Biodivers Conserv (2008) 17:261–276
DOI 10.1007/s10531-007-9250-y

ORIGINAL PAPER

"Missing" protists: a molecular prospective

Slava Epstein · Purificación López-García

Received: 13 March 2007 / Accepted: 30 May 2007 / Published online: 24 October 2007

Abstract Molecular ecology methods based on 18S rRNA amplification and sequencing have revealed an astounding diversity of microbial eukaryotes in every environment sampled so far. This is certainly true of new species and genera, as essentially every new survey discovers a wealth of novel diversity at this level. This is almost certain for taxa that are higher in taxonomic hierarchy, as many molecular surveys reported novel clades within established protistan phyla, with some of these clades repeatedly confirmed by subsequent studies. It may also be that the molecular approaches discovered several lineages of the highest taxonomic order, but this claim has not been vigorously verified as yet. Overall, the field of protistan diversity remains in its infancy. The true scale of this diversity is unknown, and so are the distribution of this diversity, its patterns, spatial and temporal dynamics, and ecological role. The sampled diversity appears to be just the tip of the iceberg, and this offers outstanding opportunities for microbial discovery for the purposes of both basic and applied research.

Keywords 18S rRNA · Cryptic species · Eukaryotic phylogeny · Molecular ecology · Species richness

Introduction

Molecular analyses based on the rRNA approach have become an affordable and practical tool that is used routinely to study microbial diversity in environmental samples. This type

Special Issue: Protist diversity and geographic distribution. Guest editor: W. Foissner.

S. Epstein
Marine Science Center, Northeastern University, Nahant, MA 01908, USA
e-mail: slava.epstein@gmail.com

P. López-García (✉)
Unité d'Ecologie, Systématique et Evolution, CNRS UMR8079, Université Paris-Sud,
Orsay Cedex 91405, France
e-mail: puri.lopez@u-psud.fr

of analyses was designed to bypass inability to cultivate the majority of prokaryotic organisms in vitro, and provided an excellent tool to classify the organisms by placing the sequences of this marker gene in a phylogenetic tree of life (Olsen et al. 1986; Woese 1987). Results accumulated over the past two decades are impressive. They are more than "just" 300,000 SSU rDNA prokaryotic sequences now available from GenBank,—they have changed our view of the biosphere (Pace 1997).

Compared to prokaryotes, the application of the SSU rRNA gene-based approach to study protistan diversity is very recent (Díez et al. 2001; López-García et al. 2001; Moon-van der Staay et al. 2001; Moreira and López-García 2002). However, even the limited information obtained so far shows certain similarities as well as dissimilarities with what is known about prokaryotic molecular diversity. Contrary to prokaryotic microbiologists, some protozoologists maintain that the principal protistan taxa have been identified during the era of α-taxonomy, and they question the very possibility of discovering novel high ranking candidate protistan taxa (e.g. Finlay 2002; Cavalier-Smith 2004). This is not necessarily so for novel subgroups, at perhaps the class level down to species, where the newly discovered diversity appears impressive (Moreira and López-García 2002; Berney et al. 2004; Richards and Bass 2005). Furthermore, although still quite unstudied, the genetic diversity within protist morphospecies might be surprisingly high as revealed by SSU rRNA gene sequences (Boenigk et al. 2005), the combination of these with intergenic spacer sequences (ITS) (Katz et al. 2005; Rodriguez et al. 2005), and multilocus sequence analysis (Katz et al. 2006; Slapeta et al. 2006a, b). These findings call into question the validity of the phenotype only-based species concept, and open new avenues to explore protist biogeography (Foissner 2006), ecological specialization, and mechanisms underlying protist speciation. In the present review, we briefly summarize recent data on the molecular diversity of microbial eukaryotes.

The molecularly identified but undescribed diversity of protists

Power and pitfalls of molecular phylogenetic identification

The first molecular eukaryotic diversity studies in the early 2000s revealed three major categories of SSU rRNA gene sequences: sequences closely related to known species and genera genes, sequences forming divergent groups within known phyla, and sequences without a clear affiliation to any described eukaryotic phyla. The latter were more frequently identified from anoxic habitats, which have been traditionally less studied, such as various anoxic sediments (Dawson and Pace 2002; Stoeck and Epstein 2003), including hydrothermal sediments (Edgcomb et al. 2002; López-García et al. 2003), and the anoxic water column in, e.g., the Cariaco Basin of the coast of Venezuela (Stoeck et al. 2003). Some authors claimed that certain highly divergent lineages could possibly represent new eukaryotic kingdoms (Dawson and Pace 2002). This claim was subsequently contested (Berney et al. 2004; Cavalier-Smith 2004). The reanalysis of sequences deposited in databanks until 2003 suggested that there had probably been an overestimation regarding the 'novelty' of many of these sequences, due to their misplacement in eukaryotic phylogenies (Berney et al. 2004). There are 3 principal factors that account for such misplacement:

1. PCR of mixed DNA templates generates, with certain frequency, chimeric sequences. Partial treeing analyses is widely used to identify such sequences, together with visual inspection of alignments, and the use of several software programs e.g., CHIMERA_CHECK and BELLEROPHON, both available at the RDP (Maidak et al. 2001).

2. Long-branch attraction due to heterogeneous rates of evolution between lineages (Philippe et al. 2000; Brinkmann et al. 2005). This artifact can be minimized by using methods that are less prone to it (e.g., maximum likelihood or Bayesian analyses), including an appropriate selection of slowly-evolving lineages and avoiding the inclusion of other fast-evolving sequences and distant outgroups. The combined use of several phylogenetic markers, such as concatenations of multiple protein-coding genes generated from EST (expressed sequence tags) of cultivated protists (O'Brien et al. 2007), or the combination of small and large subunit rRNA genes (Moreira et al. 2007) can significantly improve problematic molecular phylogenies.
3. Insufficient taxon sampling (Philippe et al. 2000), an artifact especially important in protistan phylogeny because a significant number of known eukaryotic taxa presently lack representative rRNA gene sequences in databases. A case in point is several protistan taxa that reached a stable phylogenetic position once a significant number of sequences of other members of that group were included (see Table 1 in Berney et al. 2004). Pertinent examples include AT4-11 (Apusozoa) (López-García et al. 2003), BOLA187 and BOLA366 (*'Mastigamoeba invertens'* group) (Dawson and Pace 2002), CS_E022 (jakobid) and C1_E027 *'Retortamonas/Carpediemonas'* group) (Edgcomb et al. 2002). There are also several examples of orphan SSU rRNA sequences that eventually found a "home" taxon in what originally simply lacked representative sequences (Fig. 1). For instance, DH145-KW16, a clone that branched at the base of the Acantharea initially (López-García et al. 2002), and CS_E043 (Edgcomb et al. 2002), were shown to be members of the Taxopodida once the sequence of the heliozoan-like *Sticholonche zanclea* was determined (Nikolaev et al. 2004). Further, Fig. 1 shows that the deep-sea radiolarian-like sequence DH45-HA2 (López-García et al. 2002) turned out to cluster with the Nassellaria and various other radiolarian-like sequences from deep-sea plankton (DH148-EKD30), anoxic sediments (E32, E16, E215, E191, E196) and hydrothermal sediment (CS_E004, C1_E045) with a group of single-celled spumellarians, both of which lacked representative sequences until recently (Kunitomo et al. 2006). Another example relates to the kathablepharids, an important flagellate component of marine and freshwater systems distantly related to the cryptophytes, and for which SSU rRNA gene sequences were not available until recently (Okamoto and Inouye 2005). Upon their determination, a number of environmental sequences retrieved in various previous studies could be affiliated to this flagellate group (Slapeta et al. 2006a, b). A more extreme case is the phylotypes DH145-EKD11 from deep-sea plankton (López-García et al. 2001) and CCW75 from anoxic marine sediment (Stoeck and Epstein 2003). These were initially placed as basal unresolved lineages, but once the first representative sequences (*Myrionecta* and *Mesodinium* spp.) became available, they were shown to cluster with the Mesodinidae, a group of fast-evolving ciliates (Johnson et al. 2004).

