Vol. 81: 83–94, 2018
https://doi.org/10.3354/ame01864

AQUATIC MICROBIAL ECOLOGY
Aquat Microb Ecol

Published online February 26

Improved 18S rDNA amplification protocol for assessing protist diversity in oxygen-deficient marine systems

Claudia Wylezich[1,2,*], Daniel P. R. Herlemann[1], Klaus Jürgens[1]

[1]Leibniz Institute for Baltic Sea Research Warnemünde, Biological Oceanography, Seestraße 15, 18119 Rostock, Germany

[2]*Present address:* Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Diagnostic Virology, Südufer 10, 17493 Greifswald - Insel Riems, Germany

ABSTRACT: High-throughput sequencing techniques have been increasingly used in biodiversity estimates and in the detection of rare species within prokaryotic and eukaryotic assemblages in different habitats. In studies of protists, the V4 region of the 18S rRNA gene is the genetic marker most often used. However, established primers for the variable region V4 show several mismatches when used to examine protist groups important in oxygen-deficient aquatic systems, such as euglenozoans, some stramenopiles, and ciliates (e.g. mesodiniids). In this study, we designed new general primers covering the V4 and V5 region and used them in combination with a primer mix in a 2-step PCR approach to improve our understanding of eukaryotic diversity in oxygen-deficient zones. Prior to amplification of environmental samples from oxygen-deficient water layers of the Black Sea, we tested and improved the protocol on a protist mock community (1 ciliate and 5 flagellate species). With the developed approach, we detected protist taxa of a broad taxonomic range such as several ciliate groups including the hitherto missing mesodiniids, as well as of dinoflagellates, apicomplexans, apusomonads, different stramenopiles (bicosoecids, diatoms, chrysophytes, several marine stramenopile clades, bolidophytes), and haptophytes, excavates (jakobids, different otherwise missing euglenozoan groups, and schizopyrenids), rhizarians, amoebozoans, and choanoflagellates in the Black Sea sample. Therefore, we conclude that the new primers and tested amplification protocol improve the assessment of protist diversity from oxygen-deficient waters including phylogenetic groups that could not be detected by most published primers.

KEY WORDS: Protists · Diversity · High-throughput sequencing · Mock community · Black Sea · Oxygen · Anoxic · Singleton

INTRODUCTION

The advent of next-generation high-throughput sequencing (HTS) methods, such as 454 pyrosequencing, Ion Torrent and Illumina sequencing, have initiated a new era of diversity studies using cultivation-independent methods. This approach has allowed in-depth and cost-effective studies of the environmental diversity of both dominant and rare species. The large number of short gene sequence fragments, generated in a very short time, has rapidly expanded prokaryotic and eukaryotic sequence databases. The best-studied target gene of HTS methods is the small subunit ribosomal RNA gene (SSU rRNA gene), specifically, the 16S rRNA gene in prokaryotes and the 18S rRNA gene in most eukaryotes. For protists (eukaryotic microorganisms), several variable regions of the 18S rRNA gene were evaluated in pioneering studies of different marine and freshwater systems, including the V3 (Medinger et al. 2010, Nolte et al.

*Corresponding author: claudia.wylezich@fli.de

2010), V4 (e.g. Bråte et al. 2010, Cheung et al. 2010, Stoeck et al. 2010, Engel et al. 2012, Bachy et al. 2013), and V9 regions (e.g. Amaral-Zettler et al. 2009, Stoeck et al. 2009, Edgcomb et al. 2011, Pawlowski et al. 2011). Some of these studies also provided the first HTS evaluations of protist assemblages in suboxic and anoxic marine habitats (Stoeck et al. 2009, 2010, Edgcomb et al. 2011). Here, good agreement regarding the taxon composition within different taxonomic hierarchies was found between Sanger-sequenced clone libraries and HTS studies (Edgcomb et al. 2011), and a high amount of presumably low-abundance operational taxonomic units (OTUs) indicated the suitability of HTS approaches to detect presumably rare species (Stoeck et al. 2009). This was confirmed by the study of Medinger et al. (2010), in which single-cell PCR was compared with HTS.

