Second, the relative abundance of taxonomic groups varied between ecosystems, even though OTUs affiliated to Cryptophyta, Stramenopiles (or Heterokonta) and Alveolata accounted for the majority of sequences in all

Table 1. Characteristics of freshwater systems studied and diversity and richness estimates for the different samples.

Sampled ecosystem	Gabard (MG)	Saint Robert (SR)	Etang des Vallées (EV)	Sainte Anne (RSA)	La Claye (LC)
Ecosystem type	Forest pond	Village pond	Large pond	Forest brook	Pond on peaty soil
GPS coordinates	48°39′15.83″N 1°55′20.26″E	48°39′54.82″N 1°56′45.28″E	48°41′23,0″N 001°54′59,2″E	48°36′45.91″N 1°58′16.61″E	48°36′31.72″N 1°56′17.33″E
Approximate surface (m^2)	850 (210 × 75 m)	495 (20 × 28 m)	12 880 (210 × 75 m)	1 m width at sampling point	265 (24 × 10 m)
Approximate depth (cm)	50	50	150	20	60
Sampling Date	April 5, 2012	April 5, 2012	March 30, 2012	April 6, 2012	April 6, 2012
Water temperature (°C)	9.0	10.7	13.1	7.1	7.8
pH	6.6	7.2	7.31	7.36	7
Oxygen (%)	78.9	58.9	116.8	72.8	61.2
Conductivity (μS cm^{-1})	81.4	531	288	746	520
TDS (mg l^{-1})	81	531	288	746	520
Chlorophyll (μg l^{-1})	74.9	61.9	44.7	2.6	10.7
NO_3^- (mg l^{-1})	1.41	1.17	5.08	1.45	2.02
NO_2^- (mg l^{-1})[a]	~0	0.054	0.047	0.010	~0
NH_3 (mg^{-1})	0.02	1.22	0.02	0.03	0.04
DOC (mg l^{-1})	19.0	15.7	9.8	17.7	36.3
PO_4^{3-} (mg l^{-1})	0.15	0.03	0.03	0.03	0.03
Size fraction studied	0.2–5 μm	0.2–5 μm	5–30 μm 0.2–5 μm	0.2–5 μm	0.2–5 μm

Sample Name[b]	MG25	MG25b	SR25	SR25b	EV34 (5–30 μm)	EV34b (5–30 μm)	EV33	EV33c	RSA25	RSA25b	LC25	LC25b
Number of reads before filtering	15 020	5110	85 803	6724	28 418	14 168	22 314	7938	37 766	7140	21 283	14 215
Number of quality-filtered reads	10 616	4197	42 034	4506	10 982	6906	17 670	3947	19 652	4243	11 191	10 605
Observed number of OTUs	76	55	15	37	147	132	177	87	427	198	66	68
Richness[c] (Standard error)	67.9 (2.37)	55.0 (0.15)	11.4 (1.06)	36.8 (0.44)	134.4 (2.94)	126.9 (2.07)	148.5 (3.99)	87.0 (0.00)	313.7 (7.52)	197.8 (0.44)	57.9 (2.34)	57.0 (2.67)
Diversity (Simpson Index)	0.86	0.80	0.75	0.68	0.96	0.95	0.95	0.95	0.92	0.91	0.50	0.38
Evenness	0.58	0.56	0.61	0.45	0.75	0.72	0.69	0.78	0.64	0.7	0.34	0.27

a. Nitrite concentrations in Gabard and La Claye ponds were under the kit detection limit of 0.006 mg l^{-1} and were set as 0 in analysis.
b. Samples correspond to 0.2–5 μm cell-size fractions except stated otherwise. Replicate samples for the same site are labeled "b".
c. Expected number of OTUs in random subsamples of the size of the smallest sequence library (3947 reads in EV33c).
Replicate samples are indicated by a small case letter after the sample name.
DOC, dissolved organic carbon; TDS, total dissolved solids.

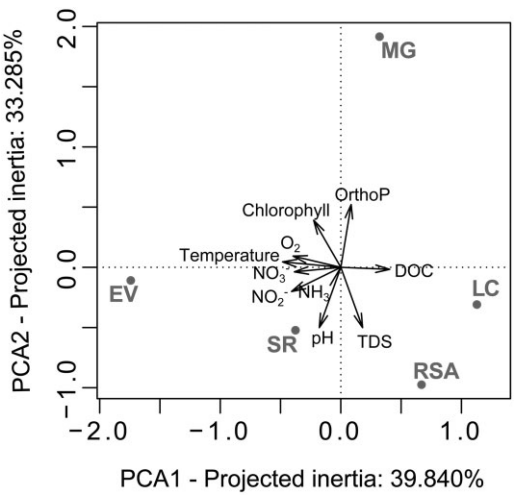


Fig. 1. Principal component analysis (PCA) plot of the measured physico-chemical parameters. Sampled ecosystems appear in grey (dots) and physico-chemical parameters in black. MG, Mare Gabard; EV, Etang des Vallees; LC, La Claye; RSA, Ru Sainte Anne; SR, Saint Robert. TDS, total dissolved solids; DOC, dissolved organic carbon; OrthoP, orthophosphate.

libraries (Figs 2 and 4). Cryptophyte sequences were remarkably abundant, representing up to 49% and 53% of the total number of reads and reads of the 0.2–5 μm size fraction respectively. Cryptophytes were the dominant group in Saint Robert and La Claye (around 75% of reads in those samples). They were also the dominant group in the brook Sainte Anne (between 32–44% of reads) and were the second most abundant group in the two other systems (Fig. 2, Table S1). Most cryptophyte reads belonged to a unique OTU affiliated to *Cryptomonas curvata* (OTU_52; Figs 4 and S3). Stramenopiles constituted the second most abundant group in our systems and represented 25% of all reads and nearly 24% of the reads from the 0.2–5 μm size fraction samples (Fig. 2, Table S1). Within this supergroup, Chrysophyceae were abundant in all samples. However, the most abundant stramenopile groups in Mare Gabard and Ru Sainte Anne were, respectively, Synurophyceae and Bacillariophyceae, two lineages producing silica skeletons or scales. Oomycetes were also relatively abundant in the brook Sainte Anne. Alveolates constituted the third most abundant supergroup in our study. It was the dominant group in Etang des Vallées (32–41% of reads in replicates

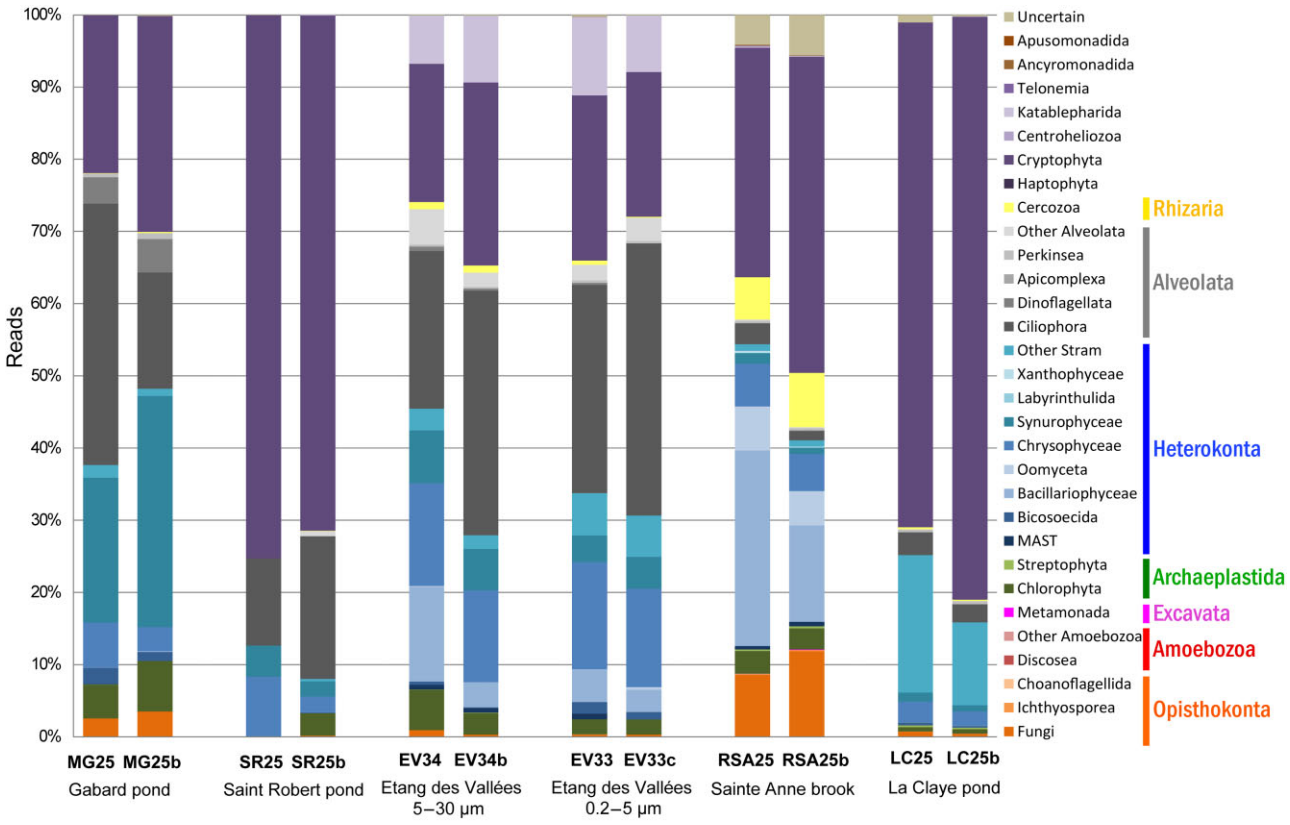


Fig. 2. Histogram showing the relative proportion of 18S rRNA gene amplicon reads assigned to high-rank taxa in the five shallow ecosystems studied. Replicate samples are labelled as 'b' or 'c'. In all cases, the distribution corresponds to protists in the 0.2–5 μm size range, except in the Etang des Vallées, where the 5–30 μm size fraction was additionally analysed.

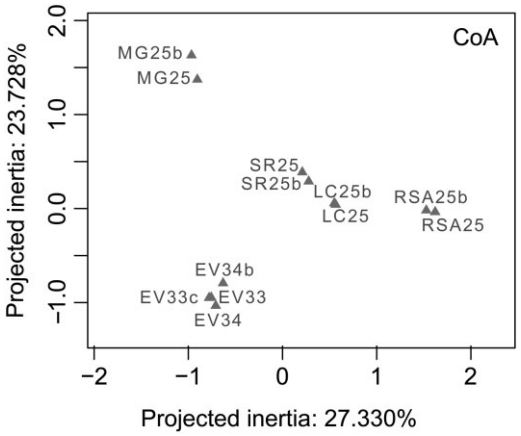


Fig. 3. Correspondence analysis (CoA) plot showing protist community composition similarities and differences among the five ecosystems studied. MG25 and MG25b, Mare Gabard; LC25 and LC25b, La Claye; SR25 and SR25b, Saint Robert; RSA25 and RSA25b, Ru Saint Anne; EV33 and EV33c, Etang des Vallées (0.2–5 μm size samples). EV34 and 34b, Etang des Vallées (5–30 μm size samples).

of the 0.2–5 μm size fraction, and similar values for the larger size fraction analysed) and one of the dominant groups in Mare Gabard (22–40% of reads in duplicate samples). It was also a major component in Saint Robert pond (12–20%). However, they represented only a small proportion of the taxa detected in the brook Sainte Anne and La Claye pond (less than 5% of reads; Fig. 2, Table S1). Among alveolates, ciliates were the most represented in all samples; although OTUs affiliated to dinoflagellates and other alveolates had occasionally significant proportions.

Along with these three abundant supergroups, members of the Opisthokonta, Amoebozoa, Excavata, Archaeplastida, Rhizaria and of groups of uncertain position in the eukaryotic tree were also detected. Because we purposefully used general eukaryotic primers biasing against Metazoa, opisthokonts were essentially represented by fungi, although choanoflagellates and ichthyosporeans were also detected. The highest abundance of fungi was observed in Mare Gabard and Ru Sainte Anne, but they generally represented less than 10% of reads. Rhizarian OTUs were retrieved in all ecosystems and were only represented by cercozoans. They were not abundant except for the Ru Saint Anne, where they reached 6–8% of the reads. Katablepharid sequences represented as much as 8–11% of reads from the 0.2–5 μm fraction in the Etang des Vallées but were less abundant elsewhere (0.03% on average in each small size fraction sample). Archaeplastida (streptophytes and, mainly, chlorophytes) were detected in all ecosys-

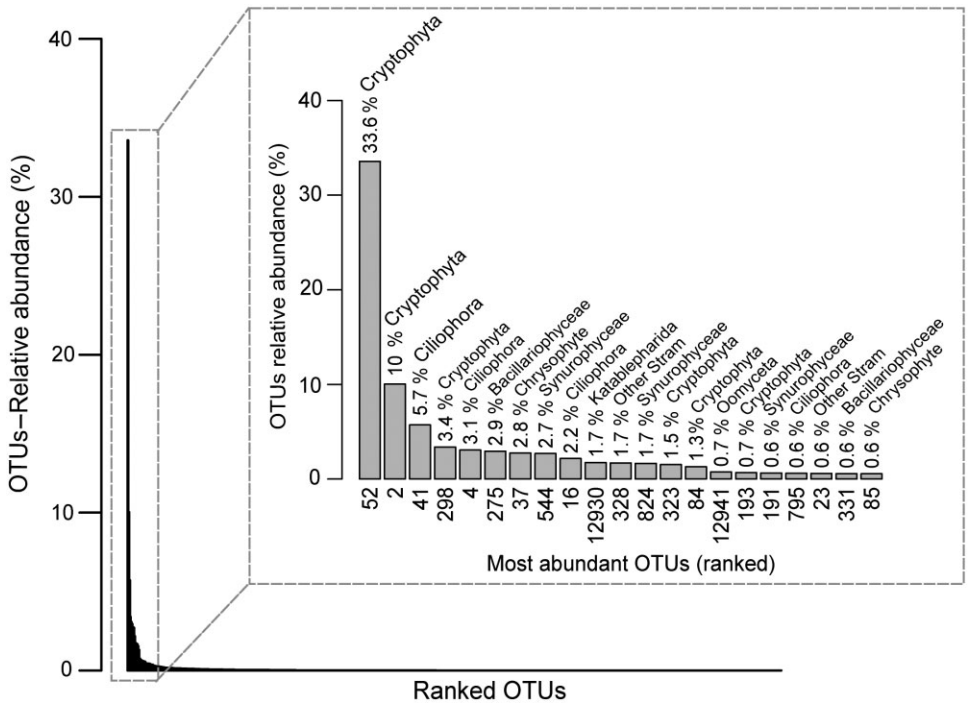


Fig. 4. Rank abundance curve for the total 768 OTUs detected collectively in the 0.2–5 μm size fraction of the five shallow freshwater systems studied. Sequence data from all samples were pooled to define OTUs with high stringency. The relative abundance of protist OTUs representing more than 0.5% of the total number of reads are shown in the inset. The identity number of the respective OTUs and their taxonomic affiliation are shown below and above the histogram bars.