An additional problematic aspect of rRNA surveys is that they do not necessarily identify live cells, and are likely to detect allochthonous species of little importance for the community under study. An alternative is to target rRNA via cDNA libraries, an approach that has only recently been employed in the study of microbial eukaryotes (Stoeck et al. 2007b).

Despite the above problems, molecular environmental analyses of protist diversity are revealing a large variety of protists within high-level taxa, and are helping to define their genetic diversity in nature. However, we can also predict that large protistan diversity

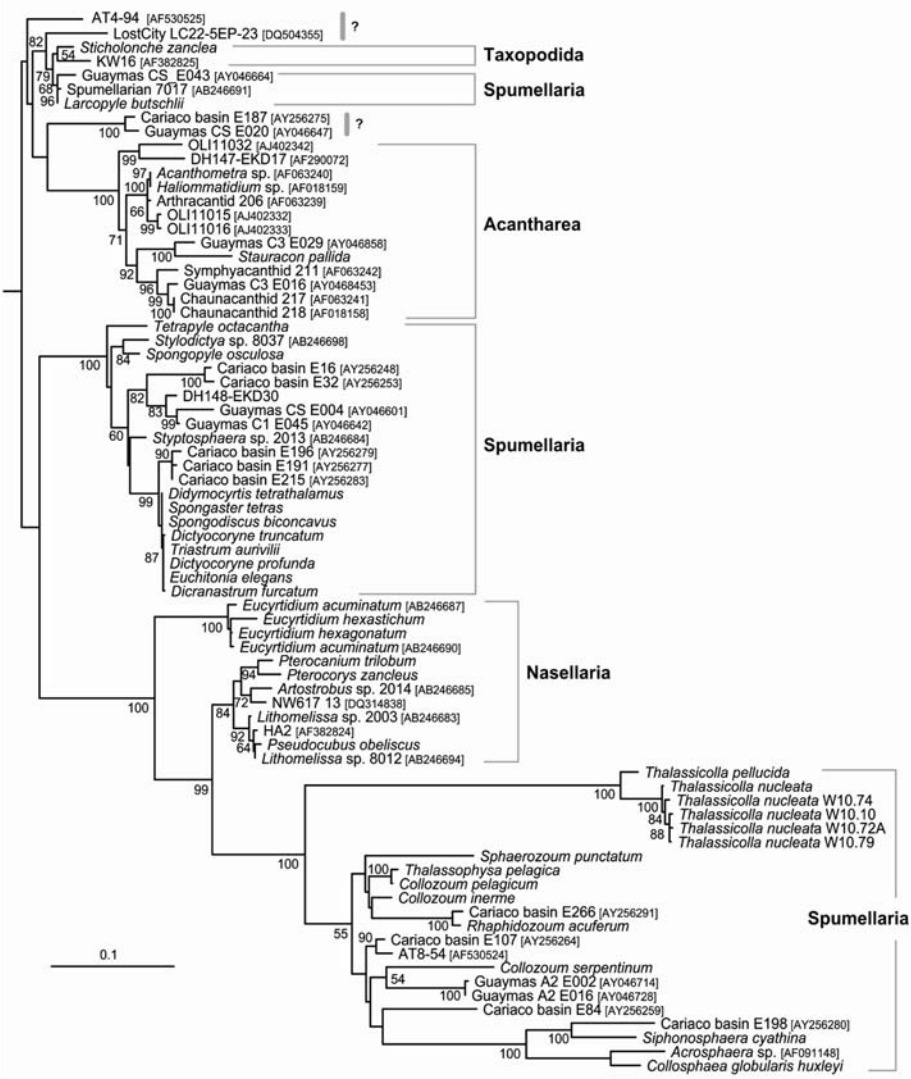

Fig. 1 Phylogenetic tree showing the position of environmental radiozoan 18S rRNA gene sequences retrieved from several oceanic locations by molecular diversity analyses. The tree was constructed by maximum likelihood (PhyML) using 901 unambiguously aligned positions, applying a GTR + Γ + I model of sequence evolution. Bootstrap values (500 replicates) higher than 50% are shown at nodes. The tree was rooted with ten cercozoan sequences

escapes molecular identification due to primer biases and other PCR-related biases (Stoeck et al. 2006; Stoeck et al. 2007a). It is quite possible that some groups are rarely, or never, captured by environmental libraries that employ general eukaryotic primers (or even combinations thereof), e.g. many fast evolving parasitic lines (gregarines, excavate parasites), many amoeboid lineages, or rhodophytes. The use of taxon-specific primers can reveal a much richer diversity, as was recently shown for the Cercozoa (Bass and Cavalier-Smith 2004).

Novel protistan lineages

Undeniably, molecular environmental surveys of eukaryotes have uncovered a wide range of SSU rRNA phylotypes that belong to previously described phyla. Some of them are represented by single sequences unrelated to others within the phylum, and since the position of singletons is often unstable, their true phylogeny remains to be validated. Others form novel clusters composed of sequences reported by independent studies, which is strongly indicative of their reality. One such example is the cluster formed by DH148-5-EKD18, CS_R003, BOLA048, Sey010 and Sey017 from deep-sea plankton (López-García et al. 2001), hydrothermal sediment (Edgcomb et al. 2002), anoxic tidal sediment (Dawson and Pace 2002) and river sediment (Berney et al. 2004, see their Fig. 2). However, the branch leading to this clade is very long suggesting that the whole group may not represent a truly novel taxon but be misplaced from its real position by the long-branch attraction artifact. Some of the clusters remain good candidates for new phylum-level groups (Berney et al. 2004), although validation by complementary approaches, including in situ fluorescence hybridization methods (FISH), and ultimately, morphological and structural characterization, will be required.

