The protistan gap in the eukaryotic tree of life[1]

Thorsten Stoeck & Alexandra Stock

Protists are unicellular eukaryotes that have conquered almost any habitat on our planet. The vast majority thrives in the oceanic realm, where they together with bacteria and archaea account for the majority of the biomass: The entire microbial food web, including protozoan microzooplankton, is typically five to ten times the mass of all multicellular marine organisms (Pomeroy et al. 2007). Also the ecological importance of protists becomes obvious when we consider their key position in pelagic food webs. As emphasized earlier, the pelagic food-web is actually microbe centric whereas the 'diatom-copepod-fish' food web is a relatively minor component (Barber 2007). Autotrophic protistan plankton belongs to major light harvesters driving primary productivity in open oceans (Liu et al. 2009): (i) by grazing phagotrophic protists control the dynamics of bacteria and archaea; (ii) by selective feeding on functional bacterial and archaeal taxa indirectly influence global biogeochemical cycles. Furthermore, protists are of critical importance to sustain higher life in the oceans by transferring energy and matter to higher trophic levels. These examples demonstrate that an adequate function and functioning of the oceans is strongly influenced by protistan biodiversity. In order to maintain a functioning ecosystem, a better knowledge and understanding of this biodiversity is essential.

Protists are probably the most abundant and diverse eukaryotes on Earth (Patterson 1999). However, even though they constitute the majority of phylogenetic groups in the eukaryotic tree of life (Fig. 1), our current knowledge of protistan diversity is decidedly incomplete, mainly due to methodological reasons.

Traditionally, protistan diversity research relies on cultivation or manual single cell collection and subsequent microscopy analyses. However, many, if not most,

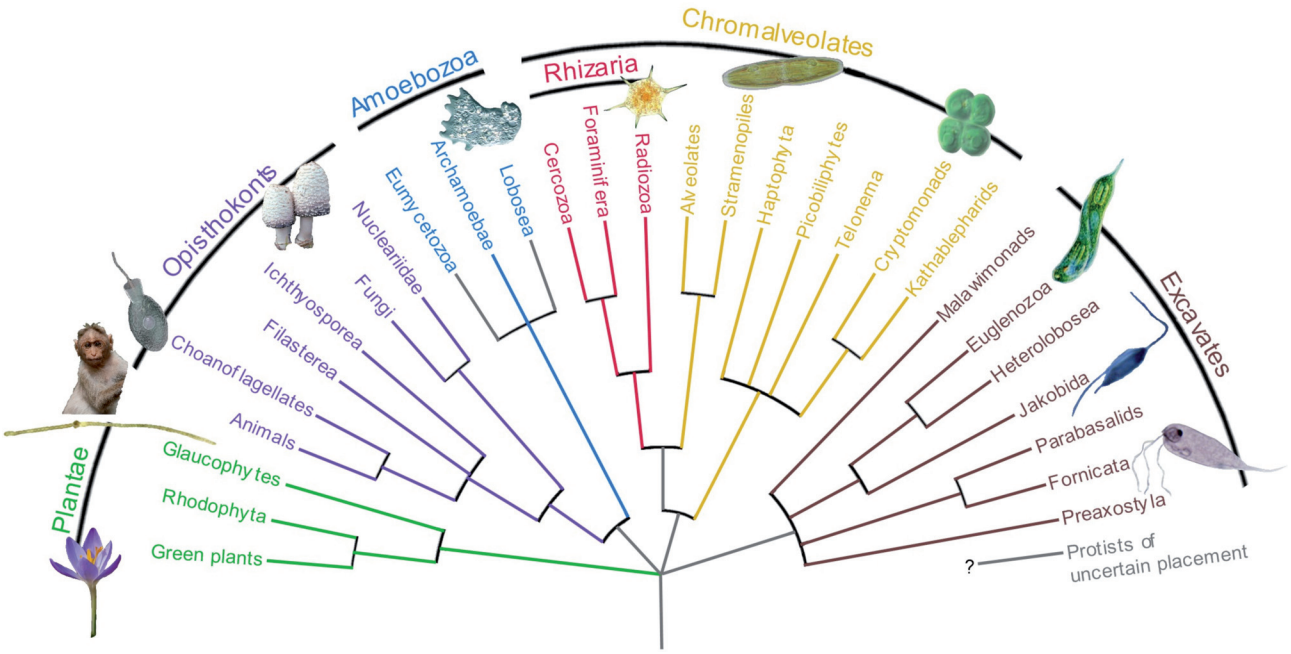

Fig. 1. Eukaryotic tree of life. Supergroups according to http://www.tolweb.org, modified. Unicellular organisms are represented in all major evolutionary lineages with the exception of plants and animals. Most lineages are composed entirely of protists.