The V4 region of the 18S rRNA gene has since been confirmed as the region best suited for HTS investigations of microbial eukaryotes (Dunthorn et al. 2012), and it is now the marker most frequently used for this purpose. The combination of the V4 with the V5 region provides the most detailed phylogenetic information of the 18S rRNA gene (Hugerth et al. 2014). However, the majority of known V4 primers discriminate against some protist groups, such as the aberrant foraminifers as well as excavates and some ciliate groups (Pawlowski et al. 2011). For instance, the V4 primers of Bråte et al. (2010) are unable to detect euglenids, whereas those used by Stoeck et al. (2010) and Cheung et al. (2010) are not well adapted to amplify jakobids, some euglenids, and ciliates. However, these often disregarded protist groups belong to important and presumably dominant taxa in suboxic and anoxic-sulfidic marine pelagic habitats (e.g. Stoeck et al. 2009, Edgcomb et al. 2011, Wylezich & Jürgens 2011). The mesodiniids for example are discriminated by the above-mentioned V4 primers, as assessed *in silico* (Stoeck et al. 2014), although these organisms are often present in high abundance, based on cell counts, in oxygen-depleted habitats (Anderson et al. 2012, Weber et al. 2014). In order to improve the recovery of important protist groups from marine oxygen-deficient water columns, we developed a new V4/V5 primer system and tested it for DNA and cDNA (expressed rRNA genes) templates derived from an artificial protist community and 2 environmental samples from the Black Sea's oxygen-deficient redox zone, which were previously Sanger sequenced (Wylezich & Jürgens 2011). The protist mock community was used to adjust the amplification conditions of the new primers, and to evaluate the OTU clustering threshold, whereas the Black Sea is a perfect case study for oxygen-deficient marine water columns for which protistan diversity has only poorly been resolved and where primers, which cover relevant groups of those systems, are still needed.

MATERIALS AND METHODS

Preparing a culture-based model community

For a protist mock community of known composition and cell abundance, 6 cultures from the Leibniz Institute for Baltic Sea Research Warnemünde (IOW) culture collection were used: 1 choanoflagellate (*Codosiga balthica* IOW94), 2 stramenopiles (*Cafeteria roenbergensis* IOW16, *Spumella* sp. IOW81), 1 kinetoplastid (*Bodo saltans* IOW92), 1 ciliate (*Cyclidium* sp. IOW89), and 1 dinoflagellate (*Heterocapsa triquetra*). The cultures originated from isolations of plankton samples from different depths around the oxic–anoxic interface (the so-called redoxcline) of the central Baltic Sea (Gotland Deep). Exceptions were *C. roenbergensis*, obtained from a coastal plankton sample (near Heiligendamm, Germany), and *H. triquetra*, isolated from the North Sea near the island of Sylt, Germany. The cultures were maintained in F/2 medium (Guillard & Ryther 1962) to which a wheat grain was added as a carbon source to sustain bacterial growth. Alternating dark/light conditions were provided for *Heterocapsa*. To determine culture densities, 1 ml of cells was fixed with glutaraldehyde (final concentration 4%). After 1 h incubation at 4°C, the cells were filtered onto black polycarbonate filters (0.8 µm pore-size; 25 mm diameter; Whatman), stained with DAPI (0.01 mg ml^{-1}), mounted, and enumerated under a Zeiss Axioskop 2 mot plus epifluorescence microscope (Carl Zeiss MicroImaging). Approximately 1000 cells $filter^{-1}$ were counted at 630× using filter set 02 (Carl Zeiss MicroImaging), except in the low-abundance *Heterocapsa* samples, in which approximately 100 cells were counted. Each culture was then filtered (~8 h post-fixation) onto a Durapore filter (Millipore, 0.2 µm pore size) such that the final cell concentration was ~3×10^5 cells $filter^{-1}$, except *Heterocapsa* (3×10^4). The filter was immediately frozen in liquid nitrogen and stored at −80°C until nucleic acid extraction.

Sampling in the Black Sea

Water samples from the redoxcline, a broad transitional zone between oxic and anoxic/sulfidic water

layers, of the southern Black Sea (Stn 23; 41° 36′ N, 37° 26′ E) were taken on board the RV 'Meteor' in May 2007 (M72/5) as described by Wylezich & Jürgens (2011). The samples, collected from 2 depths (suboxic and anoxic-sulfidic conditions at 130 and 155 m, respectively), were investigated previously using 18S rRNA gene clone libraries and using a different set of general eukaryotic primers (Wylezich & Jürgens 2011).