In addition, there are several cases of larger clusters of sequences, also generated by independent studies, that branch together and form defined monophyletic groups with no described representatives. The first molecular diversity studies carried out on marine picoplankton revealed the presence of two large clusters of alveolate sequences, initially termed marine alveolates Group I and II (Díez et al. 2001; López-García et al. 2001; Moon-van der Staay et al. 2001; Moreira and López-García 2002). Sequences from both groups have so far been identified exclusively in marine environments. The genetic diversity within both groups is substantial and comparable to that of, for example, the ciliates. The sequence of *Amoebophrya* sp., a parasite of the dinoflagellate *Gymnodinium sanguineum* (Gunderson et al. 1999) from the order Syndiniales, clusters with those of Group II, suggesting that at least some Group II alveolates may be parasites, and their registered diversity may be linked to that of their host organisms. Indeed, it had been previously suggested that, based on SSU rRNA gene sequences, *Amoebophrya ceratii* was a complex of species (Janson et al. 2000). However, while Group II can be identified with at least some known species, marine alveolate Group I cannot. Only recently, the first hint came about their biology, i.e., the possible symbiotic nature of its members, as Group I alveolate sequences were amplified from single radiolarian cells (Dolven et al. 2007).

Within the stramenopiles (heterokonts), several clusters of SSU rRNA sequences were identified with no known representatives; these originated mostly from the upper (photic) ocean environment. Massana and co-workers described up to 12 clusters of likely heterotrophic stramenopiles, with four MAST clades comprising approximately 75% of the total number of stramenopile sequences deposited in public databases (Massana et al. 2004a, b). In a recent study, Richards and Bass (2005) registered up to 16 stramenopile clusters. We note, however, that the entire clade of stramenopiles is unstable, casting doubt in the reality of so many independent stramenopile groups. Some of these newly detected clades are supposed to be heterotrophic because they branched at the base of the stramenopile clade, far from typical photosynthesizing lineages. The heterotrophic lifestyle was later confirmed by FISH, which suggested that at least some of these stramenopiles are bacterivorous (Massana et al. 2002, 2006). In addition to heterotrophic lines, a novel clade of photosynthetic stramenopiles was detected in Artic waters (Lovejoy et al. 2006). The MAST clades remain an exciting target for microbial discovery, and first steps have been

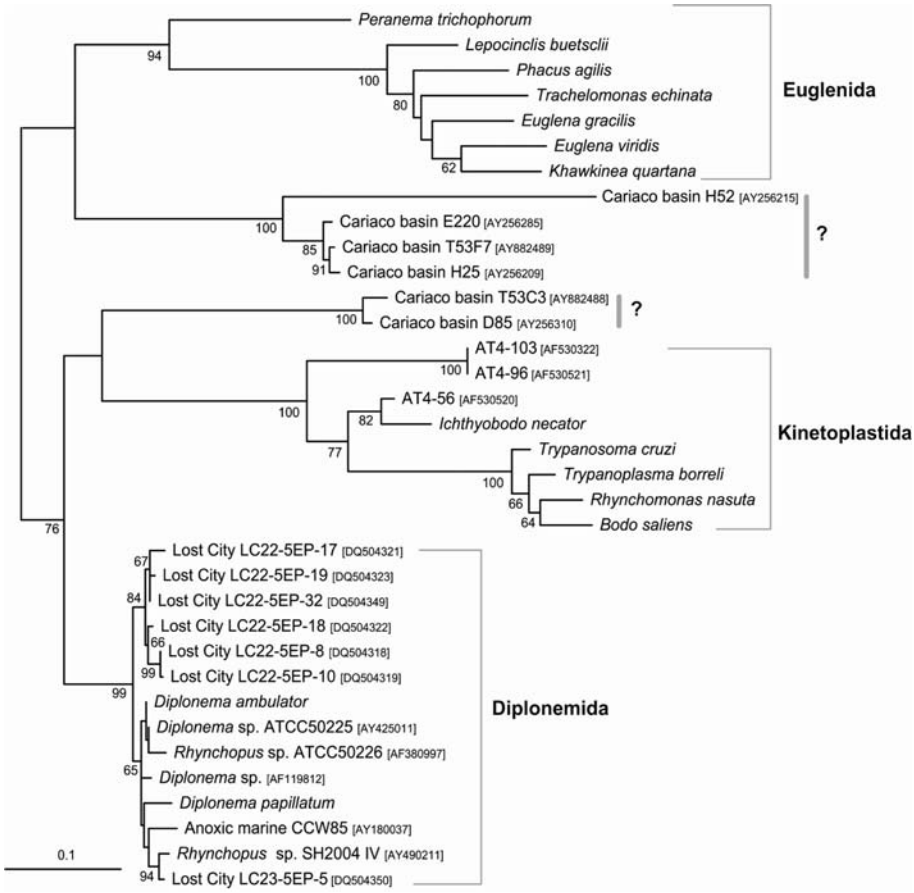

Fig. 2 Phylogenetic tree showing the position of environmental euglenozoan 18S rRNA gene sequences retrieved by molecular diversity analyses from deep-sea plankton, deep-sea hydrothermal vent areas (Lost City and Rainbow, Mid-Atlantic Ridge) and the anoxic Cariaco basin. The tree was constructed by maximum likelihood (PhyML) using 500 unambiguously aligned positions (the limited number of positions used is imposed by the presence of a few partial sequences), applying a GTR + Γ + I model of sequence evolution. Bootstrap values (500 replicates) higher than 50% are shown at nodes

taken towards recovering and visualizing cells of this elusive group (Kolodziej and Stoeck 2007).

While most of the newly discovered protistan diversity falls within known clades, such as alveolates and stramenopiles, there are several lineages possibly representing novel protistan groups of higher taxonomic rank. Typically, these lineages consist of few sequences obtained by a limited number of studies. For instance, several sequences retrieved from deep-sea plankton and fluid-seawater interfaces in the Lost City hydrothermal site formed a robust clade sister to the diplonemids within the Euglenozoa (López-García et al. 2007). This clade was clearly distinct from other euglenozoan sequences identified by Stoeck et al. (2006) in anoxic sediments (Fig. 2). The sequences RT5iin14 and RT5iin16 could represent a group at the base of the opisthokonts (Amaral-Zettler et al. 2002). Finally, various divergent phylotypes might form clades within Amoebozoa or even outside the six recognized eukaryotic supergroups (e.g. in Dawson and Pace 2002; Edgcomb et al. 2002;

López-García et al. 2003; Stoeck and Epstein 2003). However, the true extent of their novelty is not clear, primarily because of the long branch attraction phenomenon.