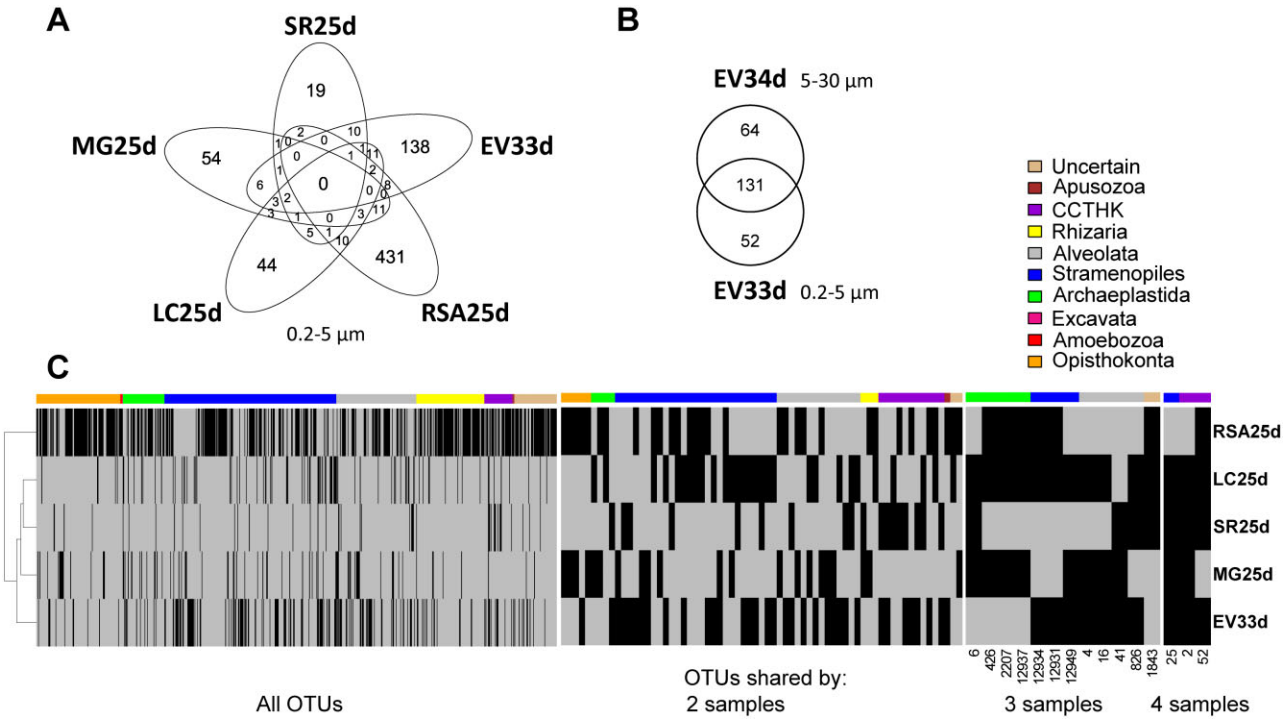


Fig. 5. Distribution of protist OTUs in the five shallow freshwater systems studied. Sequence data from replicates were pooled.
A. Five-set Venn diagram showing the number of specific and shared OTUs in the different freshwater systems (0.2–5 µm fraction size).
B. Venn diagram showing OTUs shared by the two fraction sizes analysed in Etang des Vallées.
C. Clustering analysis of the five ecosystems based on the presence of shared OTUs, as shown by the heatmap. Only 0.2–5 µm size fractions are considered. Heatmaps show all OTUs and OTUs shared by two, three or four ecosystems. No OTU is shared by all five ecosystems. Each row represents an ecosystem and each vertical bar an OTU. Black: OTU present, Grey: OTU absent. The coloured sidebar indicates the taxonomic affiliation of the OTU represented by the bar below. Colour codes representing different phylogenetic affiliation are indicated in the box. CCTHK: Cryptophyta, Centroheliozoa, Telonemia, Haptophyta and Katablepharida. For presentation reasons, the width of OTU bars is not the same in all the heatmaps. RSA25d, Ru Sainte Anne; LC25d, La Claye; SR25d, Saint Robert; MG25d, Mare Gabard; EV33d, EV34d, Etang des Vallées.

tems. Haptophytes were also identified, although in low proportions, and only in the Etang des Vallées and the Ru Sainte Anne. OTUs affiliated to Excavata, Labyrinthulida, Xanthophyceae, Apicomplexa, Centroheliozoa, Telonemida, Amoebozoa and Apusomonadida (Apusozoa) were detected only in the highly diverse Sainte Anne brook (Fig. 2, Table 1). Three additional apusozoan OTUs were detected in Sainte Anne brook and the Etang des Vallées.

Despite some similarity in the distribution of large phylogenetic groups, with cryptophytes, stramenopiles and alveolates dominating the different shallow water systems, protist communities were very different at the phylotype scale. Indeed, no OTU was shared by all the systems and the vast majority of the 768 OTUs detected in the small size fraction was specific to the libraries of a single ecosystem (Fig. 5A). Sixty-seven and twelve OTUs were shared by two and three different ecosystems respectively (Fig. 5C). Only three OTUs were shared by four systems (Fig. 5C), two of which affiliated to cryptophytes and corresponded to the most abundant OTUs (Fig. 4). Remarkably, OTU_52, affiliated to *Cryptomonas curvata*, was particularly abundant in St Robert and La Claye samples where it represented as much as 43–52% and 69–78% of reads respectively. However, OTU_52 dropped to 4–7% of reads in the 0.2–5 µm fraction of the Etang des Vallées and was not detected in Mare Gabard. OTU_25, affiliated to stramenopiles, was also shared by four systems, but represented only 0.13% of all reads in 0.2–5 µm fractions.

The composition of protist communities within the two different size fractions (0.2–5 and 5–30 µm) of the Etang des Vallées was very similar, as revealed by the NMDS and CoA analyses (Figs 3 and S2) as well as by low Bray–Curtis distances (0.35 on average, min: 0.29, max: 0.43). Moreover, the community structure was similar in both size fractions from the Etang des Vallées, which displayed the highest diversity and evenness recorded of all samples (Table 1). The taxonomic composition was also similar at high-rank taxa level, although diatoms and green algae were in slightly higher proportions in the largest size fraction (Fig. 2). At a finer level of resolution,

more OTUs were shared by both size fractions (131 OTUs) than specific to each of them (Fig. 4B).

Phylogenetic diversity

In order to get more detailed information on the different OTUs identified, we carried out phylogenetic analyses with partial 18S rRNA gene sequences representative of the different OTUs. Although cryptophytes were the most abundant group in our samples, they were not very diverse. Only 27 OTUs were affiliated to that group (3.3% of all OTUs). Most of these OTUs were closely related to existing sequences, affiliating to photosynthetic genera, especially *Cryptomonas* (containing the highly overrepresented OTU_52 corresponding to *C. curvata* mentioned above) and *Chroomonas*, but also to the phagocytic plastid-lacking *Goniomonas* (Fig. S2). However, several OTUs appeared to be quite divergent; for instance, representative sequences of OTUs 838 and 1812 shared no more than 93% of identity with their first BLAST hit in the NCBI database, the sequences HM135076 from freshwater (Luo *et al*., 2011) and AM901364 from the cultured *Cryptomonas commutata* strain M1975 respectively (Fig. S3).

A total of 268 OTUs (33% of all OTUs) were affiliated with stramenopiles, the second most abundant group in our samples. Nearly half of them were related to Chrysophyta-Synurophyta (125 OTUs) whereas the remaining ones affiliated to Eustigmatophyta, Dictyochophyta, Bicosoecida, Oomycetes, Bacillariophyta, Xanthophyta and Labyrinthulida (Fig. S4). Surprisingly, 14 OTUs seemed to belong to various MAST groups (Fig. 6), originally thought to be exclusively marine. Nine OTUs affiliated to the group MAST-12, being related to sequences previously detected in a wide variety of ecosystems, such as a suboxic Norwegian estuary (Kolodziej and Stoeck, 2007), freshwater lakes (Lefèvre *et al*., 2008; Monchy *et al*., 2011) and a peat bog (Lara *et al*., 2011). OTU_116 was closely related (99% identity) to MAST-2 sequences detected in a freshwater lake (Luo *et al*., 2011) and the Mediterranean (Diez *et al*., 2001). OTUs_1850 and 3339 affiliated to MAST-3, which so far was known to contain only sequences from marine environments. Finally, OTU_222 and OTU_3247 had similarity by BLAST to MAST-6 sequences retrieved from the Mediterranean (AF363207, Diez *et al*., 2001) and other MAST-6 sequences, with 94% and 91% identity respectively. However, from the phylogenetic analysis, their affiliation to this clade is unclear and will require a more in-depth exploration of these groups.

Alveolates were also found to be diverse (Figs S5 and S6), but the vast majority of OTUs were assigned to ciliates (88 out of 125 alveolate OTUs). Dinoflagellates were also represented in our samples, along with 17 OTUs related to sequences of putative freshwater (Amaral-Zettler *et al*., 2008; Monchy *et al*., 2011) or marine (Behnke *et al*., 2010; Scheckenbach *et al*., 2010) perkinsid parasites. Interestingly, several OTUs grouped with Apicomplexa and *Colpodellida* and clustered into two distinct groups (Fig. S5). Eight OTUs clustered with *Colpodella edax* (Leander *et al*., 2003) and many environmental sequences from freshwater lakes or ponds (Richards, 2005; Lefèvre *et al*., 2007; Nakai *et al*., 2012; Oikonomou *et al*., 2012). Three other OTUs clustered with sequences affiliated to apicomplexans (mostly *Cryptosporidium*) coming from more diverse environments, e.g. peat bog (Lara *et al*., 2011), marine sediment (Dawson and Pace, 2002), humans (Yuan *et al*., 2012) or ostrich faecal samples (R. Martinez-Diaz, unpublished).

From the 131 OTUs affiliated with opisthokonts, the great majority (125 OTUs) related to fungi, mostly to chytrids but also to ascomycetes, basidiomycetes or to the basal Rozellida/Cryptomycota lineage (Fig. S7). In addition, 11 OTUs were related to the aphelids, the sister group to fungi, which so far mostly contains highly divergent 18S rRNA gene environmental sequences (Karpov *et al*., 2014). Several OTUs clustered with ichtyosporeans and choanoflagellates (Fig. S7).

All rhizarian OTUs branched with cercozoan sequences. Even though cercozoan OTUs were not very abundant in our samples and represent only 1.2% of the total number of reads, they were composed of numerous phylotypes (106 OTUs) that represented altogether 13% of all OTUs. The vast majority of these OTUs affiliated to the Filosa (Fig. S8). Several OTUs branched with *Endomyxa*, although they shared only 93.6% identity on average with their first BLAST hits in GenBank (calculated on the 13 OTU representative sequences affiliated to *Endomyxa*, min: 89% between OTU_12985 and AB526843, max: 99% between OTU_1810 and EU910610). Furthermore, five OTUs clustered with sequences from the Novel Clade 10, and OTUs_2162 and 608 shared 92% and 91% identity respectively with sequence EU567287 from Novel Clade 11 – Tremulida, two recently defined deep-branching cercozoan lineages (Bass *et al*., 2009; Howe *et al*., 2011).

Sixty-five OTUs (8% of all OTUs) belonged to Archaeplastida (Fig. S9). Five of them strongly affiliated to *Embryophyta*. Given the pre-filtration steps, these OTUs must correspond to pollen grains, free DNA or cells from dead leaves. Strikingly, they represented altogether only 0.04% of all reads, in spite of the dense vegetation around most sampled systems. Four additional OTUs most likely corresponded to unicellular Streptophyta, affiliated to Clausteriaceae and Desmidiaceae. The remaining archaeplastid OTUs affiliated to the chlorophyte classes Mamiellophyceae, Trebouxiophyceae, Nephroselmidophyceae and, mainly, Chlorophyceae. In addition, we