[1] Contribution to the Willi-Hennig-Symposium on Phylogenetics and Evolution, University of Hohenheim, 29 September – 2 October 2009.

protists belong to the picoplankton size class ranging in size from 2 to 5 µm. Such small cells are easily overlooked in natural samples and often times have only few morphological characters that allow taxonomic identification. Numerous of these specifically small protists have body-doubles, so called "cryptic species" (Bickford et al. 2006; Hausmann et al. 2006; Stoeck et al. 2008). As evolution is not necessarily accompanied by morphological alteration, many different evolutionary lineages or sexually isolated taxon groups are morphologically indistinguishable. Cultivation of protists is often very difficult, because we simply do not know the growth requirements of most free-living microbes. Similarly unfruitful is direct accessibility of individual cells. The abundances of specific taxa may be as low as one cell per liter of seawater, which makes some taxa difficult to find and restricts manual collections to the most abundant morphotypes. Furthermore, protists from extreme environments, that are highly adapted to their natural habitat, hardly survive manual collection and microscopy.

A first promising tool to overcome these difficulties was a cultivation-independent molecular strategy (for a review see Epstein & López-García 2008). Originally developed to bypass the inability to cultivate most bacteria and archaea and to reconstruct evolutionary histories of microbes, this approach amplifies taxonomic marker genes (predominantly the small subunit of the ribosomal RNA gene, SSU rRNA) from genomic DNA (or RNA) that was extracted from environmental samples. Amplified marker genes are then cloned into plasmids, in which they are multiplied using bacteria as incubators for subsequent Sanger sequencing. The obtained sequences are then placed into a phylogenetic context by constructing evolutionary trees using a variety of algorithms and evolutionary models. This strategy changed our view of protistan diversity and revealed that what we knew thus far, was only the tip of the iceberg of a yet undiscovered richness of protists (Epstein & López-García 2008). This approach detected formerly unknown groups of organisms up to the highest taxonomic levels in the eukaryotic tree of life (Dawson & Pace 2002) and uncovered an unexpectedly high richness of protists in even the most extreme habitats surveyed to date (Epstein & López-García 2008). As a matter of fact, protistan communities emerged as so rich in phylotypes that even the largest molecular clone library surveys were not able to sample these communities entirely to saturation (Stoeck et al. 2007). Thus, even though this molecular strategy evolved to the gold-standard in microbial diversity research complementing traditional microscopy-based approaches, it certainly is no panacea because a number of central questions in ecology and evolution still remain unresolved. Such questions include for example:

– What is the true richness of protists in natural habitats?
– What is the cellular identity of organisms behind novel sequences?
– What is the abundance and geographic distribution of these organisms?
– What are their physiological capabilities?
– What are their ecological roles in an ecosystem?

Following we will succinctly discuss some recent methodological developments that may be helpful to address these questions.

In order to reveal the organism behind an environmental sequence, one can take advantage of the primary structure of the respective sequence. A number of different tools like, for example, implemented in the ARB software package (Ludwig et al. 2004) compares the query SSU rDNA sequence (or a group of query sequences) to a large database of SSU rDNA sequences in order to identify a short nucleotide region, usually ranging between 18 and 24 bases, that is unique to the query sequence. An oligonucleotide (probe) consisting of this stretch of bases with a fluorochrome attached to its 5'-end is then hybridized to fixed cells from an environmental sample, usually drawn on a membrane. In presence of the organism of desire, the

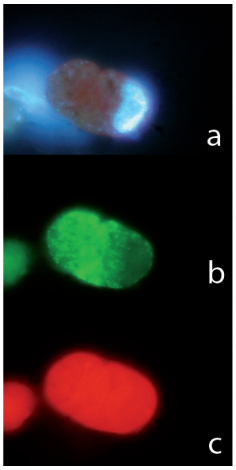

Fig. 2. Epifluorescence microscopy pictures from a fluorescence in situ hybridization (FISH) of the dinoflagellate *Gymnodinium* sp. with a universal eukaryote oligonucleotide probe labeled with the fluorochrome FITC (green). – **a**. Counterstaining with the fluorochrome DAPI, staining DNA structures in the nucleus of the target cell. DAPI can be visualized at excitation of 365 nm. At this wavelength the chl*a* autofluorescence showing in red in the cytoplasm. **b**. Excitation with a FITC-specific wavelength (450–490 nm) indicates a positive FISH signal in the target cell showing in bright green. **c**. Chl*a* autofluorescence: excitation of a pigmented organism with a wavelength of 546 nm activates electrons of pigments (chl*a*) emitting their energy in the reddish spectrum (ca. 650–750 nm).