Oceanic planktonic protists

The ocean was among the first ecosystems to be explored by the rRNA approach, and remains one of the better studied environments. Most environmental surveys have focused on the photic zone picoplankton (<2 μm) and the smallest fractions of nanoplankton (2–20 μm), usually separated by prefiltration steps. The existence of picoeukaryotic algae had been recognized for some time (Courties et al. 1994; Guillou et al. 1999; Moon-van der Staay et al. 2000), but the real extent and ecological importance of picoeukaryotes was not revealed until the first molecular studies of the smallest planktonic fractions (Worden et al. 2004). In general, the same taxa are identified in various oceanic locations, though the relative proportions of lineages fluctuate between studies. Members of the Prymnesiophyta (haptophytes), Pelagophyta (stramenopiles) and Prasinophyta (green algae) appear among the more abundant picoeukaryotic photosynthesizes in SSU rRNA gene libraries from various photic oceanic locations, including the equatorial Pacific, Mediterranean, Arctic, and the Baltic Sea (Díez et al. 2001; Moon-van der Staay et al. 2001; Guillou et al. 2004; Lovejoy et al. 2006; Medlin et al. 2006; Worden 2006). Prasinophytes, especially the genera *Ostreococcus* or *Micromonas*, appear particularly frequently in the libraries, and although common PCR-based methods are not quantitative, their relative abundance in clone libraries is suggestive of their importance in the marine ecosystem (Guillou et al. 2004; Worden 2006). A few studies have analyzed the SSU rRNA gene diversity in planktonic fractions larger than 5 μm. In this case, the photosynthetic component in clone libraries is dominated by diatoms or photosynthetic dinoflagellates (Savin et al. 2004), although members of picoplanktonic size were also be detected (Yuan et al. 2004).

The likely heterotrophic fraction of surface picoplankton encompasses sequences belonging to early-branching stramenopiles, alveolates Group I and II, ciliates and radiolarians. Groups present less frequently include cercozoans, cryptophytes, and fungi. Dinoflagellate phylotypes are detected regularly but it is not clear whether they are photosynthetic, heterotrophic, or mixotrophic. Stramenopiles appear to dominate heterotrophic picoplankton, accounting for up to 35% of clone libraries targeting the total heterotrophic flagellates (Díez et al. 2001; Massana et al. 2006). In several other cases, ciliates and Group II (Syndiniales) alveolates were the most abundant in libraries (Medlin et al. 2006). These differences can be explained by environmental variables, as well as variation in spatial and temporal parameters.

Deep-sea environment

Molecular diversity of deep-sea picoplankton has been addressed in only a handful of studies. López-García et al. (2001) focused on the aphotic water column from 250 to 3,000 m depth in the Antarctic Polar Front, as well as picoplankton from the fluid-seawater interface around deep-sea vents in the Atlantic (López-García et al. 2003; López-García et al. 2007). Compared to surface waters, deep-sea samples are characterized by a lack of typical photosynthetic lineages, whose occasional presence may be attributed to sinking organisms. Alveolates, particularly Groups I and II, appear to dominate deep-sea planktonic samples, together with radiolarians, a group also well represented in clone libraries constructed from deep-sea samples.

Protists from hydrothermal vent communities have attracted considerable attention. Deep-sea vents are generally associated with mid oceanic ridges typically found between 1,500 and 3,000 m below the surface. They have fascinated the scientific community because they might be analogous to the earliest ecosystems (Reysenbach and Cady 2001). The study of protists from vent areas started only several years ago (Atkins et al. 2000), and has been limited to hydrothermal sediments in the Guaymas basin in the Pacific (2,000 m depth) (Edgcomb et al. 2002), hydrothermal sediments, fluid-seawater mixtures, and experimental substrates exposed in the Mid-Atlantic Ridge in 1,600 and 2,200 m depth (López-García et al. 2003), the anoxic surroundings of shallow fumaroles at the Kagoshima bay (200 m depth) (Takishita et al. 2005), and the carbonates and fluid-seawater interface in the alkaline Lost City vent field (750 m depth) located off-axis at the Mid Atlantic Ridge (López-García et al. 2007). Of these, only hydrothermal anoxic sediments appear to contain potentially new lineages of the highest taxonomic level, some of which have been detected more that once by independent researchers (Berney et al. 2004). The rRNA signatures from other hydrothermal habitats vary by site, and are often dominated by alveolates, especially of apparently parasitic life style, such as Syndiniales, Perkinsozoa and Gregarina (Moreira and López-García 2003), and ciliates. The latter may be bacterivorous as they are typical in surfaces colonized by bacteria. These are followed by the Euglenozoa, both kinetoplastids and diplonemids, fungi, radiolarian and cercozoan sequences (López-García et al. 2007).

Sediments and other oxygen-depleted environments

Anoxic environments hold a special promise for the discovery of microbial eukaryotes (Sogin et al. 1989) because eukaryotic cells might have evolved before oxygen in the atmosphere reached its present level oxygen concentration (Fenchel and Finlay 1995; Brocks et al. 1999). The exploration of 18S rRNA gene diversity of anoxic habitats started only 5 years ago, and is represented by 10 published surveys of anoxic marine water column (Stoeck et al. 2003; Behnke et al. 2006; Stoeck et al. 2006; Zuendorf et al. 2006), shallow marine sediments (Dawson and Pace 2002; Stoeck and Epstein 2003; Takishita et al. 2005), deep-sea hydrothermal vent systems (Edgcomb et al. 2002; López-García et al. 2003), and freshwater sulfidic spring (Luo et al. 2005). While in no way comprehensive, these surveys indicate a degree of uniqueness of protistan diversity uncovered (Richards and Bass 2005), and their collective information allows some preliminary analyses of this diversity.

The most obvious feature of microbial eukaryotes from anoxic environments is a very substantial richness of their communities. The published clone libraries tend to be relatively large, often in excess of 1,000 clones sequenced, with some constructed using multiple PCR primers (Behnke et al. 2006; Stoeck et al. 2006; Zuendorf et al. 2006). Nonetheless, none of surveys came even close to phylotype saturation, indicating a substantial diversity missed (Fig. 3; see more in last section). Equally suggestive is an extremely low overlap between species lists obtained using different PCR primer sets (Behnke et al. 2006; Stoeck et al. 2006; Zuendorf et al. 2006), further indicating that more comprehensive surveys would have uncovered richness several to many times that actually reported.