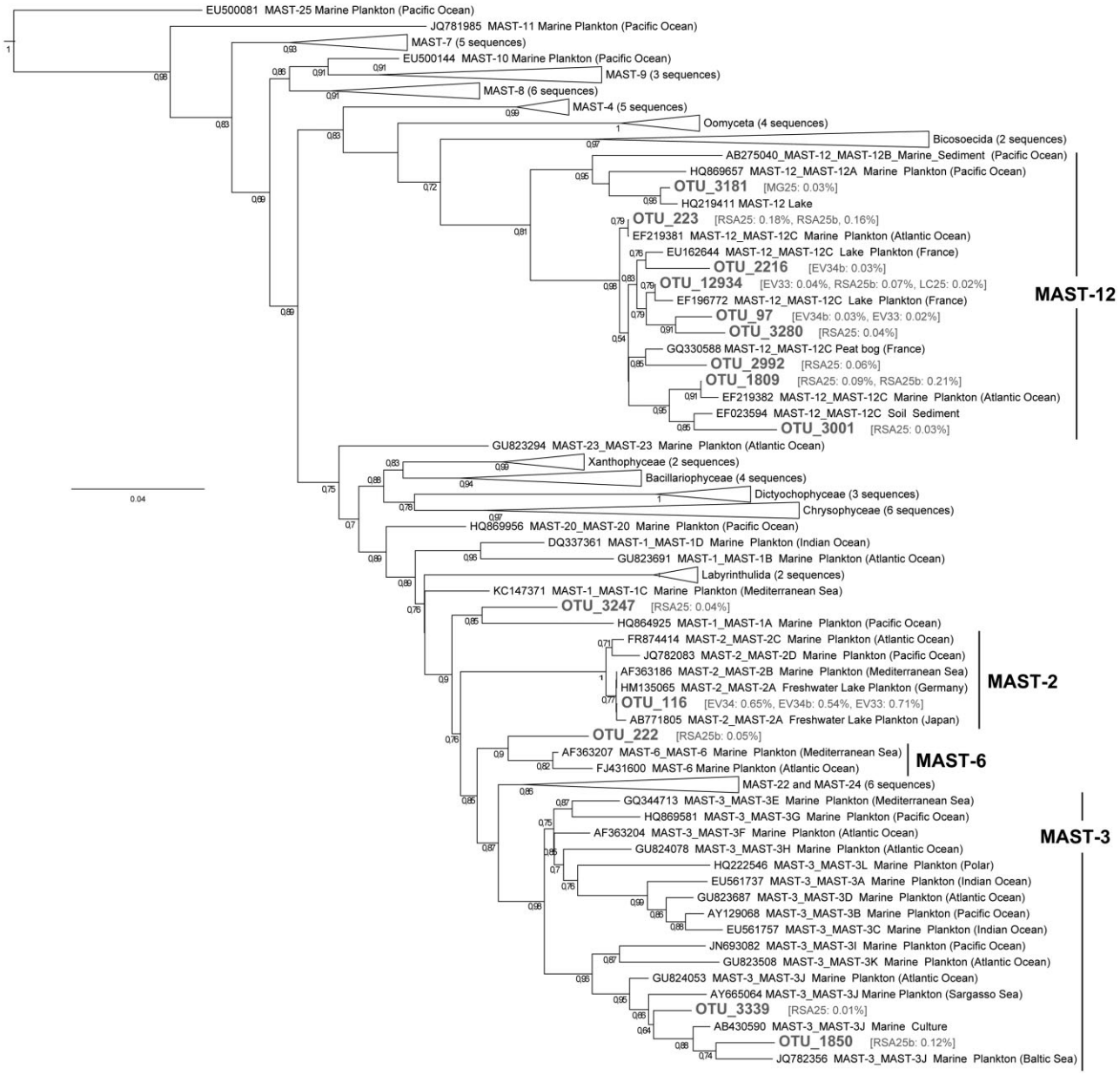


Fig. 6. Approximate maximum likelihood (ML) phylogenetic tree of partial 18S rRNA gene stramenopile sequences showing the presence of several MAST clades in shallow freshwater systems. Non-MAST stramenopile lineages have been collapsed. A total of 392 unambiguously aligned positions were used to reconstruct the tree. Two alveolate sequences were used to root the tree. Representative sequences of OTUs from this work are shown in bold. The name of samples where MAST OTUs were detected and the respective proportion of reads are shown within brackets. The scale bar represents the number of estimated substitutions per position for a unit branch length. RSA25/RSA25b, Ru Saint Anne; MG25/MG25b, Mare Gabard; EV33/ EV33b (0.2–5 μm) and EV34/EV34b (5–30 μm), Etang des Vallées.

detected a few OTUs belonging to other, less abundant eukaryotic groups. Thus, four, one and six OTUs affiliated to katablepharids, telonemids and haptophytes respectively (Fig. S3). Most haptophyte OTUs were very close to other freshwater sequences, but the OTU_3295 had a sequence nearly identical to *Jomonlithus littoralis* (AM490979; 99% identity), a coastal species and to two environmental sequences (JX680345, JX680344) from a brackish pond (Simon *et al.*, 2013). Sequences belonging to the basal haptophyte group HAP-1 (Šlapeta *et al.*, 2005) were detected in the Ru Sainte Anne (OTU_3063; Fig. S3). Six additional OTUs belonged to centroheliozoans (Fig. S3). Their closest relatives were sequences from soil or freshwater sediment and were only retrieved in Sainte Anne brook, the most narrow and less deep ecosystem (Table 1). That observation, along

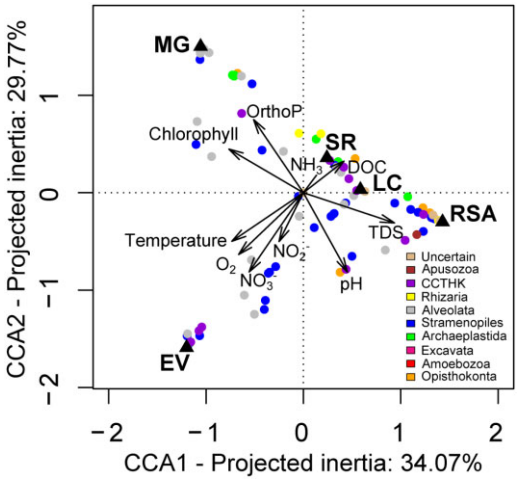


Fig. 7. Canonical correspondence analysis (CCA) plot. Only 0.2–5 µm size fractions are considered. Each dot represents an OTU. The colours indicate the taxonomic affiliation. Black triangles indicate samples. Duplicate samples appear superimposed. CCTHK: Cryptophyta, Centroheliozoa, Telonemia, Haptophyta and Katablepharida. TDS: total dissolved solutes, DOC: dissolved organic carbon, OrthoP: orthophosphate.

with the higher proportions and diversity in Sainte Anne brook of cercozoa, fungi and diatoms, known to be usually composed of large and/or benthic cells, could be explained by a higher influence or contribution of benthic communities in this system. One excavate OTU (OTU_1857; 99% identical to *Trimastix marina*) and one apusomonad OTU were also detected in Sainte Anne brook. Three amoebozoan OTUs from the brook and two ancyromonad OTUs could also be detected in our small freshwater systems (Fig. S9). Finally, 66 of the total 812 OTUs determined could not be affiliated with confidence to any group.

Effects of physico-chemical parameters and distance on community structure and composition

The five sampled ecosystems were separated by distances between 2 and 9.5 km (Fig. S1). To see whether the differences in community composition increased with geographical distance, we pooled sequence data from replicate samples for the 0.2–5 µm size fraction, calculated Bray–Curtis distances between all systems based on the pooled data and performed a Mantel analysis to test the correlation between Bray–Curtis and geographical distance matrices. The Mantel test showed no correlation between differences in protist community composition and distance between ecosystems ($r = 0.2151$, P-value = 0.236). As an illustration, OTU_52 (*Cryptomonas curvata*) reached 44% of all reads (43–52% in replicates) in the Saint Robert pond, but it could not be detected in the Gabard pond (Fig. 5), its nearest system (2.1 km away). However, this OTU was detected in all other systems, up to 9.5 km away.

As mentioned above, each of our five ecosystems was characterized by a set of specific environmental variables, with the Mare Gabard and the Ru Saint Anne being, respectively, the less and most charged in TDS and Saint Robert being highly enriched in ammonia (Table 1; Fig. 1). All these differences in environmental parameters correlated to the structure of protist communities in these ecosystems as revealed by a Mantel test linking environmental and community pairwise distances between samples ($r = 0.7323$, P-value = 0.014). However, additional canonical correspondence analyses (CCA) did not reveal any clear relationship between any of the detected eukaryotic supergroups and one or several environmental variables, although some OTUs appeared to be associated with particular samples according to their specific presence (Fig. 5) and distribution in CCA (Fig. 7). At the phylum or class level (Fig. S10), haptophytes, katablepharids, choanoflagellates, cercozoans and cryptophytes appeared to reach higher proportions where pH was the highest and phosphate concentration the lowest, the two latter variables being tightly correlated in PCA and CCA (Figs 1 and 7). Streptophytes were associated with high values of DOC whereas bacillariophytes and fungi were more abundant where total dissolved solutes were the highest. These features could thus partially explain the differences in composition of the small eukaryotes among our contrasting ecosystems.

Discussion

Shallow freshwater systems, reservoirs of protist diversity

Compared with oceans, molecular protist diversity surveys in freshwater ecosystems are still scarce. Paradoxically, inland water bodies are collectively much more heterogeneous than oceans and much more sensitive to environmental variation. Despite so, only a handful of studies on freshwater protist communities exist, mainly from a variety of lakes, from temperate regions and mountains to polar areas (e.g. Charvet *et al.*, 2012; Lepère *et al.*, 2013; Taib *et al.*, 2013). However, nearly nothing is known from smaller, shallower and often ephemeral freshwater systems, despite their global distribution and ecological importance (Downing *et al.*, 2006). Recent approaches based on massive sequence analysis, although prone to a variety of methodological errors and biases, are called to facilitate comparative studies among those ecologically relevant ecosystems. This is especially true for tiny protists, because morphology-based exploration very often miss their phylogenetic diversity (Massana *et al.*, 2002; Moreira and López-García, 2002; Šlapeta

et al., 2006). To contribute to that task, we carried out a study based on massive pyrosequencing of amplified 18S rRNA gene fragments of protists in the 0.2–5 μm size range in five shallow freshwater ecosystems from a temperate area. Many of the high-rank taxa detected occur, although with variable relative abundances, in other freshwater systems, such as lakes. Cryptophytes, stramenopiles and alveolates were by far the most abundant supergroups detected, although their internal diversity varied greatly among ecosystems (Fig. 2). Although less abundant in the open sea (Shi *et al*., 2009; Kirkham *et al*., 2013), small cryptophytes seem to be common in freshwater (Lefranc and Thénot, 2005; Šlapeta *et al*., 2005; Lepère *et al*., 2008; Mangot *et al*., 2012; Taib *et al*., 2013). Stramenopiles, as well as ciliates, which dominate the alveolate diversity in our samples, are also common in protist communities from freshwater lakes (Lefranc and Thénot, 2005; Lepère *et al*., 2008; Taib *et al*., 2013). Fungi, Ichthyosporea and choanoflagellates within the Opisthokonta, cercozoans within the Rhizaria, and members of the Amoebozoa, Archaeplastida, Excavata as well as members of the phylogenetically unresolved Haptophyta, Centrohelida, Katablepharida, Telonemida- and Apusozoa were also detected.

Surprisingly, fungi and their relatives, which are often detected in very large proportions in some lakes (Lefranc and Thénot, 2005; Lepère *et al*., 2008), accounting for up to 94% of all the 454-pyrosequences in some of them (Taib *et al*., 2013), were in very low proportions (generally much less than 10%; Fig. 2) in our systems. This may reveal a real difference between shallow ecosystems and deeper lakes, but it could also reflect varying population dynamics along the year (Nolte *et al*., 2010). Our samples were collected in early spring and many of the fungal OTUs detected correspond to potential parasitic fungi or related lineages, such as chytrids, rozellids/cryptomycota or aphelids (Fig. S7), which might experience fluctuation depending on host population dynamics along the year. Temporal surveys would be needed to test such a hypothesis. Several OTUs in our freshwater systems branched deeply within known eukaryotic supergroups. Thus, the Rozellida-Cryptomycota and the aphelids have been considered the deepest lineages of fungi (Lara *et al*., 2010; Jones *et al*., 2011), although recently, they have been re-classified as forming part of the superphylum Opisthosporidia together with Microsporidia and would be the sister group to fungi (Karpov *et al*., 2014). Other detected deep-branching lineages include the Colpodellida (Leander *et al*., 2003) within the alveolates, the poorly known novel Clade-10 Cercozoa and Tremulida (Bass *et al*., 2009; Howe *et al*., 2011) or the early diverging haptophyte lineage HAP-1, so far only detected in freshwater systems (Šlapeta *et al*., 2005; Shalchian-Tabrizi *et al*., 2011).

Although protist community composition was different in the five analysed systems, the comparison of protist diversity in two cell fraction sizes (0.2–5 and 5–30 μm) in the most diverse site, the Etang des Vallées, did not differ much in terms of high-rank taxa (Figs 2 and 3). Indeed, despite the presence of OTUs apparently confined to each cell-size fraction (21% and 25%, for the smaller and larger size ranges, respectively), a considerable proportion of OTUs (53%) were shared between the two cell fractions (Fig. 5B). Several explanations are possible and non-mutually exclusive. One is that many small organisms may be retained in filters of larger pore-diameter if the filtered biomass is high, because they might be entangled with larger organisms. On the opposite, flexible organisms of relatively bigger sizes than those of filter pores may pass through them under the filtration pressure applied. Finally, many organisms may encompass a size range that spans the 5 μm diameter barrier imposed by our filters. This may be due to either differential sizes during their life cycle (e.g. gametes or spores and vegetative forms) or to size variability under a given life stage.

The broad protist diversity unveiled in the five freshwater systems also reflects a wide ecological diversity of functions. At a very general level, typical photosynthetic, heterotrophic and parasitic groups were detected. Among photosynthesizers, cryptophytes followed by photosynthetic stramenopiles (diatoms, chrysophytes, synurophytes, xantophytes) largely exceeded green algae. Many of these lineages may also contribute to heterotrophic activities, because many photosynthetic protists are mixotrophs (Zubkov and Tarran, 2008; Massana, 2011; Hartmann *et al*., 2012). Typical predators span most of the eukaryotic diversity identified, from choanoflagellates and amoeba to bicosoecids, ciliates, cercozoa, centrohelids, katablepharids or apusozoa (Fig. 2). However, heterotrophic activities also encompass the degrading activity of osmotrophic taxa, including several fungi and labyrinthulids and that of parasitic or parasitoid protists (many chytrids, cryptomycota, aphelids, oomycetes, apicomplexa, perkinsids).

Altogether, our extensive 18S rDNA-based survey of small protists suggests that shallow freshwater systems are important reservoirs of eukaryotic diversity. Not only members of all supergroups are present, but also several members of uncertain, poorly known or deep-branching lineages. Given the heterogeneity of this kind of systems on Earth, it can be advanced that further studies on shallow freshwater systems will uncover yet-to-characterize protists, the study of which will be of relevance to understand the ecology of these ecosystems and, from an evolutionary perspective, to reconstruct poorly resolved areas in the eukaryotic tree.

Marine-freshwater barriers transgressed

One of the potential surprises that the study of protist diversity in varied freshwater systems may bring is the increasing awareness that salinity barriers can be overcome more easily than previously thought. Although lineages with members in both freshwater and marine systems exist, such as dinoflagellates, haptophytes, perkinsids or stramenopiles (Logares *et al*., 2007; Bråte *et al*., 2010; Shalchian-Tabrizi *et al*., 2011; Simon *et al*., 2013), the marine-freshwater transition is thought to be rare (Logares *et al*., 2009). However, this view may be biased by current undersampling of freshwater systems. Indeed, in our study, we have detected protist lineages thought to be exclusively marine in the past, notably several MAST lineages. Thus, we did not only identify members of MAST-2 and MAST-12 lineages, which have recently been detected in other freshwater systems (Massana *et al*., 2014), but also members of MAST lineages never identified in freshwater systems before. This was the case of MAST-3, for which we identified two bona fide OTUs and, potentially, of MAST-6 and even MAST-1 lineages, although OTUs related to the latter are of much more uncertain affiliation (Fig. 6).