probe anneals to its complementary target in this organism resulting in a fluorescence signal under excitation. Subsequently, epifluorescence microscopy will detect this signal (Fig. 2). Using this so called Fluorescence In Situ Hybridization (FISH, Pernthaler et al. 2001) strategy Massana and colleagues visualized a novel stramenopile organism that was formerly known exclusively from environmental sequencing (Massana et al. 2002). At the same time, epifluorescence microscopy allows to distinguish pigmented from non-pigmented protists, because chlorophyll *a* (chl*a*) displays a strong autofluorescence signal under UV-excitation (Fig. 2).

Taking advantage of this chl*a* autofluorescence, it is possible to assign a trophic mode to a previously unknown and uncultivated protist (Massana et al. 2002). Stoeck et al. (2003) developed this approach a step further, namely by preparing an environmental sample in a way that allows subsequent scanning electron microscopy of the same individual cell that was previously identified using FISH. This strategy allowed the morphological description of a novel MAST-organism (uncultured marine stramenopile) from an intertidal microbial mat (Kolodziej & Stoeck 2007). Furthermore, FISH enables to survey the abundance, and spatial and temporal dynamics of a specific target organism by screening a large number of samples in a relatively short period of time. For example, FISH revealed the abundance and distribution of heterotrophic flagellates (uncultured marine stramenopiles) in a variety of oceanic provinces (Massana et al. 2006). In another application, a specific FISH probe may be helpful in finding the appropriate enrichment strategy of a natural sample in order to stimulate growth of the targeted organism in this sample. Such a targeted enrichment may ultimately lead to the isolation and pure culturing of organisms that were previously known exclusively by their SSU rDNA sequences, enabling detailed ultrastructural descriptions and physiological profilings. As an example, Stoeck et al. (2005) succeeded to characterize a novel kinetoplastid *(Actuariola framvarensis)* from oxygen-depleted waters of a Norwegian Fjord.

Natural microbial communities harbor many more organisms than can actually be detected at reasonable efforts and costs. In order to assess the extent of protistan diversity in a specific sample, statistical techniques are available for extrapolating the total number of taxa (phylotypes, genotpyes, operational taxonomic units) from a finite sample. Generally, these techniques apply non-parametric estimators like ACE or Chao (Chao & Shen 2003–2005) or parametric models like inverse Gaussian or mixed-Pareto (Hong et al. 2006). Because non-parametric estimators only perform well, when a large proportion of a community's richness has been detected, parametric models are more appropriate for estimating taxon richness based on clone library data (Chao et al. 2006; Jeon et al. 2006). However, this argument is largely ignored by most microbial ecologists and due to its ease of use, non-parametric estimators are most commonly employed. A more direct way to access the unseen majority of taxa in natural microbial communities are high throughput sequencing strategies like massively parallel tag sequencing (454 sequencing, pyrosequencing). This strategy allows the analysis of millions of short sequence reads (current sequencing platforms read up to ca. 500 bp) in a reasonable amount of time and at a fraction of the costs of the traditional cloning and Sanger sequencing strategy. For the first time, this approach was applied in 2006 for the diversity and richness analyses of bacterial communities in oceanic systems (Sogin et al. 2006). Recently, this strategy has been adopted for microbial eukaryotes (Stoeck et al. 2009, 2010). Even though pyrosequencing detected many more organisms and taxon groups in an individual sample compared to the clone library approach, a number of (presumably low-abundance) populations still remained undetected even after sequencing more than 400,000 tags in an individual sample. Thus, a combination of statistical tools with high-throughput sequencing data currently seems to be the most efficient combination to assess protistan taxon richness in natural samples.

The examples mentioned above demonstrate that the technologies are available to close the protistan gap in the eukaryotic tree of life and to address the ecological and evolutionary significance of microbial biodiversity. The resurgence of the "organismic discovery era" makes it a very exciting time to be a microbial ecologist.

References

Barber, R. T. A. (2007): OCEANS: Picoplankton do some heavy lifting. – Science, **315**: 777–778.

Bickford, D., Lohman, D. J., Sohdi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. & Das, I. (2006): Cryptic species as a window on diversity and conservation. – Trends in Ecology and Evolution, **22** (3): 148–155.

Chao, A. & Shen, T. J. (2003–2005): Program SPADE (Species Prediction and Diversity Estimates). – In: http://chao.stat.nthu.edu.tw.

Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T. J. (2006): Abundance-based similarity indices and their estimation when there are unseen species in samples. – Biometrics, **62**: 361–371.

Dawson, S. C. & Pace, N. R. (2002): Novel kingdom-level eukaryotic diversity in anoxic environments. – Proceedings of the National Academy of Sciences of the United States of America, **99**: 8324–8329.

Epstein, S. & López-García, P. (2008): "Missing" protists: a molecular prospective. – Biodiversity and Conservation, **17** (2): 261–276.

Hausmann, K., Selchow, P., Scheckenbach, F., Weitere, M. & Arndt, H. (2006): Cryptic species in a morphospecies complex of heterotrophic flagellates: the case study of *Caecitellus* spp. – Acta Protozoologica, **45**: 415–431.

Hong, S. H., Bunge, J., Jeon, S. O. & Epstein, S. S. (2006): Predicting microbial species richness. – Proceedings of the National Academy of Sciences of the United States of America, **103**: 117–122.

Jeon, S. O., Bunge, J., Stoeck, T., Barger, K., Hong, S.-H. & Epstein, S. (2006): Synthetic statistical approach reveals a high degree of richness of microbial eukaryotes in an anoxic water column. – Applied and Environmental Microbiology, **72**: 6578–6583.

Kolodziej, K. & Stoeck, T. (2007): Cellular identity of a novel uncultured MAST-12 lineage and phylogeny of the uncultured marine stramenopile sequence clade MAST-12. – Applied and Environmental Microbiology, **73**: 2718–2726.

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F. & de Vargas, C. (2009): Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. – Proceedings of the National Academy of Sciences of the United States of America, **106**: 12803–12808.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüssmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. & Schleifer, K. H. (2004): ARB: a software environment for sequence data. – Nucleic Acids Research, **32**: 1363–1371.

Massana, R., Guillou, L., Diez, B. & Pedrós-Alío, C. (2002): Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. – Applied and Environmental Microbiology, **68**: 4554–4558.

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

Patterson, D. J. (1999): The Diversity of Eukaryotes. – The American Naturalist, **154**: S96–S124.

Pernthaler, J., Glöckner, F. O., Schönhuber, W. & Amann, R. (2001): Fluorescence *in situ* hybridization (FISH) with rRNA-targeted oligonucleotide probes. – In: Paul, J. H. (ed.): Methods in Microbiology: 207–226; San Diego (Academic Press).

Pomeroy, L. A., Williams, P. J. L., Azam, F. & Hobbie, J. E. (2007): The microbial loop. – Oceanography, **20**: 28–33.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G: J. (2006): Microbial diversity in the deep sea and the underexplored "rare biosphere". – Proceedings of the National Academy of Sciences of the United States of America, **103**: 12115–12120.

Stoeck, T., Fowle, W. H. & Epstein, S. S. (2003): Methodology of protistan discovery: from rRNA detection to quality scanning electron microscope images. – Applied and Environmental Microbiology, **69**: 6856–6863.

Stoeck, T., Jost, S. & Boenigk, J. (2008): Multigene phylogeny of clonal Spumella-like strains, a cryptic heterotrophic nanoflagellate isolated from different geographic regions. – International Journal of Systematics and Evolutionary Microbiology, **58**: 716–724.

Stoeck, T., Schwarz, M. V. J., Boenigk, J., Schweikert, M., von der Heyden, S. & Behnke, A. (2005): Cellular identity of an 18S rRNA gene sequence clade within the class Kinetoplastea: the novel genus *Actuariola* gen. nov. (Neobodonida) with description of the type species *Actuariola framvarensis* sp. nov. – International Journal of Systematics and Evolutionary Microbiology, **55**: 2623–2635.

Stoeck, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V. & Epstein, S. (2007): Protistan diversity in the arctic: a case of paleoclimate shaping modern biodiversity? – PLoS One, **2** (8): e728.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D., Breiner, H. W. & Richards, T. A. (2010): Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. – Molecular Ecology, **19**: 21–31.

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M. J., Chistoserdov, A., Orsi, W. & Edgcomb, V. P. (2009): Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. – BMC Biology, **7**: 72.

Address of the authors:

Thorsten Stoeck, Alexandra Stock, Department of Ecology, University of Kaiserslautern, Erwin-Schrödinger Str. 14/143, 67663 Kaiserslautern, Germany
E-mail: stoeck@rhrk.uni-kl.de
astock@rhrk.uni-kl.de

Received and accepted as extended summary of oral presentation: 15 April 2010.