It is of particular interest to analyze which specific groups of protists were discovered in the anoxic environments. From the beginning of such research (Dawson and Pace 2002; Edgcomb et al. 2002), the surveys of anoxic environments typically reported representatives of essentially all known eukaryotic clades, with an especially high diversity of alveolate, stramenopile, and fungal 18S rRNA gene sequences.

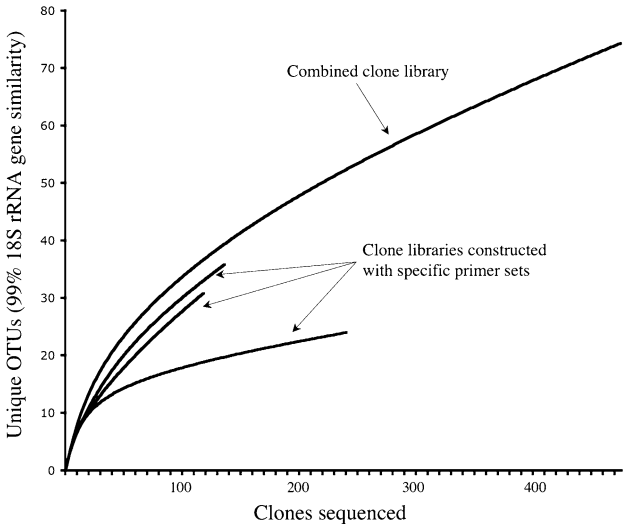

Fig. 3 Accumulation of unique phylotypes as a function of the number of clones sequenced from 18S rRNA gene clone libraries (Bunge et al. unpublished). The sequence data are from Stoeck et al. (2006). Three clone libraries were constructed using different eukaryotic primer sets. The DNA source was a single 3-l sample from anoxic water column in Cariaco Basin, off the coast of Venezuela in the Caribbean

Several specific groups appear particularly well represented in these surveys: Alveolate Group I and ciliates (López-García et al. 2003; Stoeck and Epstein 2003; Stoeck et al. 2003, 2006; Luo et al. 2005; Behnke et al. 2006; Zuendorf et al. 2006); Alveolate Group 3 (Stoeck et al. 2003); Euglenozoa, including Diplonemida (López-García et al. 2003; Stoeck et al. 2003, 2006), Cercozoa (Luo et al. 2005; Stoeck and Epstein 2003; Takishita et al. 2005), and Jakobida (Behnke et al. 2006). Perhaps more interestingly, several studies reported 18S rRNA gene sequences that appeared at the base of several important clades, such as Cercozoa and/or cercomonads (Dawson and Pace 2002; Stoeck and Epstein 2003), Labyrinthulida (Stoeck et al. 2003; Behnke et al. 2006), possibly even stramenopiles in general (Edgcomb et al. 2002; Stoeck and Epstein 2003); as well as several ciliate classes (Stoeck et al. 2003; Behnke et al. 2006).

Very exciting, and equally controversial, are the claims of discovery of rRNA gene signatures unrelated to known extant eukaryotes, as well of novel clades that appear emerge at the basal region of the 18S rRNA tree. Dawson and Pace (2002) proposed "… seven lineages … at the kingdom-level", and pointed out that several rRNA sequences obtained not only branched deeply on the eukaryotic tree, but also appeared rather slowly evolving. Over a dozen of deeply branching lineages unrelated to known eukaryotes were reported from hydrothermal vent and shallow water volcanic systems (Edgcomb et al. 2002; López-García et al. 2003; Takishita et al. 2005). The rRNA signatures were claimed to be found representing sister groups to Euglenozoa, Entamoeba, and diplonemids (Stoeck et al. 2003, 2006) as well as euglenids in general (López-García et al. 2003; Behnke et al. 2006; Stoeck et al. 2006), along with reports on novel clades of an apparently more recent origin (Stoeck et al. 2003; Zuendorf et al. 2006). There is little doubt that at least some of these findings will prove artifactual, and better taxon sampling will uncover the true position on the evolutionary tree of the sequences in question (see section *Power and pitfalls of molecular phylogenetic identification* above). Today, it remains plausible, however, that anoxic

environments do indeed harbor an undescribed diversity of microbial eukaryotes of the highest taxonomic order, including perhaps "true" early branching lineages.

Freshwater systems

The rRNA surveys tend to focus on extreme environments, and the first freshwater system studied was the highly acidic (pH ~ 2.5) Rio Tinto (Amaral-Zettler et al. 2002). A surprisingly high diversity of all major eukaryotic groups was found, with green and red algae, diatoms and *Euglena* spp. as photosynthesizes, and fungi, nucleariid-related members, cercozoans, heterotrophic stramenopiles and Amoebozoa as heterotrophic constituents. Berney et al. (2004) studied the eukaryotic diversity in the river Seymaz (Switzerland), finding sequences in their libraries mostly related to the fungi, ciliates, stramenopiles and Cercozoa, and to a lesser extent, apicomplexans, metazoa, green algae, Amoebozoa, and a lineage of unclear affiliation previously identified in deep-sea plankton and anoxic settings. Freshwater anoxic or suboxic running waters, such as those of a sulfur spring (Zodletone) revealed sequences of stramenopiles, fungi, alveolates, and, at lower frequency, Cercozoa, jakobids and diplonemids (Luo et al. 2005). Several studies surveyed standing waters of ponds and lakes. Slapeta et al. (2005) compared the microbial eukaryotic diversity in sediment and plankton of two different size range (>5 μm and <5 μm) from two—oxic and suboxic—ponds from the same geographic area. Ciliates dominated libraries from the suboxic system, and included members of the ciliate classes Oligohymenophorea, Prostomatea, Plagiopylea, Phyllopharyngea, Litostomatea and Spirotrichea, whereas planktonic cryptophytes dominated those from the oxic system. In the oligotrophic lake George (USA), rRNA surveys detected an abundance of stramenopiles, cryptomonads and alveolates. Interestingly, several rRNA gene sequence clusters appeared specific to freshwater environments, while others contained signatures of marine species indicating a wider distribution (Richards et al. 2005). A more recent study of the picoplankton of the meromictic lake Pavin (France) also suggested the presence of freshwater-specific clades, with alveolates, fungi and cercozoans being the most represented taxa in clone libraries (Lefèvre et al. 2007). Takishita et al. (2006) analyzed the anoxic sediment from a coastal meromictic lake in Japan, detecting phylotypes known from other anoxic environments, i.e. a variety of cercozoans and several sequences related to parasitic lineages. These data are still fragmentary, preventing a thorough comparison and ecological analyses. In this regard, noteworthy is the first comparative study of eukaryotic diversity in the picoplankton of three lakes (oligotrophic, oligomesotrophic and eutrophic) and showing differences between them as a function of their trophic status (Lefranc et al. 2005).