Moreover, in addition to the occurrence of the emblematic MAST groups in these shallow freshwater systems, we also identified many OTUs widespread in the eukaryotic tree that share 99% identity or more with sequences retrieved from marine environments. Examples are the haptophyte OTU_3295, practically identical to *Jomonlithus littoralis* (Fig. S3), the stramenopile OTU_116 (Fig. S4), the ciliate OTU_255 (Fig. S6) or the excavate OTU_1857 affiliating to *Trimastix marina* (Fig. S9). But there are many other OTUs whose closest relatives are sequences from marine systems, even if similarity is slightly lower.

In fact, even if salinity seems to be a relevant ecological determinant structuring microbial communities (Lozupone and Knight, 2007), the marine-freshwater barrier seems to be relatively easy to cross. On the one hand, several protists, notably cryptophytes, are osmotolerant and can cope with various salinity concentrations by the means of contractile vacuole regulation (Hoef-Emden, 2014). On the other hand, many freshwater systems contain relatively high levels of dissolved solutes (organic and/or inorganic) requiring similar adaptations as those require for life in seawater salts. In this sense, it is interesting to note that most of the 'typically marine' lineages detected in our freshwater systems were identified in the Ru Sainte Anne, which has the highest TDS content and seems greatly influenced by this parameter (Table 1; Fig. 7).

Elements of protist biogeography

There are two essentially opposed views with regard to microbial and, more specifically, protist biogeography. Most classical views posit that small free-living protists would find little barriers to dispersal and, hence, be widely distributed. Differences in protist community structure would then be essentially explained by local environmental parameters. This 'everything is everywhere, but the environment selects' view seems supported by the occurrence of cosmopolitan protist species (Baas-Becking, 1934; Finlay, 2002; Šlapeta *et al*., 2006). On the opposite extreme, a variety of studies seem to suggest that endemic protists (Foissner, 2006) and taxa–area relationships exist for microbial eukaryotes (Green *et al*., 2004). Comparing community compositions sidesteps undersampling and the nearly impossible task of demonstrating true microbial endemisms (Martiny *et al*., 2006). A recent study suggested that the beta-diversity of the small eukaryotes between lakes was linked to the geographic distances between ecosystems (Lepère *et al*., 2013). However, it is extremely difficult to disentangle the effect of local environmental parameters from physical distance. Furthermore, other factors, such as the temporal (Nolte *et al*., 2010) and the phylogenetic scale (Ragon *et al*., 2012) need to be taken into account. In our case, we sampled the five systems at the same or correlative dates to limit temporal effects and, although the distances involved were different from those in the study of Lepère and colleagues (2013) (10 versus 100 km scale), our study clearly rejects geographic distance as a driver of community composition. On the contrary, Mantel tests show significant correlations between differences in community structure and physico-chemical parameters. However, clear associations between high taxon levels and environmental parameters were not identified (Fig. 7 and S10). Differences at finer, OTU scale might provide hints about the role of environmental selection in determining community structure. At any rate, testing which factors more greatly influence the composition of these communities would require the inclusion of biotic parameters, notably the diversity and relative abundance of bacteria, archaea and viruses from the same systems.

Experimental procedures

Sampling and measurement of physico-chemical parameters

Samples were collected in spring 2012 from five small and shallow freshwater ecosystems at the Natural Regional Park of the Chevreuse Valley (France, South of Paris) (Table 1 and Fig. S1). The systems were chosen to represent a variety of conditions at local scale, from forest ponds rich in organic matter (Mare Gabard) to more urban and agricultural-influenced systems (Saint Robert). Surface water was collected using sterile plastic carboys and processed immediately back in the laboratory. Water samples were serially filtered through 30 μm pore-size nylon filters (Millipore), and through 5 and 0.2 μm pore-size diameter

Nucleopore membranes (Whatman). Filters were stored frozen at –20°C until DNA extraction. The water temperature, pH, the concentration of total dissolved solutes and the level of dissolved oxygen were measured *in situ* using a multiparameter probe (Multi 350i, WTW). The concentrations of dissolved nitrate (NO_3^-), nitrite (NO_2^-), ammonia (NH_3), orthophosphate (PO_4^{3-}) and DOC were measured in water samples filtered through 0.2 μm pore-size diameter Nucleopore membranes on the same day of sampling using manufactured colorimetric tests (Hach-Lange). Chlorophyll *a* concentrations were determined after harvesting plankton biomass on glass microfiber filters (GF/F, Whatman) that were stored at –20°C until ethanol pigment extraction. For chlorophyll extraction, filters were dried then ground in 7 ml of absolute ethanol, and heated at 70°C for 20 min. After centrifugation, 1 ml of supernatant was collected and optical densities at 665 and 750 nm were measured (spectrophotometer DR5000 Hach-Lange). Chlorophyll *a* concentration was determined on pigment extract by spectrophotometry, as follows: $[Chla] = 11.9 * (OD_{665} - OD_{750}) / w * Ve / Vf$ with [Chla]: $[Chla] = 11.9 * (OD_{665} - OD_{750}) / w * Ve / Vf$ chlorophyll *a* concentration (μg l^{-1}), 11.95: Reciprocal specific absorbance coefficient of chlorophyll *a* at 665 nm (μg cm ml^{-1}), OD_{665}: optical density at 665 nm, OD_{750}: optical density at 750 nm, w: width of the spectroscopic cuvette (1 cm), Ve: volume of the pigment extract (in ethanol, ml) and Vf: volume of water filtrated on the glass microfiber filter (l). The protocol was adapted from Ritchie (2006).

DNA extraction, amplification and sequencing of 18S rRNA genes

DNA was extracted from cells collected onto filters that were cut into pieces using the PowerSoil DNA extraction kit (MoBio) according to the manufacturer's instructions. DNA was eluted in 80 μl of 10 mM Tris, pH 8.0. 18S rRNA gene fragments of approximately 550 bp, encompassing the V4 hypervariable region, were amplified using the newly designed primer EK-565F (5′-GCAGTTAAAAAGCTCG TAGT) and primer 18s-EUK- 1134-R- UNonMet (5′-TTTAA GTTTCAGCCTTGCG) biased against Metazoa (Bower *et al.*, 2004). Both forward and reverse primers were tagged with 20 different 10 bp molecular identifiers (MIDs) to allow pooling and later differentiation of polymerase chain reaction (PCR) amplification products from 20 distinct samples. PCR amplifications were conducted in a total reaction volume of 25 μl using 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleotide (dNTP) mix (PCR Nucleotide Mix, Promega), 0.3 μM of each primer, 0.3–2 μl of DNA sample and 0.5 U HotStart Taq polymerase (Taq Platinum, Invitrogen). The amplification conditions consisted of 25 cycles (94°C for 30 s, 58°C for 45 s and 72°C for 90 s), preceded by 3 min of denaturation at 94°C and ending with a 10 min final extension step at 72°C. Amplicons from 5–7 independent PCR products for each sample were pooled together and then purified using the QIAquick PCR purification kit (Qiagen), according to the manufacturer's instructions. The same amounts (around 200 ng) of purified amplicons from 20 samples were pooled. Amplicons were pyrosequenced using the 454 GS FLX Titanium technology from Roche (Beckman Coulter Genomics). Sequences have been deposited at NCBI under the BioProject number PRJNA259710.

454 Pyrosequence analysis

We obtained a total of 265 899 pyrosequences (Table 1). A series of filters were applied in order to retain only high-quality sequences. First, pyrosequences containing errors in the primer region and positions with undetermined bases were eliminated using a local pipeline (Bachy *et al.*, 2013). The remaining sequences were analysed with AmpliconNoise (Quince *et al.*, 2011) to further eliminate errors introduced during PCR reactions or 454 sequencing, and build OTUs. Filtered reads were clustered by pairwise alignment and average linkage into OTUs with a 98% similarity cut-off using AMPLICONNOISE integrated in our local pipeline (Bachy *et al.*, 2013). Singletons, i.e. OTUs composed of only one read, were eliminated for precaution. The most abundant sequence in each OTU was used as reference. OTU reference sequences were blasted against the Silva SSU111 database (Pruesse *et al.*, 2007) and assigned to taxonomic groups based on sequence similarity. The sequences in all OTUs were then attributed to the different samples according to their MIDs. Chimerical OTUs were eliminated by a stringent procedure combining automated and manual steps. OTUs including sequences from at least two different samples were considered to be real. OTUs composed of sequences from only one sample were checked for chimeras using KEYDNATOOL (http://KeyDNAtools.com). The sequences considered suspect by the software were double checked by comparing BLAST hits recovered from independent sequence fragments (sequences were split in two and three fragments). Finally, OTUs whose representative sequence had a coverage of less than 90% with its first BLAST hit were eliminated if they were present in only one sample; they were kept but with their taxonomic affiliation changed to 'uncertain' if they were present in at least two samples. OTUs affiliated to cryptophyte nucleomorphs were excluded from our analysis. After filtering, we kept 146 549 correct reads (Table 1).

Phylogenetic analyses

Phylogenetic trees were built for each eukaryotic super-group or for several high-rank taxonomic lineages if they comprised only a few OTUs. Analyses included representative sequences of OTUs, their first BLAST hit and sequences from the closest cultured members. Sequences were aligned using PROBCONS (Do *et al.*, 2005). Positions retained to build trees were selected from the multiple alignments using GBLOCKS (Castresana, 2000) with the less stringent parameters. Phylogenetic reconstructions were then carried out by maximum likelihood approximation using FASTTREE (Price *et al.*, 2010). Trees were visualized using FIGTREE (http://tree.bio.ed.ac.uk/software/figtree/). Sequences with particularly interesting positions in trees were then blasted against the NCBI database (http://blast.ncbi.nlm.nih.gov) to have an insight of their similarity with additional sequences in databases. The taxonomic indications given in trees are based on the taxonomic affiliation proposed in the PR2 database (http://ssu-rrna.org/index.html).

Statistical analyses

Statistical analyses were conducted using the R SOFTWARE (http://cran.r-project.org) (R Development Core Team, 2013). To assess overall differences between microbial community compositions, we calculated pair-wise Bray–Curtis distances between all samples on the basis of OTU proportions (number of reads from each OTU for each sample normalized to the total number of reads in the corresponding sample) among replicates or sampling sites (Bray and Curtis, 1957). This would avoid heterogeneity due to different numbers of sequences generated per sample. NMDS ordination analyses were conducted based on Bray–Curtis distances (after applying a Wisconsin standardization to balance the influence of the most and least abundant OTUs) using the 'Vegan' R PACKAGE (Oksanen *et al*., 2013). CoA on all OTU frequencies for all 12 samples were done using the 'Ade4' R PACKAGE (Dray and Dufour, 2007). Diversity and richness indices were determined using the 'Vegan' package. Richness was estimated by rarefaction analysis as the estimated number of OTUs in a random subsample of each sequence library (raw counts of OTUs), of the same size as the smallest one (Hurlbert, 1971). Simpson index is defined as $D' = 1 - \sum_{i=1}^{S}(f_i^2)$ (Simpson, 1949) and evenness was calculated as $e = {-\sum_{i=1}^{S} f_i \ln(f_i)} \Big/ {\ln(S)}$ (Pielou, 1966) with S being the observed number of OTUs and f_i the frequency of each OTU_i in the sample. Venn diagrams showing the number of OTUs shared by, or exclusive to, the different samples, and heatmaps showing the presence/absence of OTUs were built using the 'GPLOTS' PACKAGE (Bolker *et al*., 2012). To test whether community composition correlated with environmental parameters in the different samples, we constructed a matrix of Bray–Curtis dissimilarities based on OTU frequencies (sequences from replicate samples were pooled given that they clustered together in previous analyses) and a matrix of Euclidean distances based on physico-chemical parameters for all ecosystems using the 'Vegan' package. Both matrices were compared using a Mantel test. The Bray–Curtis matrix was also compared with a matrix of geographical distances between ecosystems. Geographical distances were estimated based on coordinates (http://biodiversityinformatics.amnh.org/open_source/gdmg/), using the mean earth radius (Moritz, 2000) as spheroid. PCA and CCA were conducted on centred and scaled physico-chemical parameters and OTU frequencies in replicate samples (0.2–5 μm size fractions) using the 'ADE4' PACKAGE.

Acknowledgements

We are thankful to F. Hardy, the Parc Naturel Régional de la Haute Vallée de Chevreuse and the Office National des Forêts du Parc de Rambouillet. We also thank Giselle Walker for critical reading of the manuscript. The research leading to these results received funding from the CNRS EC2CO program and the European Research Council under the European Union's Seventh Framework Program ERC Grant Agreement 322669 'ProtistWorld'.

References

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., *et al*. (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59:** 429–493.

Amaral-Zettler, L.A., Rocca, J.D., Lamontagne, M.G., Dennett, M.R., and Gast, R.J. (2008) Changes in microbial community structure in the wake of Hurricanes Katrina and Rita. *Environ Sci Technol* **42:** 9072–9078.

Baas-Becking, L. (1934) *Geobiologie of inleiding tot de milieukunde*. Hague, the Netherlands: W.P. van Stock & Zoon.

Bachy, C., Dolan, J.R., López-García, P., Deschamps, P., and Moreira, D. (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *ISME J* **7:** 244–255.

Bass, D., Chao, E.E.-Y., Nikolaev, S., Yabuki, A., Ishida, K., Berney, C., *et al*. (2009) Phylogeny of novel naked Filose and Reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea revised. *Protist* **160:** 75–109.

Behnke, A., Barger, K.J., Bunge, J., and Stoeck, T. (2010) Spatio-temporal variations in protistan communities along an O/HS gradient in the anoxic Framvaren Fjord (Norway). *FEMS Microbiol Ecol* **72:** 89–102.

Bolker, B., Lodewijk, B., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., *et al*. (2012) *gplots: Various R programming tools for plotting data., v2.12.1* [WWW document]. URL http://cran.r-project.org/web/packages/gplots/index.html.

Bower, S.M., Carnegie, R.B., Goh, B., Jones, S.R.M., Lowe, G.J., and Mak, M.W. (2004) Preferential PCR amplification of parasitic protistan Small subunit rDNA from metazoan tissues. *J Eukaryot Microbiol* **51:** 325–332.

Bray, J.R., and Curtis, J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **27:** 325–349.

Bråte, J., Logares, R., Berney, C., Ree, D.K., Klaveness, D., Jakobsen, K.S., and Shalchian-Tabrizi, K. (2010) Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J* **4:** 1144–1153.

Caron, D.A. (1994) Inorganic nutrients, bacteria, and the microbial loop. *Microb Ecol* **28:** 295–298.

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

Charvet, S., Vincent, W.F., Comeau, A., and Lovejoy, C. (2012) Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. *Front Microbiol* **3:** 422.