Preparation of nucleic acids

All samples were analyzed for both genomic DNA and RNA. Nucleic acids were extracted using a protocol combining manual cell disintegration and phenol chloroform extraction, as described in detail by Wylezich & Jürgens (2011). The resulting nucleic acid solution was used for DNA-based analyses. For RNA-based studies, co-precipitated DNA was removed by digesting an aliquot of the extracted nucleic acids (40 µl) with DNase I (Ambion) for 1 h at 37°C. The remaining RNA was quantified using a NanoDrop ND-100 spectrometer (NanoDrop Technologies). Polymerase chain reactions (PCRs) were then performed using the general eukaryotic primers EukA (Medlin et al. 1988) and Euk516rGC (Díez et al. 2001) to confirm that the RNA was free of residual DNA. The confirmed RNA was reverse-transcribed into cDNA using the iScript Select DNA synthesis kit (Bio-Rad) according to the manufacturer's instructions and both the random primer mix and the gene-specific primer EukB (Medlin et al. 1988).

Primer design, preliminary amplification, and sequencing tests

We modified published V4 reverse primers to make them more general and to cover important protist groups known to occur in oxygen-deficient marine waters (Wylezich & Jürgens 2011). The published reverse primers TAReukREV3 and V4_euk_R2 were modified and resulted in the new primers E981R and E1007R, respectively (see Table 1 for all primers and Fig. S1 for binding sites, available in the Supplement at www.int-res.com/articles/suppl/a081p083_supp.pdf). Both were then used in combination with the newly designed forward primer (E561F) to analyze the mock community and the 2 environmental Black Sea samples, each with DNA and cDNA templates (6 samples in total). Using PCR, the primers (1) did not result in any bands or (2) resulted in many bands or (3) in smears but not in 1 (or 2) clear band. PCR optimization steps, such as gradients of $MgCl_2$, temperature, DMSO, or BSA, and the use of touch-down protocols, did not improve the results. We then additionally designed a new general reverse primer (E1164R) covering the V4 and V5 region and used it in combination with the forward primer E561F. *In silico* analyses of the primer pair E561F and E1164R revealed highest coverage for protist diversity (81%) in contrast to V4 primer combinations. After sequencing of the excised amplicon at the expected size, sequencing resulted in a huge amount of non-target (prokaryotic) reads (Fig. S2). Therefore, we developed the 2-step PCR protocol with eukaryotic pre-amplification and following E561F-E1164R amplifi-

Table 1. Sequences of the primers used in this study. Numbers in the names of the V4/V5 amplification primers indicate the sequence position with respect to *Saccharomyces cerevisiae* accession number AY251629. Compare Fig. S1 in the Supplement at www.int-res.com/articles/suppl/a081p083_supp.pdf for binding sites

Primer name	Primer sequence	Reference
Primers used for pre-amplification (16 cycles)		
25F	CATATGCTTGTCTCAAAGATTAAGCCA	Bass & Cavalier-Smith (2004)
EukA	AACCTGGTTGATCCTGCCAGT	Medlin et al. (1988)
18S-For-n2	GATCCTGCCAGTAGTCATAYGC	Wylezich et al. (2012)
kineto14F	CTGCCAGTAGTCATATATGCTTGTTTCAAGGA	von der Heyden & Cavalier-Smith (2005)
18S-1630Rev	CGACGGGCGGTGTGTACAA	Wylezich & Jürgens (2011)
Primers used for V4/V5-amplification (25 cycles)		
E561F	GGTGCCAGCASCCGCGGTAA	This study
E981R[a]	ACTYTSGTYCTTGATYR	This study
E1007R[b]	GGTVTCTRATCNTCTTYGA	This study
E1164R	GTCAATTYCTTTAAGTTTCAG	This study

[a]Modified TAReukREV3 (Stoeck et al. 2010)
[b]Modified V4_euk_R2 (Bråte et al. 2010)

cation in order to reduce the number of prokaryotic reads. The 2-step approach was the most successful, and several cycle numbers of pre-amplification (10, 13, 16, 18, and 20) were combined with 25 cycles of E561F-E1164R amplification in preliminary tests. Additionally, different amounts of template (1 or 5 µl of undiluted or 1 µl of 1:10 diluted PCR product from the pre-amplification step) were tested. Further, a prokaryotic control was included in the PCR in order to check for amplification of 16S rRNA genes (see Fig. 2). Blank controls from the pre-amplification were also re-amplified for all PCR reactions and produced no amplicons.