Soils

Despite the recognized richness of microbial soil communities, and in contrast to aquatic systems and their associated sediments, the diversity of microbial eukaryotes in soils has been rarely approached by molecular methods. This may be partly explained by the technical difficulty imposed by the overwhelming presence of fungal hyphae, and abundance of animals masking unicellular eukaryotes in clone libraries. In fact, molecular surveys targeting specifically fungal communities show that fungi are extremely diverse in soils, and particularly forest soils (Anderson and Cairney 2004; O'Brien et al. 2005). Nonetheless, several recent molecular analyses show the feasibility of group-specific primer approach also in studying the rest of the eukaryotic microbial community. For example, Moon-van der Staay and colleagues amplified the 18S rRNA genes from a wide variety of eukaryotic phyla from

agricultural soils, including fungi, cercozoans, green algae, ciliates, and a large variety of amoeboid protists including heteroloboseans and amoebozoans (Moon-van der Staay et al. 2006). In spite of this progress, soils remain among the most poorly studies habitat of unicellular eukaryotes.

The unidentified diversity of protists

If there is anything that the molecular approaches to microbial diversity showed with certainty is that the earlier notion of protozoologists having access to "… a full deck, or at least as full as we could" (Baldauf 2003) might have been rather naive. However, attaching specific numbers to this qualitative observation has proved challenging.

Estimates of the total number of protistan species are not based on statistical theory and are emphatically speculative. This is one reason why they vary so widely: while some argue that the global richness of free-living protozoa is less than 20,000 species (Finlay and Fenchel 1999), others maintain that a single phylum of ciliates comprises more than 30,000 species (Foissner 1999). There are multiple reasons behind this disagreement. One of them is the well known problem with species definition in microorganisms. This problem goes outside of the scope of this paper, but we note that it can be pragmatically resolved by considering rRNA gene sequence-based operational taxonomic units (OTUs). Of relevance here is the universal observation that the observed OTU frequency distribution is characterized by a large number of OTUs registered only once. Accumulation curves in molecular surveys are far from saturation (Fig. 3). This indicates that the complete inventory is a daunting task, and that the protistan richness even in relatively simple communities can be only estimated.

There are two principal groups of statistical approaches for estimating how many OTUs have been missed by a survey. Both have been used in macro- and prokaryotic biology, but it was not until 2004 that protozoologists began their adaptation to the needs of protistan biodiversity research. Nonparametric methods typically employ nonparametric coverage based estimators developed by Anne Chao, such as S_{Chao1}, ACE and ACE1 (Chao 1984, 2005). Massana et al. (2004a, b) employed S_{Chao1} and estimated the total eukaryotic (including metazoan) richness in four 10-l samples of surface waters in NW Mediterranean as seasonally varying between 72 and 171 OTUs per sample (OTUs defined as RFLP patterns with the 18SrRNA gene sequence similarity within phylotype ranging from 96.3% to 99.9%). Countway et al. (2005) used a similar approach and applied it to a collection of eight 4-l surface seawater samples from NW Atlantic and predicted the total eukaryotic (including metazoan) richness as 162–282 OTUs (defined as groups of sequences sharing over 95% 18S rRNA gene similarity). It is not clear, however, if the use of nonparametric estimators was optimal in these cases. These estimators are known to perform well under conditions of high coverage of the total richness by the empirical data (Chao and Bunge 2002). This condition is almost never met in microbial diversity studies (e.g., the number of OTUs registered only once is very high). Under these circumstances, the coverage-based estimators are likely to underestimate the total richness, and do so by an unknown factor. A notable exception may be a recent study of soil ciliate diversity by Chao et al. (2006). The painstakingly collected empirical dataset reported there provided an unusually high coverage of the total ciliate diversity, possibly justifying the application of nonparametric estimators.

The second group of methods employs parametric distribution to describe the frequency distribution of detected OTUs, and project the distribution so as to estimate how many OTUs must have been missed (Bunge and Fitzpatrick 1993). The application of these

methods in biodiversity (including microbial diversity) research typically does not include estimation of model parameters using maximum likelihood, goodness-of-fit assessment, or correct maximum likelihood standard errors (Hong et al. 2006). It is also common to assume that the frequency distribution of the detected OTUs follows the log-normal model, which has little justification in microbial diversity research (i.e., even if frequency distribution of species in nature is lognormal, the same is not necessarily true of that of PCR products, or clones in clone libraries). In light of that, we developed an empirical approach (Hong et al. 2006), informed by modern statistical theory, which makes no a priori assumptions on the nature of OTU frequency distribution. The approach is based on systematic application of several parametric models to the datasets, choosing the one that fits the data the best and gives a biologically meaningful standard error. Jeon et al. (2006) used this approach to estimate protistan richness in a single 3-l sample from anoxic waters of the Cariaco Basin off the coast of Venezuela, and predicted the total richness of the sample to be 398 ± 156 OTUs (defined as clusters sharing over 99% of 18S rRNA gene sequence similarity). Behnke et al. (2006) applied the same approach to the 18S rRNA gene survey data on stratified water column of a Norvegian fiord, and estimated the total protistan richness in habitats with different oxygen and sulfide regimes: 64 ± 15 OTUs in sulfide-free layer, 147 ± 46 OTUs at the upper sulfide boundary, and 27 ± 8 OTUs in the highly sulfidic layer.

The amount of information on the pool of species "missed" by standard clone libraries is thus very limited, and this prevents a thorough analysis of protistan richness at any spatial scale. As a result, we do not really know the total number of protistan species per sample, habitat, environment, etc. The few available statistical estimates, and qualitative picture of protistan diversity that transpires from the collection curves published, provide only a suggestive view of this diversity. A recent study (Stoeck et al. 2007a) compiled and re-analyzed via a collection of parametric and nonparametric approaches the data for marine sediments and marine anoxic environments (Dawson and Pace 2002; Edgcomb et al. 2002; López-García et al. 2003; Stoeck and Epstein 2003, Stoeck et al. 2003, 2006). An emerging pattern is that samples from more extreme environments (High Arctic and hydrothermal vents) contain more to many more phylotypes at the approximately the species level (OTUs formed at 97–99% rRNA gene sequence identity) than temperate zone tidal flats (hundreds vs dozens in one to several grams of sediment). Most of these phylotypes seem unique and rarely exhibit an exact match to organisms reported from elsewhere. If even small volume individual samples show degrees of uniqueness, the grand total of global protistan richness must be very large indeed. Presently, it would be imprudent to attach a number to this richness, but it does seem likely that genetic diversity exceeds projections based on morphological criteria—compare for example (Fenchel et al. 1990, 1995 vis a vis Zuendorf et al. 2006).