Cheung, M.K., Au, C.H., Chu, K.H., Kwan, H.S., and Wong, C.K. (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *ISME J* **4:** 1053–1059.

Corpe, W.A., and Jensen, T.E. (1992) An electron microscopic study of picoplanktonic organisms from a Small Lake. *Microb Ecol* **24:** 181–197.

Costanza, R. (1999) The ecological, economic, and social importance of the oceans. *Ecol Econ* **31:** 199–213.

Dawson, S.C., and Pace, N.R. (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc Natl Acad Sci USA* **99:** 8324–8329.

Diez, B., Pedrós-alió, C., and Massana, R. (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67:** 2932–2941.

Do, C.B., Mahabhashyam, M.S.P., Brudno, M., and Batzoglou, S. (2005) ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res* **15:** 330–340.

Downing, J.A., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., *et al*. (2006) The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol Oceanogr* **51:** 2388–2397.

Dray, S., and Dufour, A.-B. (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* **22:** 1–20.

Fenchel, T. (1982) Ecology of heterotrophic microflagellates. I. Some important forms and their functional morphology. *Mar Ecol Prog Ser* **8:** 211–223.

Finlay, B. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296:** 1061–1063.

Foissner, W. (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozool* **45:** 339–366.

Gómez, F., Moreira, D., Benzerara, K., and López-García, P. (2011) *Solenicola setigera* is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. *Environ Microbiol* **13:** 193–202.

Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., *et al*. (2004) Spatial scaling of microbial eukaryote diversity. *Nature* **432:** 747–750.

Groisillier, A., Massana, R., Valentin, K., Vaulot, D., and Guillou, L. (2006) Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquat Microb Ecol* **42:** 277–291.

Guillou, L., Chrétiennot-Dinet, M.-J., Medlin, L.K., Claustre, H., Goër, S.L., and Vaulot, D. (1999) Bolidomonas: a new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *J Phycol* **35:** 368–381.

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R.M., Kirkham, A.R., 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., and Horiguchi, T. (2007) Species of the parasitic genus duboscquella are members of the enigmatic Marine Alveolate Group I. *Protist* **158:** 337–347.

Hartmann, M., Grob, C., Tarran, G.A., Martin, A.P., Burkill, P.H., Scanlan, D.J., and Zubkov, M., V (2012) Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc Natl Acad Sci USA* **109:** 5756–5760.

Hoef-Emden, K. (2014) Osmotolerance in the Cryptophyceae: jacks-of-all-trades in the Chroomonas Clade. *Protist* **165:** 123–143.

Howe, A.T., Bass, D., Scoble, J.M., Lewis, R., Vickerman, K., Arndt, H., and Cavalier-Smith, T. (2011) Novel cultured protists identify deep-branching environmental DNA clades of cercozoa: New genera *Tremula*, *Micrometopion*, *Minimassisteria*, *Nudifila*, *Peregrinia*. *Protist* **162:** 332–372.

Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* **52:** 577–586.

Jardillier, L., Zubkov, M.V., Pearman, J., and Scanlan, D.J. (2010) Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* **4:** 1180–1192.

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

Jones, M.D.M., Forn, I., Gadelha, C., Egan, M.J., Bass, D., Massana, R., and Richards, T.A. (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474:** 200–203.

Karpov, S.A., Mamkaeva, M.A., Aleoshin, V.V., Nassonova, E., Lilje, O., and Gleason, F.H. (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front Microbiol* **5:** 112.

Kirkham, A.R., Lepère, C., Jardillier, L.E., Not, F., Bouman, H., Mead, A., and Scanlan, D.J. (2013) A global perspective on marine photosynthetic picoeukaryote community structure. *ISME J* **7:** 922–936.

Knight-Jones, E.W., and Walne, P.R. (1951) Chromulina pusilla Butcher, a dominant member of the ultraplankton. *Nature* **167:** 445–446.

Kolodziej, K., and 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.

Lara, E., Moreira, D., and López-García, P. (2010) The environmental clade LKM11 and Rozella form the deepest branching clade of fungi. *Protist* **161:** 116–121.

Lara, E., Mitchell, E.A.D., Moreira, D., and López-García, P. (2011) Highly diverse and seasonally dynamic protist community in a pristine peat bog. *Protist* **162:** 14–32.

Leander, B.S., Kuvardina, O.N., Aleshin, V.V., Mylkinov, A.P., and Keeling, P.J. (2003) Molecular phylogeny and surface morphology of *Colpodella edax* (Alveolata): insights into the phagotrophic ancestry of apicomplexans. *J Eukaryot Microbiol* **50:** 334–340.

Lefèvre, E., Bardot, C., Noël, C., Carrias, J.-F., Viscogliosi, E., Amblard, C., and Sime-Ngando, T. (2007) Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ Microbiol* **9:** 61–71.

Lefèvre, E., Roussel, B., Amblard, C., and Sime-Ngando, T. (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS ONE* **3:** 10.

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

Lepère, C., and Boucher, D. (2006) Succession and regulation factors of small eukaryote community composition in a

lacustrine ecosystem (Lake Pavin). *Appl Environ Microbiol* **72:** 2971–2981.

Lepère, C., Domaizon, I., and Debroas, D. (2008) Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Appl Environ* **74:** 2940–2949.

Lepère, C., Domaizon, I., Taib, N., Mangot, J.-F., Bronner, G., Boucher, D., and Debroas, D. (2013) Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS Microbiol Ecol* **85:** 85–94.

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

López-García, P., and Moreira, D. (2008) Tracking microbial biodiversity through molecular and genomic ecology. *Res Microbiol* **159:** 67–73.

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

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

Logares, R., Brate, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K., and Rengefors, K. (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* **17:** 414–422.

Lozupone, C.A., and Knight, R. (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* **104:** 11436–11440.

Luo, W., Bock, C., Li, H., Padisák, J., and Krienitz, L. (2011) Molecular and microscopic diversity of planktonic eukaryotes in the oligotrophic Lake Stechlin (Germany). *Hydrobiologia* **661:** 133–143.

Mangot, J.-F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G., and Debroas, D. (2012) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environ Microbiol* **15:** 1745–1758.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

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

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

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

Massana, R., Castresana, J., Balagué, V., Romari, K., Groisillier, A., Valentin, K., *et al*. (2004b) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol* **70:** 3528–3534.

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

Massana, R., Del Campo, J., Sieracki, M.E., Audic, S., and Logares, R. (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J* **8:** 854–866.

Monchy, S., Sanciu, G., Jobard, M., Rasconi, S., Gerphagnon, M., Chabé, M., *et al*. (2011) Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environ Microbiol* **13:** 1433–1453.

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

Moreira, D., and López-García, P. (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol* **10:** 31–38.

Moreira, D., and López-García, P. (2014) The rise and fall of Picobiliphytes: how assumed autotrophs turned out to be heterotrophs. *Bioessays* **36:** 468–474.

Moritz, H. (2000) Geodetic reference system 1980. *J Geod* **74:** 128–133.

Nakai, R., Abe, T., Baba, T., Imura, S., Kagoshima, H., Kanda, H., *et al*. (2012) Eukaryotic phylotypes in aquatic moss pillars inhabiting a freshwater lake in East Antarctica, based on 18S rRNA gene analysis. *Polar Biol* **35:** 1495–1504.

Nolte, V., Pandey, R., and Jost, S. (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol Ecol* **19:** 2908–2915.

Oikonomou, A., Katsiapi, M., Karayanni, H., Moustaka-Gouni, M., and Kormas, K.A. (2012) Plankton microorganisms coinciding with two consecutive mass fish kills in a newly reconstructed lake. *ScientificWorldJournal* **2012:** 504135.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, R.P., O'Hara, R.B., *et al*. (2013) vegan: Community Ecology Package.

Patterson, D.J., and Larsen, J. (1991) *Biology of Free-Living Heterotrophic Flagellates. Systematics Association Special*, Vol. 45. Oxford, England: Clarendon Press.

Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. *J Theor Biol* **13:** 131–144.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5:** e9490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12:** 38.

R Development Core Team (2013) R: A language and environment for statistical computing.

Ragon, M., Fontaine, M.C., Moreira, D., and López-García, P. (2012) Different biogeographic patterns of prokaryotes and

microbial eukaryotes in epilithic biofilms. *Mol Ecol* **21:** 3852–3868.
Richards, T.A. (2005) The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. *Environ Microbiol* **7:** 1413–1425.
Ritchie, R.J. (2006) Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth Res* **89:** 27–41.
Romari, K., and Vaulot, D. (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49:** 784–798.
Scheckenbach, F., Hausmann, K., Wylezich, C., Weitere, M., and Arndt, H. (2010) Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proc Natl Acad Sci USA* **107:** 115–120.
Seenivasan, R., Sausen, N., Medlin, L.K., and Melkonian, M. (2013) *Picomonas judraskeda* gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as 'picobiliphytes. *PLoS ONE* **8:** e59565.
Shalchian-Tabrizi, K., Reier-Røberg, K., Ree, D.K., Klaveness, D., and Bråte, J. (2011) Marine–freshwater colonizations of haptophytes inferred from phylogeny of environmental 18S rDNA sequences. *J Eukaryot Microbiol* **58:** 315–318.
Shi, X., Marie, D., Jardillier, L., Scanlan, D., and Vaulot, D. (2009) Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS ONE* **4:** e7657.
Simó, R. (2001) Production of atmospheric sulfur by oceanic plankton: biogeochemical, ecological and evolutionary links. *Trends Ecol Evol* **16:** 287–294.
Simon, M., López-García, P., Moreira, D., and Jardillier, L. (2013) New haptophyte lineages and multiple independent colonizations of freshwater ecosystems. *Environ Microbiol Rep* **5:** 322–332.
Simpson, E.H. (1949) Measurement of diversity. *Nature* **163:** 688.
Stockner, J.G. (1988) Phototrophic picoplankton: an overview from marine and freshwater ecosystems. *Limnol Oceanogr* **33:** 765–775.
Stoeck, T., Taylor, G.T., and Epstein, S.S. (2003) Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). *Appl Environ Microbiol* **69:** 5656–5663.
Šlapeta, J., Moreira, D., and López-García, P. (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. *Proc Biol Sci* **272:** 2073–2081.
Šlapeta, J., López-García, P., and Moreira, D. (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol* **23:** 23–29.
Taib, N., Mangot, J.-F., Domaizon, I., Bronner, G., and Debroas, D. (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. *PLoS ONE* **8:** e58950.
Triadó-Margarit, X., and Casamayor, E.O. (2012) Genetic diversity of planktonic eukaryotes in high mountain lakes (Central Pyrenees, Spain). *Environ Microbiol* **14:** 2445–2456.
Wright, R.T., and Coffin, R.B. (1984) Measuring microzooplankton grazing on planktonic marine bacteria by its impact on bacterial production. *Microb Ecol* **10:** 137–149.
Yuan, C.L., Keeling, P.J., Krause, P.J., Horak, A., Bent, S., Rollend, L., and Hua, X.G. (2012) Colpodella spp.-like parasite infection in woman, China. *Emerg Infect Dis* **18:** 125–127.
Zhao, B., Chen, M., Sun, Y., Yang, J., and Chen, F. (2011) Genetic diversity of picoeukaryotes in eight lakes differing in trophic status. *Can J Microbiol* **57:** 115–126.
Zubkov, M.V., and Tarran, G.A. (2008) High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **455:** 224–226.

Supporting information

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

Fig. S1. Shallow freshwater systems sampled in this study in April 2011.
A. Localization of the five ecosystems in the Parc Naturel Régional de la Haute Vallée de Chevreuse, SouthWest of Paris, France (http://www.parc-naturel-chevreuse.fr).
B. Distances as the crow flies between the sampled ponds and brook (km). Photographs of the sites are shown on the left.
Fig. S2. Non-metric multidimensional scaling (NMDS) plot showing protist community composition similarities and differences among the five ecosystems studied. MG25 and MG25b: Mare Gabard, SR25 and SR25b: Saint Robert, EV33 and EV33c: Etang des Vallées (0.2–5 μm size sample), EV34 and EV34b: Etang des Vallées (5–30 μm size samples), RSA25 and RSA25b: Sainte Anne, LC25 and LC25b: La Claye.
Fig. S3. Approximate maximum likelihood (ML) phylogenetic tree of partial 18S rRNA gene sequences of cryptophytes, haptophytes, telonemids, katablepharids and centroheliozoans. A total of 354 unambiguously aligned positions were used to reconstruct the tree. Two chlorophyte sequences were used as outgroup (not shown) to root the tree. Representative sequences of OTUs from this work are shown in bold green. For abundant OTUs (> 0.5% reads), their proportions in term of total number of reads and reads from the 0.2–5 μm size fraction are given within square brackets. Statistical local support values higher than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.
Fig. S4. Approximate ML phylogenetic tree of partial 18S rDNA sequences of stramenopiles. A total of 336 unambiguously aligned positions were used to reconstruct the tree. Two alveolate sequences were used as outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values greater than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.
Fig. S5. Approximate ML phylogenetic tree of partial 18S rRNA gene sequences of alveolates. Diloflagellate and ciliate branches are shown collapsed. A total of 325 unambiguously aligned positions were used to reconstruct the tree. Four

stramenopile sequences were used as outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values higher than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

Fig. S6. Approximate ML phylogenetic tree of partial 18S rRNA sequences of alveolates. Branches leading to groups other than dinoflagellates and ciliates (see Fig. S4) are shown collapsed. A total of 325 unambiguously aligned positions were used to reconstruct the tree. Four stramenopile sequences were used as outgroup to root the tree. Representative sequences of OTUs from that work are shown in bold green. Local support values greater than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

Fig. S7. Approximate ML phylogenetic tree of 18S rDNA partial sequences of opisthokonts. A total of 342 unambiguously aligned positions were used to reconstruct the tree. Two amoebozoan sequences were used as outgroup to root the tree. Representative OTU sequences from our work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit of branch length.

Fig. S8. Approximate ML phylogenetic tree of partial 18S rDNA sequences of cercozoans. A total of 353 unambiguously aligned positions were used to reconstruct the tree. Two ciliate sequences were used as an outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit branch length.

Fig. S9. Approximate ML phylogenetic tree of partial 18S rDNA sequences of Archaeplastida. A total of 450 unambiguously aligned positions were used to reconstruct the tree. Two haptophyte sequences were used as an outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit branch length.