The optimized PCR combination consisted of 16 cycles of eukaryotic pre-amplification (20 µl reaction volume) followed by 25 cycles of V4/V5-amplification (50 µl final reaction volume, including 0.5 µl DMSO). For pre-amplification (pre-PCR), the eukaryote-specific reverse primer 18S-1630Rev (0.12 mM final volume) was combined with a mixture of the forward primers 25F, EukA, 18S-For-n2, and kineto14F (0.03 mM each, for primer information, see Table 1). The reaction conditions were annealing at 56°C for 45 s and elongation at 70°C for 2 min. Re-amplification (re-PCR) was then carried out using the primers E561F and E1164R (54°C annealing and 72°C elongation, 15 s each). One µl of the pre-amplification product (5 µl for the cDNA of the sulfidic Black Sea sample) was used in the re-PCR, which contained the V4/V5 primers tagged with 454 adaptors, a sample-specific multiplex identifier (MID). For all PCR reactions, Herculase II Fusion DNA polymerase (Agilent Technologies) was used. The PCR products were gel purified, visualized, and excised from the agarose gel on a blue light table (wavelength 460–490 nm; Flu-O-Blu, Biozym), and extracted using the NucleoSpin gel and PCR clean-up kit (Machery-Nagel). The DNA concentrations of 6 samples were determined in a Picogreen assay (Molecular Probes), and equimolar amounts of the V4/V5 fragments were pooled for each sample. Two µg of the amplicon mix was sent to a next-generation sequencing service (LGC Genomics) and sequenced from the reverse primer direction using Roche/454 GS FLX Titanium technology.

Sequence data processing

Quality checks were done with the obtained sequence reads using the program Acacia (Bragg et al. 2012). With this program, we checked the reads for the reverse primer sequence, and de-multiplexed them to their sample origin based on the sample-specific ID. Sequences that contained Ns and with a quality score <25 were removed. Homopolymers were also corrected using Acacia (denoising program). The software package JAguc (Nebel et al. 2011) was used to remove singletons across the whole data set (later partly re-included, see below) after which we clustered those sequences with >200 bp and average similarities into OTUs using different cut-offs (90, 95, 97, 98, 99%) based on the unweighted pair group method with arithmetic mean (UPGMA). The mock community data sets (DNA and cDNA) were first clustered into OTUs to find an appropriate data-base-adapted OTU definition. Based on the results of this data set, we then applied a very conservative cut-off of 95% for the environmental Black Sea data for all following steps. A preliminary taxonomic affiliation was assigned using the BLAST algorithm as implemented in the JAguc package. The clustered and blasted data sets were finally checked for chimeras using keyDNAtools (Guillou et al. 2008). Sequence reads that matched with metazoans, plants, or prokaryotes were excluded. A representative sequence of each obtained $OTU_{95\%}$ and all the initially rejected singletons were manually checked again by BLAST (Altschul et al. 1997) using the NCBI and PR^2 databases (Guillou et al. 2013). Singletons that represented an already known sequence (with >95% sequence similarity to sequences from GenBank) were re-included in the data set. Additionally, the raw data set was treated using the Silva next-generation sequencing pipeline (https://www.arb-silva.de/ngs/) with a 95% cut-off and the exclusion of all sequences <200 nucleotides in length. For the purpose of comparison, the Shannon-Wiener diversity index, previously calculated for the clone libraries constructed for the 2 Black Sea samples (Wylezich & Jürgens 2011), was also computed for the 454 data sets (DNA and cDNA). Sequences of Black Sea OTUs have been deposited in the NCBI GenBank under accession numbers KY990054–KY990102.

RESULTS

Final V4/V5 amplification using new primers

A new primer system was designed using very conservative sequence regions, which resulted in a primer system (E561F and E1164R) that covers both the V4 and the V5 region. The expected amplicon size, as detected *in silico*, is ~600 base pairs (bp) for

most eukaryotes (including stramenopiles, alveolates, jakobids, and cercozoans), 750–800 bp for euglenids, and ~820 bp for bodonids. For amoebozoans, several fragment lengths are possible. The overlap with bacterial and archaeal sequences was minor (<10% of the reference database was covered with one of the primers). The primers were tested using TestPrime (Klindworth et al. 2013) as implemented in the web-based Silva database (Silva 128). The newly designed primers E561F and E1164R have an expected coverage of 83% for the 18S rRNA gene sequences contained in the Silva 128 database (ambiguities were taken into account, Fig. 1). For comparison, the primer pair consisting of TAReuk-454FWD1 and TAReukRev3 (Stoeck et al. 2010) and the modified Primer E981R combined with the forward primer E561F had a lower expected coverage (63 and 66% coverage, respectively), while the modified V4 reverse primer (E1007R) of Bråte et al. (2010), had lower sequence coverage (59%; see Fig. 1). The amplification success with the modified primers reported in the literature and the newly designed V4/V5 primers is shown in Fig. 2.