Additionally, recent observations show that even organisms that are essentially identical morphologically may be phylogenetically very different (Boenigk et al. 2005; Katz et al. 2005, 2006; Rodriguez et al. 2005; Slapeta et al. 2006a, b) and/or ecologically quite distinct (see for example Nanney et al. 1998; Foissner 1999; Coleman 2002; Lowe et al. 2005). The number of these so-called 'cryptic species' is completely unknown. Collectively, this makes it reasonable to hypothesize that pre-molecular inventories significantly underestimated global protistan richness. This might have produced a skewed picture of protists as a group of cosmopolitan forms of limited diversity ("everything is every where", Beijerinck 1913). The recent data lend a tentative support to the opposing view on protists as a remarkable world of diverse species with biogeographies (Baldauf 2003; Foissner 1999, 2006).

References

Amaral Zettler LA, Gomez F, Zettler E et al (2002) Microbiology: eukaryotic diversity in Spain's River of Fire. Nature 417:137

Anderson IC, Cairney JW (2004) Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. Environ Microbiol 6:769–779

Atkins MS, Teske AP, Anderson OR (2000) A survey of flagellate diversity at four deep-sea hydrothermal vents in the Eastern Pacific Ocean using structural and molecular approaches. J Eukaryot Microbiol 47:400–411

Baldauf SL (2003) The deep roots of eukaryotes. Science 300:1703–1706

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

Behnke A, Bunge J, Barger K et al (2006) Microeukaryote community patterns along an O2/H2S gradient in a supersulfidic anoxic fjord (Framvaren, Norway). Appl Environ Microbiol 72:3626–3636

Beijerinck MW (1913) De infusies en de ontdekking der backteriën. Müller, Amsterdam, the Netherlands

Berney C, Fahrni J, Pawlowski J (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. BMC Biol 2:13

Boenigk J, Pfandl K, Stadler P et al (2005) High diversity of the 'Spumella-like' flagellates: an investigation based on the SSU rRNA gene sequences of isolates from habitats located in six different geographic regions. Environ Microbiol 7:685–697

Brinkmann H, van der Giezen M, Zhou Y et al (2005) An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst Biol 54:743–757

Brocks JJ, Logan GA, Buick R et al (1999) Archaean molecular fossils and the early rise of eukaryotes. Science 285:1025–1027

Bunge J, Fitzpatrick M (1993) Estimating the number of species—a review. J Am Stat Assoc 88:364–373

Cavalier-Smith T (2004) Only six kingdoms of life. Proc R Soc Lond B Biol Sci 271:1251–1262

Chao A (1984) Non-parametric estimation of the number of classes in a population. Scand J Statist 11:265–270

Chao A (2005) Species richness estimation. In: Balakrishnan C, Read B, Vidakovic B (eds) Encyclopedia of statistical sciences. Wiley, New York, pp 7907–7916

Chao A, Bunge J (2002) Estimating the number of species in a stochastic abundance model. Biometrics 58:531–539

Chao A, Li PC, Agatha S et al (2006) A statistical approach to estimate soil ciliate diversity and distribution based on data from five continents. Oikos 114:479–493

Coleman AW (2002) Microbial eukaryote species. Science 297:337; author reply 337

Countway PD, Gast RJ, Savai P et al (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. J Eukaryot Microbiol 52:95–106

Courties C, Vaquer A, Troussellier M et al (1994) Smallest eukaryotic organisms. Nature 370:255

Dawson SC, Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. Proc Natl Acad Sci USA 99:8324–8329

Díez B, Pedrós-Alió C, 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

Dolven JK, Lindqvist C, Albert VA et al (2007) Molecular diversity of alveolates associated with Neritic North Atlantic radiolarians. Protist 158:65–76

Edgcomb VP, Kysela DT, Teske A et al (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. Proc Natl Acad Sci USA 99:7658–7662

Fenchel T, Finlay BJ (1995) Ecology and evolution in anoxic worlds. Oxford Univ Press, Oxford

Fenchel T, Kristensen LD, Rasmussen L (1990) Water column anoxia: vertical zonation of planctonic protozoa. Mar Ecol Prog Ser 62:1–10

Fenchel T, Bernard C, Esteban G et al (1995) Microbial diversity and activity in a Danish fjord with anoxic deep water. Ophelia 43:45–100

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

Finlay BJ, Fenchel T (1999) Divergent perspectives on protist species richness. Protist 150:229–233

Foissner W (1999) Protist diversity: estimates of the near-imponderable. Protist 150:363–368

Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool 45:111–136

Guillou L, Chretiennot-Dinet MJ, Medlin LK et al (1999) *Bolidomonas*: a new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). J Phycol 35:368–381

Guillou L, Eikrem W, Chretiennot-Dinet MJ et al (2004) Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. Protist 155:193–214

Gunderson JH, Goss SH, Coats DW (1999) The phylogenetic position of *Amoebophrya* sp. infecting *Gymnodinium sanguineum*. J Eukaryot Microbiol 46:194–197

Hong SH, Bunge J, Jeon SO, Epstein SS (2006) Predicting microbial species richness. Proc Natl Acad Sci USA 103:117–122

Janson S, Gisselson LA, Salomon PS et al (2000) Evidence for multiple species within the endoparasitic dinoflagellate *Amoebophrya ceratii* as based on 18S rRNA gene-sequence analysis. Parasitol Res 86:929–933

Jeon SO, Bunge J, Stoeck T et al (2006) Synthetic statistical approach reveals a high degree of richness of microbial eukaryotes in an anoxic water column. Appl Environ Microbiol 72:6578–6583

Johnson MD, Tengs T, Oldach DW et al (2004) Highly divergent SSU rRNA genes found in the marine ciliates *Myrionecta rubra* and *Mesodinium pulex*. Protist 155:347–359

Katz LA, McManus GB, Snoeyenbos-West OLO et al (2005) Reframing the 'Everything is everywhere' debate: evidence for high gene flow and diversity in ciliate morphospecies. Aquat Microbiol Ecol 41: 55–65

Katz LA, Snoeyenbos-West O, Doerder FP (2006) Patterns of protein evolution in *Tetrahymena thermophila*: implications for estimates of effective population size. Mol Biol Evol 23:608–614

Kolodziej K, Stoeck T (2007) Cellular identification of a novel uncultured marine stramenopile (MAST-12 Clade) small-subunit rRNA gene sequence from a norwegian estuary by use of fluorescence in situ hybridization-scanning electron microscopy. Appl Environ Microbiol 73:2718–2726