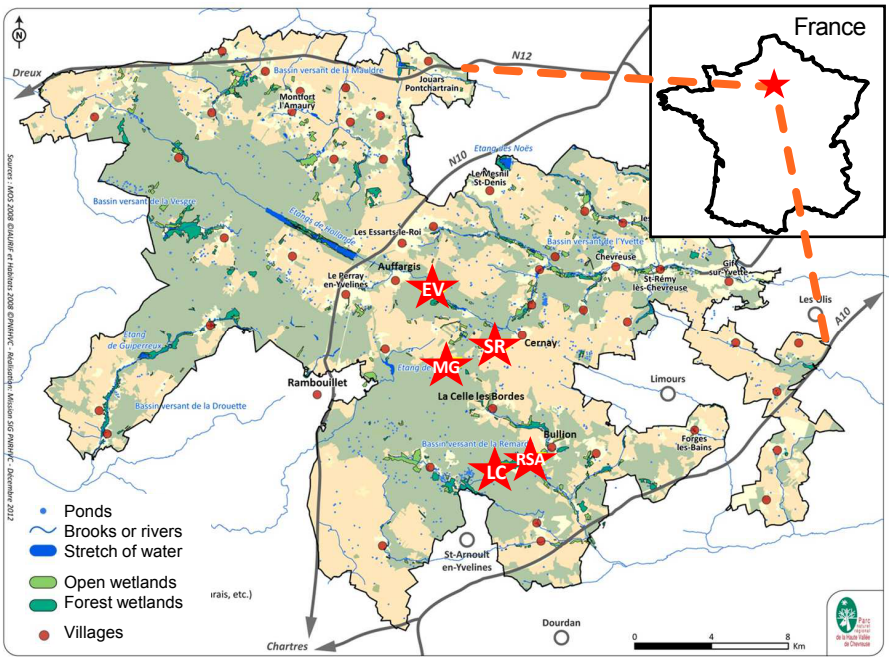
Fig. S10. Canonical correspondance analysis (CCA) plot. All plots are from the same analysis as in Fig. 6, with ellipses emphasizing the distribution of OTUs affiliated to (A) alveolates, (B) rhizarians, archaeplastids and excavates, (C) cryptophytes, (D) centroheliozoans, telonemids, haptophytes, katablepharids and apusozoans, stramenopiles and (E) unikonts. Projected inertia on CCA1: 34.07% and on CA: 29.77%. Black dots correspond to OTUs. Grey triangles indicate the samples. SR, Saint Robert; EV, Etang des Vallées; MG, Mare Gabard; LC, La Claye; RSA, Ru Sainte Anne. Replicate samples are superimposed. DOC, dissolved organic carbon; OrthoP, ortho phosphate; TDS, total dissolved solids.

Table S1. Number and percentage of reads assigned to different taxa in the five studied ecosystems. Replica samples are labelled 'b'.

Table S1. Number and percentage of reads assigned to different taxa in the five studied ecosystems. Replica samples are labeled 'b'.

	Sampled ecosystem		Gabard pond		Saint Robert pond		Etang des Vallées				Sainte Anne brook		La Claye pond	
	Size fraction		0.2 - 5 µm		0.2 - 5 µm		5 - 30 µm		0.2 - 5 µm		0.2 - 5 µm		0.2 - 5 µm	
	Sample names		MG25	MG25b	SR25	SR25b	EV34	EV34b	EV33	EV33c	RSA25	RSA25b	LC25	LC25b
	Nbr of reads per sample →		10616	4197	42034	4506	10982	6906	17670	3947	19652	4243	11191	10605
	Nbr of reads per group ↓													
Opisthokonta	Fungi	2958	2.6	3.5	0.1	0.2	0.9	0.3	0.3	0.3	8.6	11.9	0.7	0.4
Opisthokonta	Ichthyosporea	30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.1	0.0
Opisthokonta	Choanoflagellida	16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Amoebozoa	Discosea	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Amoebozoa	Other Amoebozoa	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Excavata	Metamonada	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Archaeplastida	Chlorophyta	3081	4.7	7.0	0.0	3.1	5.6	3.0	2.1	2.1	3.2	2.9	0.6	0.6
Archaeplastida	Streptophyta	115	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.2	0.3	0.3	0.2
Stramenopiles	MAST	368	0.0	0.0	0.0	0.0	0.6	0.6	0.8	0.0	0.5	0.6	0.0	0.0
Stramenopiles	Bicosoecida	757	2.3	1.2	0.0	0.1	0.5	0.2	1.6	1.0	0.0	0.0	0.4	0.2
Stramenopiles	Bacillariophyceae	8473	0.0	0.0	0.0	0.0	13.1	3.5	4.4	3.1	27.1	13.3	0.0	0.0
Stramenopiles	Oomyceta	1471	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.4	6.1	4.8	0.0	0.0
Stramenopiles	Chrysophyceae	11872	6.2	3.4	8.2	2.2	14.2	12.7	14.9	13.6	5.9	5.1	2.9	2.0
Stramenopiles	Synurophyceae	7964	20.1	32.0	4.3	2.1	7.3	5.7	3.7	4.4	1.4	0.9	1.2	0.8
Stramenopiles	Labyrinthulida	14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Stramenopiles	Xanthophyceae	56	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0
Stramenopiles	Other Stram	5541	1.8	1.0	0.0	0.4	3.0	1.9	5.8	5.7	0.9	0.8	19.1	11.5
Alveolata	Ciliophora	23037	36.2	16.1	12.1	19.8	21.8	33.9	28.8	37.7	2.9	1.3	3.1	2.5
Alveolata	Dinoflagellata	730	3.7	4.6	0.0	0.0	0.6	0.3	0.3	0.0	0.0	0.0	0.0	0.0
Alveolata	Apicomplexa	14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Alveolata	Perkinsea	362	0.5	0.8	0.0	0.0	0.3	0.2	0.3	0.3	0.3	0.3	0.3	0.4
Alveolata	Other Alveolata	1273	0.0	0.0	0.0	0.7	4.9	2.0	2.2	3.3	0.2	0.1	0.1	0.0
Rhizaria	Cercozoa	1829	0.1	0.1	0.0	0.1	1.0	1.0	0.6	0.1	5.9	7.5	0.3	0.2
	Haptophyta	24	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
	Cryptophyta	71564	21.9	29.9	75.2	71.4	19.1	25.3	22.8	20.0	31.7	43.7	69.9	80.7
	Centroheliozoa	38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0
	Katablepharida	3630	0.0	0.0	0.0	0.1	6.7	9.2	10.9	7.8	0.1	0.0	0.0	0.0
	Telonemida	18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	Ancyromonadida	17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0
	Apusomonadidae	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Uncertain	1284	0.0	0.2	0.0	0.0	0.1	0.1	0.2	0.1	4.2	5.6	1.1	0.2

A



B

	Gabard	Saint Robert	Etang des Vallées	Sainte Anne	La Claye
Gabard	0	2.11	3.95	5.87	5.2
Saint Robert		0	3.48	6.13	6.30
Etang des Vallées			0	9.46	9.14
Sainte Anne				0	2.48
La Claye					0

Figure S1. Shallow freshwater systems sampled in this study in April 2011. A, localization of the five ecosystems in the Parc Naturel Régional de la Haute Vallée de Chevreuse, SouthWest of Paris, France (http://www.parc-naturel-chevreuse.fr). B, distances ias the crow flies between the sampled ponds and brook (km). Photographs of the sites are shown on the left.

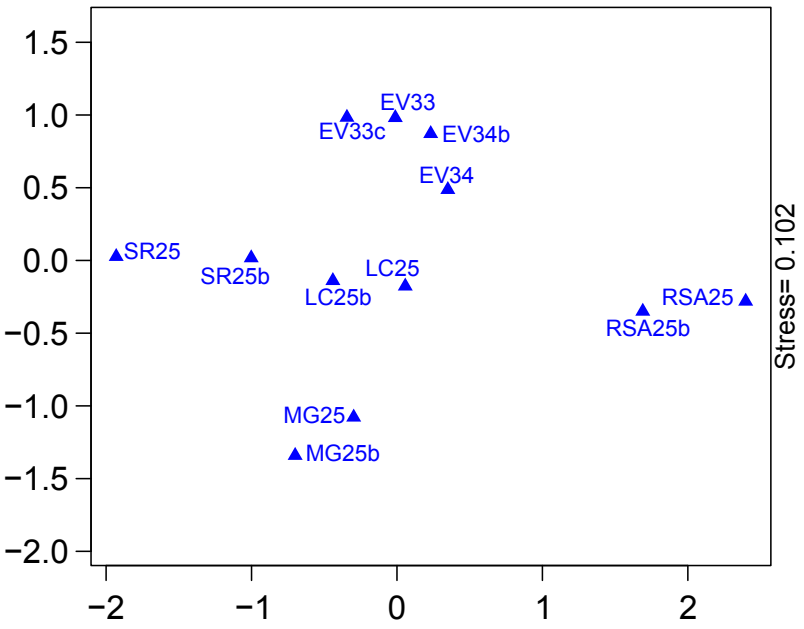


Fig. S2. Nonmetric Multi-Dimensional Scaling (NMDS) plot showing protist community composition similarities and differences among the five ecosystems studied. MG25 and MG25b: Mare Gabard, SR25 and SR25b: Saint Robert, EV33 and EV33c: Etang des Vallées (0.2-5 µm size sample), EV34 and EV34b: Etang des Vallées (5-30 µm size samples), RSA25 and RSA25b: Sainte Anne, LC25 and LC25b: La Claye

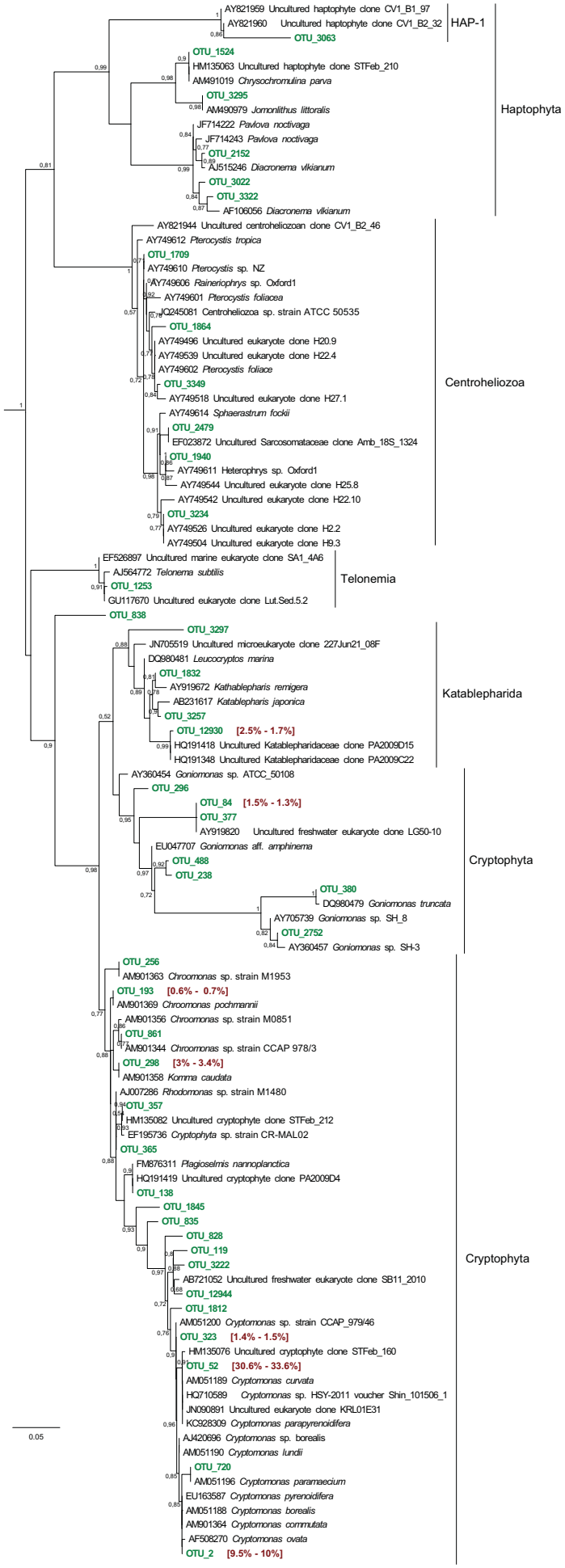


Figure S3. Approximate Maximum Likelyhood (ML) phylogenetic tree of partial 18S rRNA gene sequences of cryptophytes, haptophytes, telonemids, katablepharids and centroheliozoans. A total of 354 unambiguously aligned positions were used to reconstruct the tree. 2 chlorophyte sequences were used as outgroup (not shown) to root the tree. Representative sequences of OTUs from this work are shown in bold green. For abundant OTUs (> 0.5% reads), their proportions in term of total number of reads and reads from the 0.2-5 µm size-fraction are given within square brackets. Statistical local support values higher than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

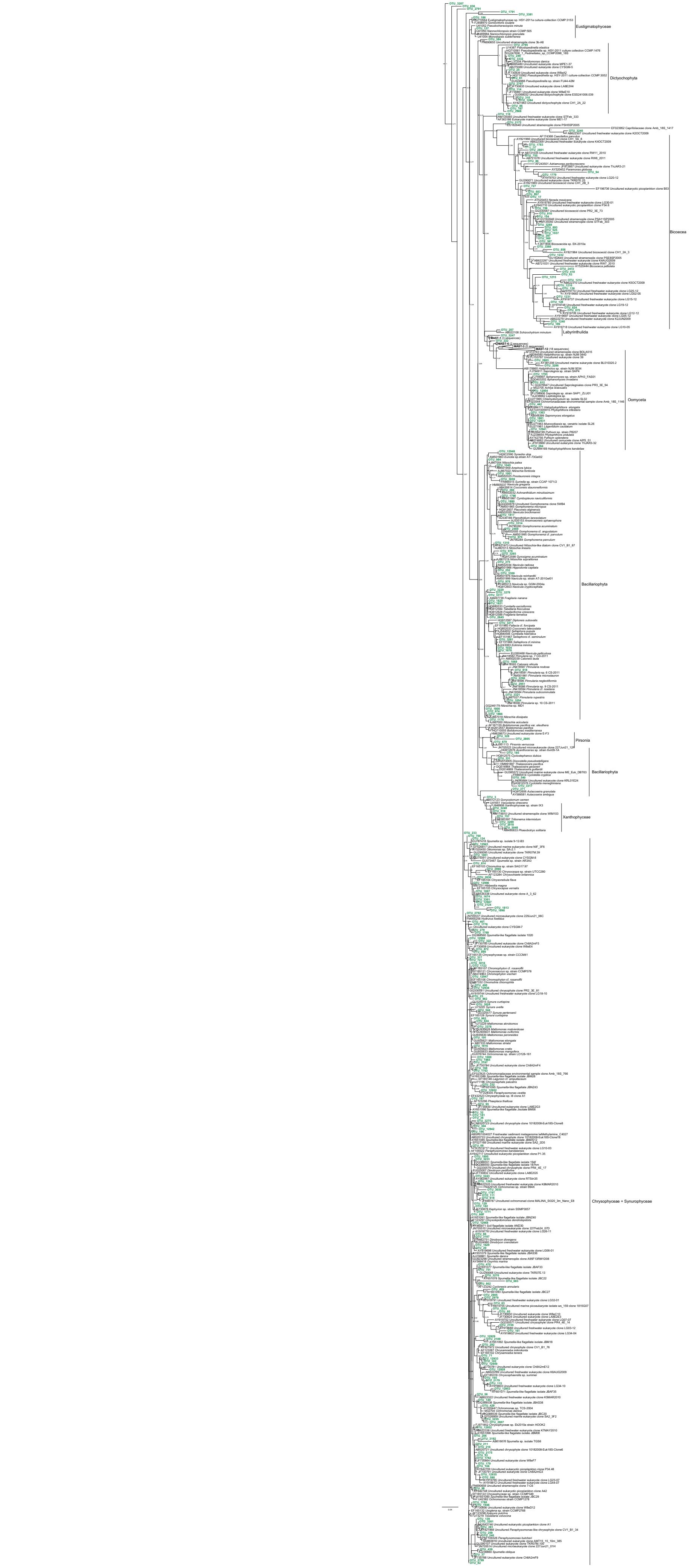


Figure S4. Approximate ML phylogenetic tree of partial 18S rDNA sequences of stramenopiles. A total of 336 unambiguously aligned positions were used to reconstruct the tree. Two alveolate sequences were used as outgroup to root the tree. Representative sequences of OTUs form this work are shown in bold green. Local support values grater than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

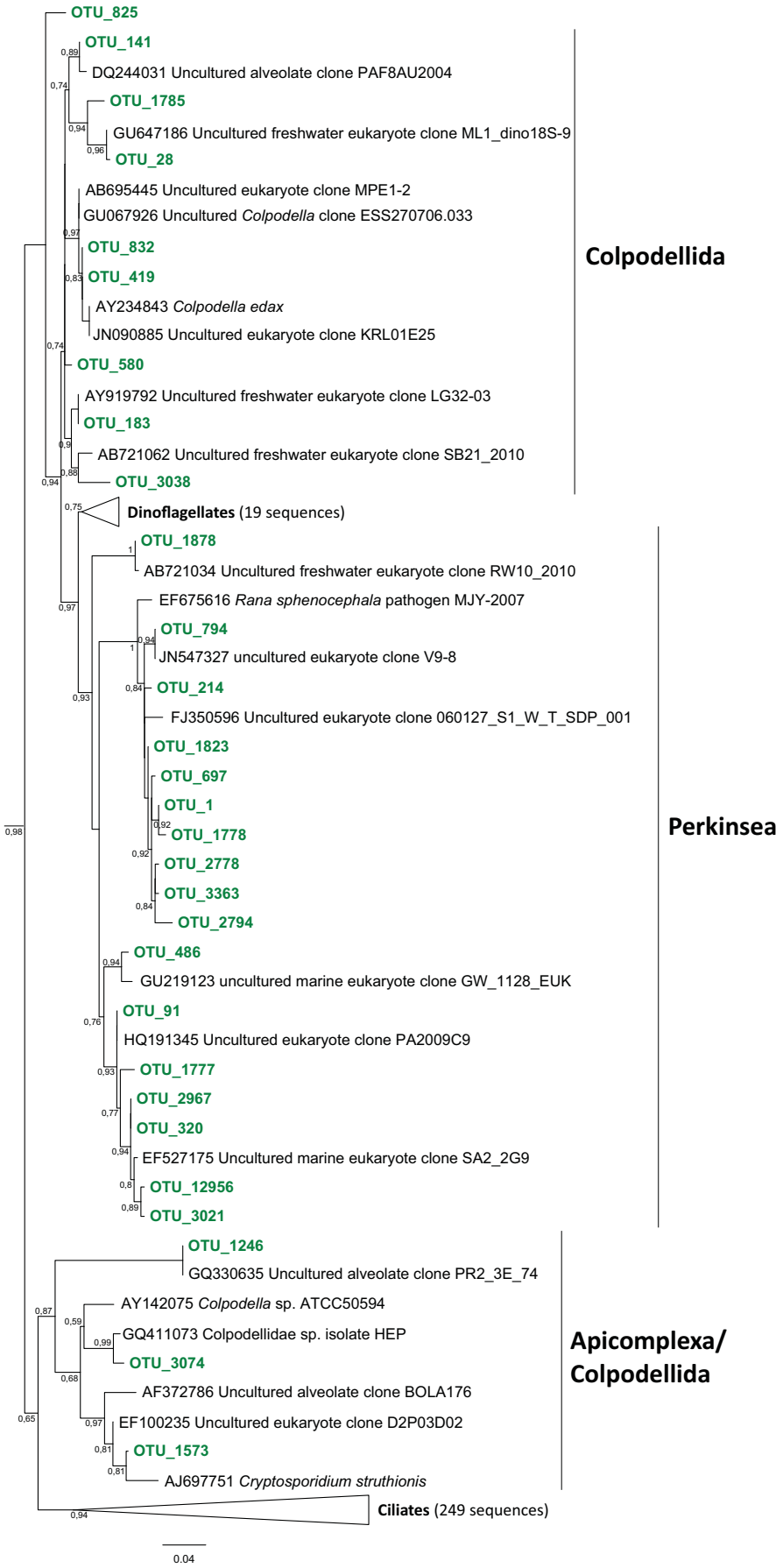


Figure S5. Approximate ML phylogenetic tree of partial 18S rRNA gene sequences of alveolates. Diloflagellate and ciliate branches are shown collapsed. A total of 325 unambiguously aligned positions were used to reconstruct the tree. Four stramenopile sequences were used as outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values higher than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

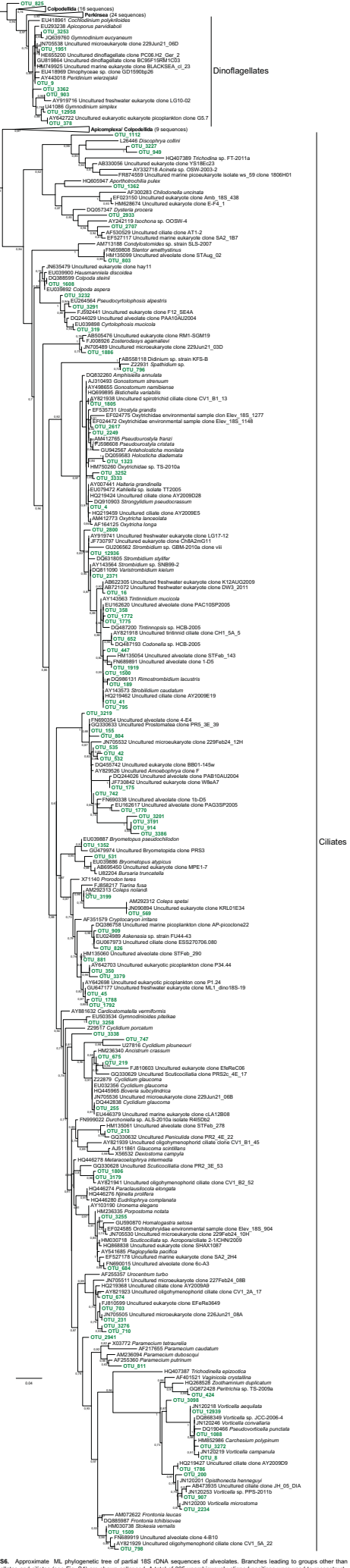


Figure S6. Approximate ML phylogenetic tree of partial 18S rDNA sequences of alveolates. Branches leading to groups other than dinoflagellates and ciliates (see Fig. S4) are shown collapsed. A total of 325 unambiguously aligned positions were used to reconstruct the tree. Four stramenopile sequences were used as outgroup to root the tree. Representative sequences of OTUs from that work are shown in bold green. Local Support values greater than 0.5 are shown at nodes. The scale bar represents the number of estimated substitutions per position for a unit branch length.

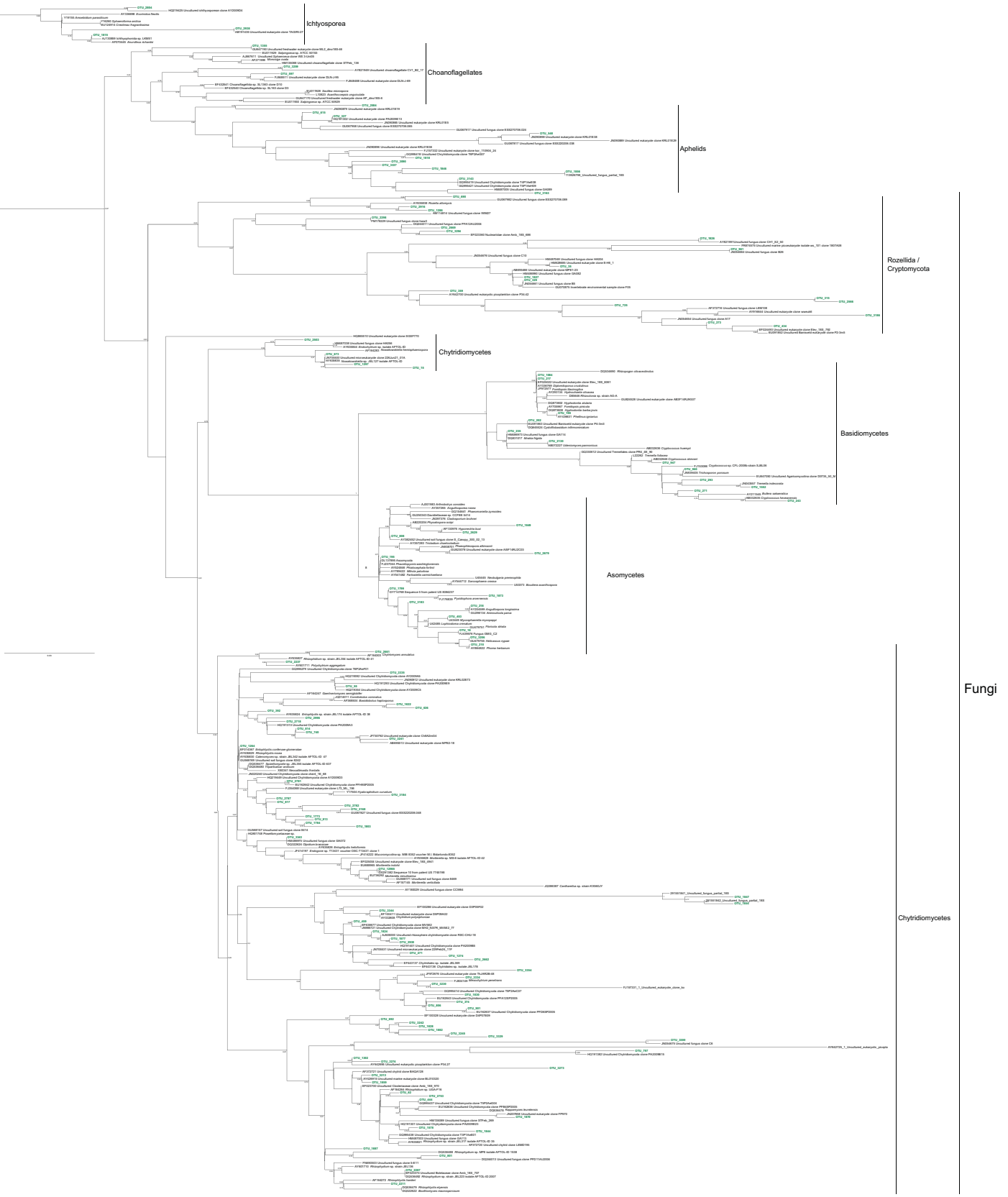


Figure S7. Approximate ML phylogenetic tree of 18S rDNA partial sequences of opisthokonts. A total of 342 unambiguously aligned positions were used to reconstruct the tree. Two amoebozoan sequences were used as outgroup to root the tree. Representative OTU sequences from our work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit of branch length.

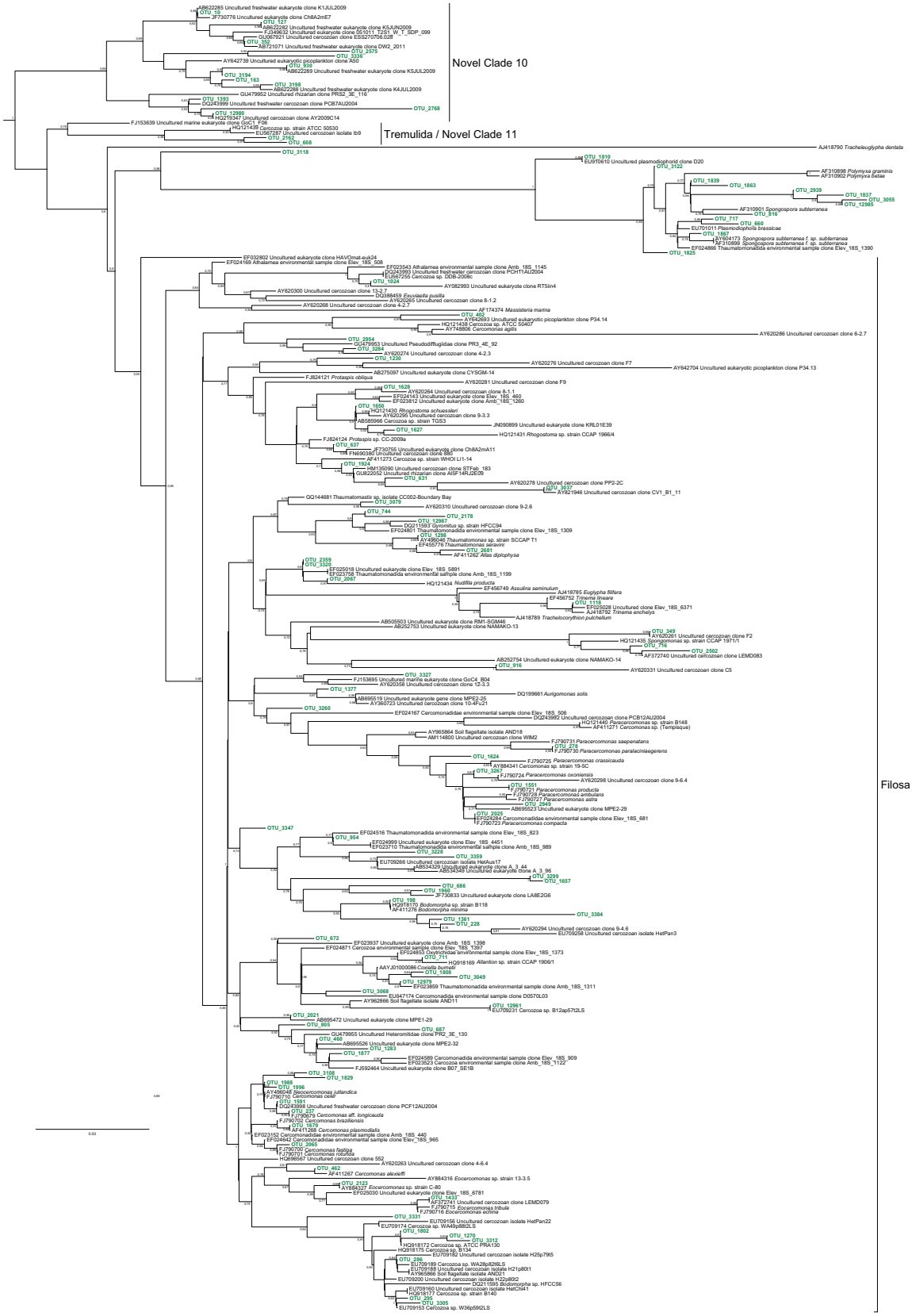


Figure S8. Approximate ML phylogenetic tree of partial 18S rDNA sequences of cercozoans. A total of 353 unambiguously aligned positions were used to reconstruct the tree. Two ciliate sequences were used as an outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit branch length.