Kunitomo Y, Sarashina I, Iijima M et al (2006) Molecular phylogeny of acantharian and polycystine radiolarians based on ribosomal DNA sequences, and some comparisons with data from the fossil record. Eur J Protistol 42:143–153

Lefèvre E, Bardot C, Noël C et al (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

Lefranc M, Thenot A, Lepere C et al (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. Appl Environ Microbiol 71:5935–5942

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

López-García P, Rodríguez-Valera F, Moreira D (2002) Towards the monophyly of Haeckel's Radiolaria: 18S rRNA environmental data support the sisterhood of Polycystinea and Acantharea. Mol Biol Evol 19:118–121

López-García P, Philippe H, Gaill F et al (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental micro-colonizers at the Mid-Atlantic Ridge. Proc Natl Acad Sci USA 100:697–702

López-García P, Vereshchaka A, Moreira D (2007) Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. Environ Microbiol 9:546–554

Lovejoy C, Massana R, Pedros-Alio C (2006) Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. Appl Environ Microbiol 72:3085–3095

Lowe CD, Day A, Kemp SJ et al (2005) There are high levels of functional and genetic diversity in *Oxyrrhis marina*. J Eukaryot Microbiol 52:250–257

Luo Q, Krumholz LR, Najar FZ et al (2005) Diversity of the microeukaryotic community in sulfide-rich Zodletone Spring (Oklahoma). Appl Environ Microbiol 71:6175–6184

Maidak BL, Cole JR, Lilburn TG et al (2001) The RDP-II (Ribosomal database project). Nucleic Acids Res 29:173–174

Massana R, Guillou L, Diez B et al (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 et al (2004a) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. FEMS Microbiol Ecol 50:231–243

Massana R, Castresana J, Balague V et al (2004b) Phylogenetic and ecological analysis of novel marine stramenopiles. Appl Environ Microbiol 70:3528–3534

Massana R, Terrado R, Forn I et al (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. Environ Microbiol 8:1515–1522

Medlin LK, Metfies K, Mehl H et al (2006) Picoeukaryotic plankton diversity at the Helgoland time series site as assessed by three molecular methods. Microb Ecol 52:53–71

Moon-van der Staay A-Y, van der Staay GWM, Guillou L et al (2000) Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. Limnol Oceanogr 45:98–109

Moon-van der Staay SY, De Wachter R, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409:607–610
Moon-van der Staay SY, Tzeneva VA, van der Staay GW et al (2006) Eukaryotic diversity in historical soil samples. FEMS Microbiol Ecol 57:420–428
Moreira D, López-García P (2002) Molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol 10:31–38
Moreira D, López-García P (2003) Are hydrothermal vents oases for parasitic protists? Trends Parasitol 19:556–558
Moreira D, von der Heyden S, Bass D et al (2007) Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. Mol Phylogenet Evol: doi:10.1016/j.ympev.2006.11.001
Nanney DL, Park C, Preparata R et al (1998) Comparison of sequence differences in a variable 23S rRNA domain among sets of cryptic species of ciliated protozoa. J Eukaryot Microbiol 45:91–100
Nikolaev SI, Berney C, Fahrni JF et al (2004) The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA 101:8066–8071
O'Brien HE, Parrent JL, Jackson JA et al (2005) Fungal community analysis by large-scale sequencing of environmental samples. Appl Environ Microbiol 71:5544–5550
O'Brien EA, Koski LB, Zhang Y et al (2007) TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res 35:D445–D451
Okamoto N, Inouye I (2005) The katablepharids are a distant sister group of the Cryptophyta: a proposal for Katablepharidophyta divisio nova/Kathablepharida phylum novum based on SSU rDNA and beta-tubulin phylogeny. Protist 156:163–179
Olsen GJ, Lane DJ, Giovannoni SJ et al (1986) Microbial ecology and evolution: a ribosomal RNA approach. Annu Revue Microbiol 40:337–365
Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734–740
Philippe H, Lopez P, Brinkmann H et al (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc Lond B Biol Sci 267:1213–1221
Reysenbach AL, Cady SL (2001) Microbiology of ancient and modern hydrothermal systems. Trends Microbiol 9:79–86
Richards TA, Bass D (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. Curr Opin Microbiol 8:240–252
Richards TA, Vepritskiy AA, Gouliamova DE, Nierzwicki-Bauer SA (2005) The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. Environ Microbiol 7:1413–1425
Rodriguez F, Derelle E, Guillou L et al (2005) Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). Environ Microbiol 7:853–859
Savin MC, Martin JL, LeGresley M et al (2004) Plankton diversity in the Bay of Fundy as measured by morphological and molecular methods. Microb Ecol 48:51–65
Slapeta J, López-García P, Moreira D (2006a) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol Biol Evol 23:23–29
Slapeta J, López-García P, Moreira D (2006b) Present status of the molecular ecology of kathablepharids. Protist 157:7–11
Slapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci 272:2073–2081
Sogin ML, Gunderson JH, Elwood HJ et al (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia. Science 243:75–77
Stoeck T, Epstein S (2003) Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. Appl Environ Microbiol 69:2657–2663
Stoeck T, Taylor GT, Epstein SS (2003) Novel eukaryotes from the permanently anoxic Cariaco Basin (Caribbean Sea). Appl Environ Microbiol 69:5656–5663
Stoeck T, Hayward B, Taylor GT et al (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. Protist 157:31–43
Stoeck T, Kasper J, Bunge J, Leslin C, Ilyin V, Epstein S (2007a) Protistan diversity in the arctic: a case of paleoclimate shaping modern biodiversity? PLoS ONE 2:e728
Stoeck T, Zuendorf A, Breiner HW, Behnke A (2007b) A molecular approach to identify active microbes in environmental eukaryote clone libraries. Microb Ecol 53:328–339
Takishita K, Miyake H, Kawato M et al (2005) Genetic diversity of microbial eukaryotes in anoxic sediment around fumaroles on a submarine caldera floor based on the small-subunit rDNA phylogeny. Extremophiles 9:185–196

Takishita K, Tsuchiya M, Kawato M et al (2006) Genetic diversity of microbial eukaryotes in anoxic sediment of the saline meromictic lake Namako-Ike (Japan): on the detection of anaerobic or anoxic-tolerant lineages of eukaryotes. Protist 2006 Sep 1; [Epub ahead of print]

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Worden AZ (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquat Microbiol Ecol 43:165–175

Worden AZ, Nolan JK, Palenik B (2004) Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. Limnol Oceanogr 49:168–179

Yuan J, Chen MY, Shao P et al (2004) Genetic diversity of small eukaryotes from the coastal waters of Nansha Islands in China. FEMS Microbiol Lett 240:163–170

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