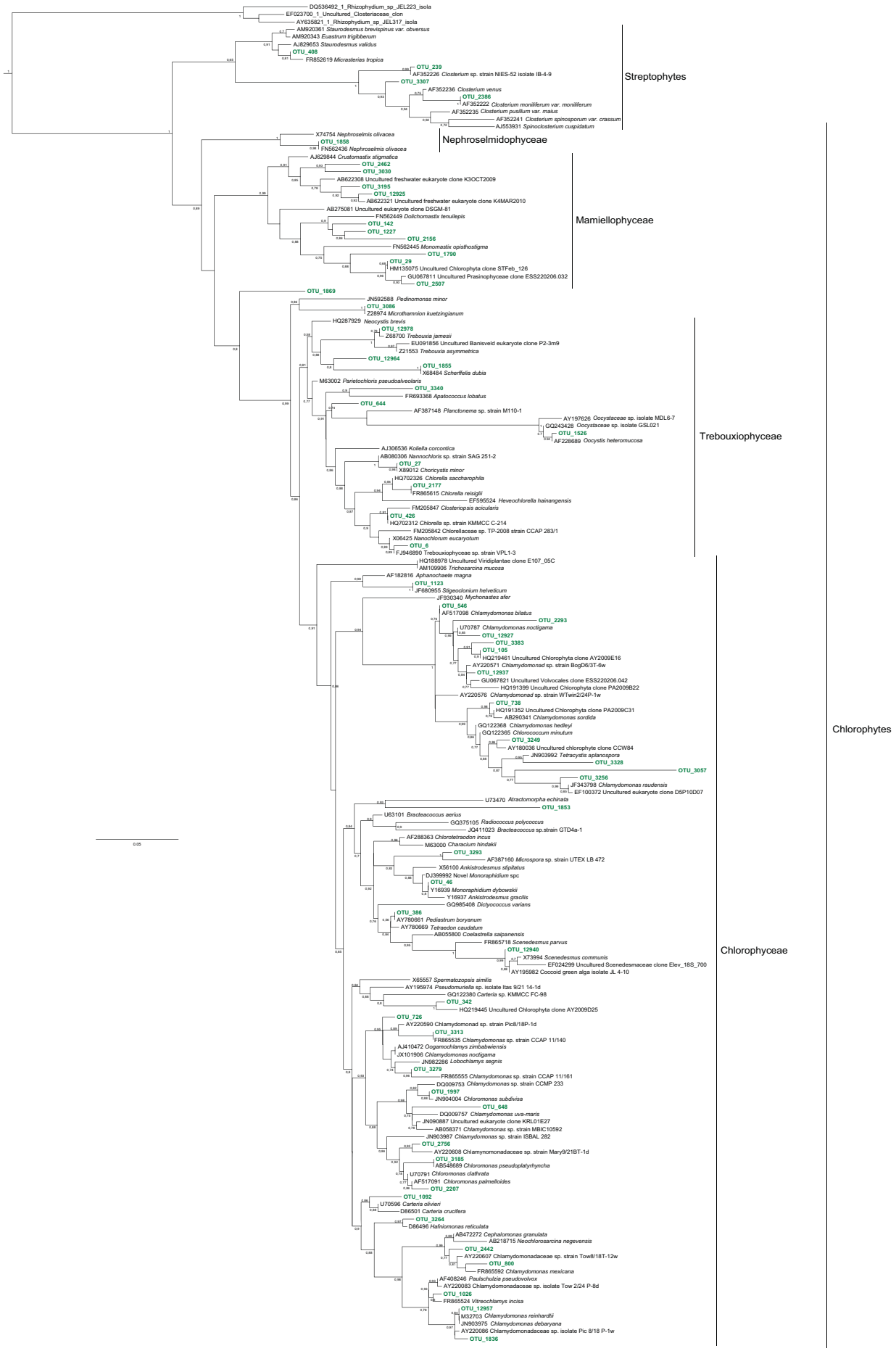


Figure S9. Approximate ML phylogenetic tree of partial 18S rDNA sequences of Archaeplastida. A total of 450 unambiguously aligned positions were used to reconstruct the tree. Two haptophyte sequences were used as an outgroup to root the tree. Representative sequences of OTUs from this work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit branch length.

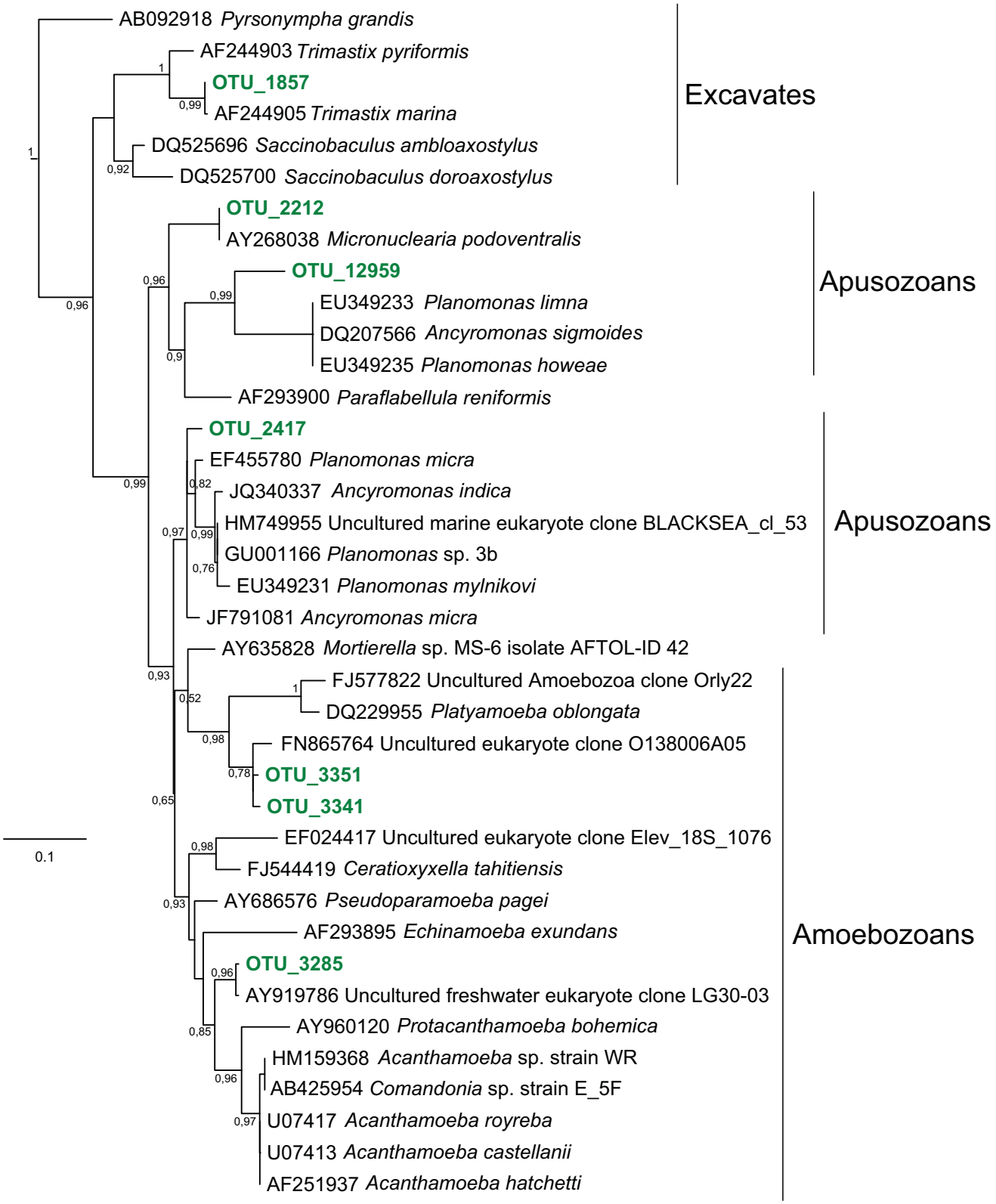


Figure S9. Approximate ML phylogenetic tree of partial 18S rDNA sequences of excavates, apusozoans and amoebozoans. A total of 339 unambiguously aligned positions were used to reconstruct the tree. Two bacterian sequences were used as an outgroup to root the tree. Representative sequences of OTUs from that work are shown in bold green. Local support values greater than 0.5 are shown at nodes. Scale bar represents the number of estimated substitutions per position for a unit branch length.

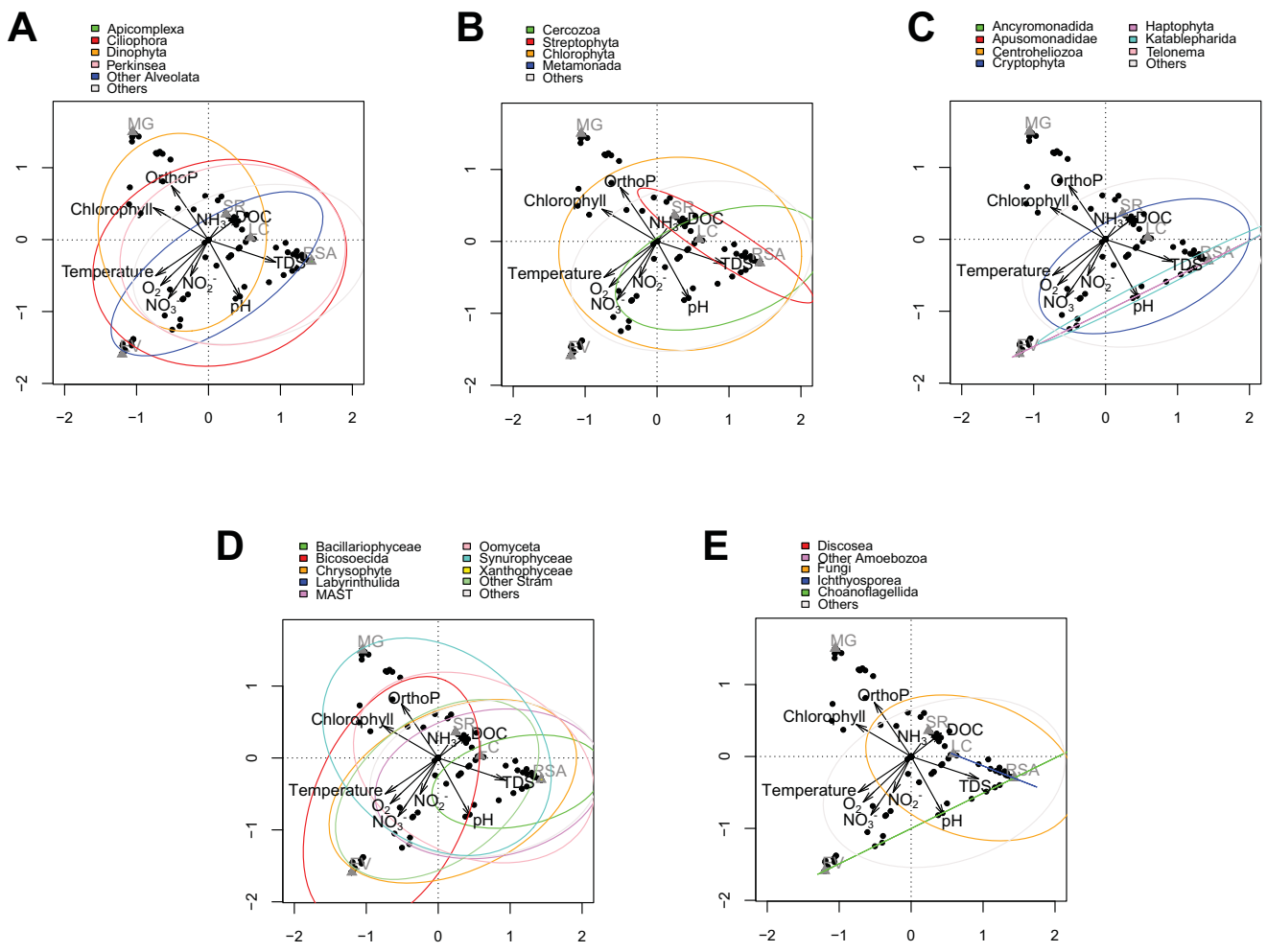


Figure S10. Canonical Correspondance Analysis (CCA) plot. All plots are from the same analysis as in Fig. 6, with ellipses emphasizing the distribution of OTUs affiliated to alveolates (A), rhizarians, archaeplastids and excavates (B), cryptophytes, centroheliozoans, telonemids, haptophytes, katablepharids and apusozoans (C), stramenopiles (D) and unikonts (E). Projected inertia on CCA1: 34.07% and on CA: 29.77%. Black dots correspond to OTUs. Grey triangles indicate the samples. SR, Saint Robert; EV, Etang des Vallées; MG, Mare Gabard; LC, La Claye; RSA, R u Sainte Anne. Replicate samples are superimposed. DOC, Dissolved Organic Carbon; OrthoP, Ortho Phosphate; TDS, Total Dissolved Solutes.