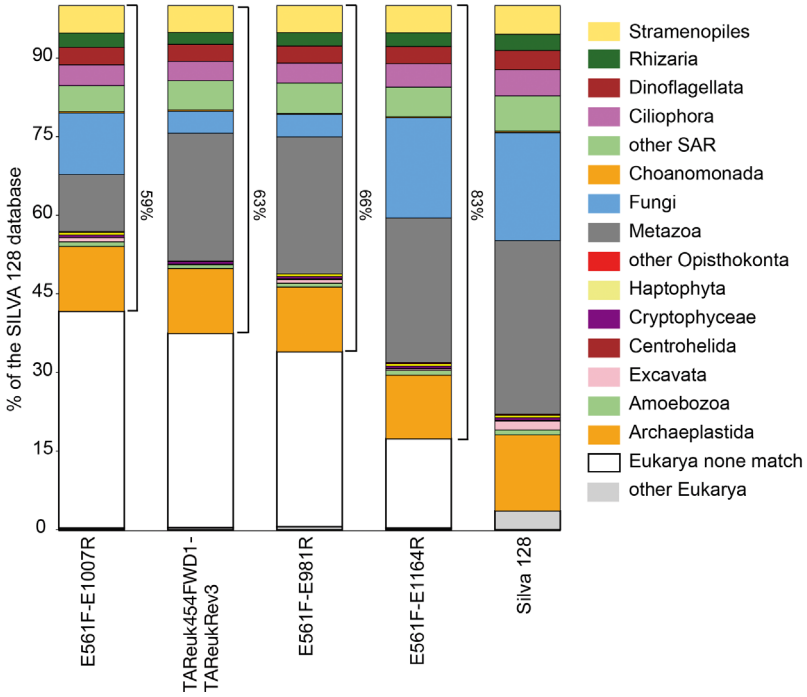

Fig. 1. Overview of the *in silico* coverage of the different V4 and V4/V5 primer sets used with eukaryotic targets (Silva 128 database). Ambiguities were taken into account. Forward primer E561F with E1007R (modified after Bråte et al. 2010) and E981R (modified after Stoeck et al. 2010) covers 59 and 66% of the eukaryotic sequences, respectively. With the new primer pair (E561F and E1164R), this was increased to 83%. For comparison, the distribution of different taxonomic eukaryotic groups in the Silva 128 database is shown (right column). The portion that is not covered by the respective primer pair is shown in white. SAR: Stramenopiles, Alveolata, Rhizaria

Sixteen pre-amplification cycles were optimal because no PCR products were obtained using fewer (10 or 13) cycles for pre-PCR, and a prokaryotic PCR product appeared when increasing the re-PCR cycle number to >25 cycles (see Fig. 2). Optimal results for the re-amplification were obtained with 1 µl of pre-amplified DNA, except for the sulfidic Black Sea sample, which required, probably due to low protist numbers, 5 µl of the template generated by pre-PCR. A slightly longer product could be removed by preparative gel electrophoresis and purification of the eukaryotic band. Amplicon sequencing was then conducted with the PCR products obtained with E561F and E1164R.

Analysis of the mock community

In general, all taxa included in the mock community were amplified by the newly designed primers and detected in the HTS result. However, relative read abundances of the individual taxa differed strongly from initial cell abundances (see Table S1 in the Supplement), especially using cDNA as template.

The mock community was used to evaluate clustering thresholds (carried out with DNA and cDNA) by a series of OTU-callings (90, 95, 97, 98, and 99%). Independent of the cut-off level, the quality-checked mock community reads resulted in a higher number of OTUs than of taxa present in the microscopic determination (Fig. 3). Especially *Cyclidium* sp. (ciliate) contained a much higher number of OTUs than expected in the DNA sample. A relatively good match was found for *Cafeteria roenbergensis* and *Codosiga balthica* and the DNA template of *Bodo*.

Analysis of the environmental Black Sea samples

Among the several amplicons obtained from the Black Sea sample by using the newly designed E561F and E1164R primers (Fig. 2), only amplicons of the expected length were sequenced due to the preparative gel electrophoresis. Without pre-amplifi-

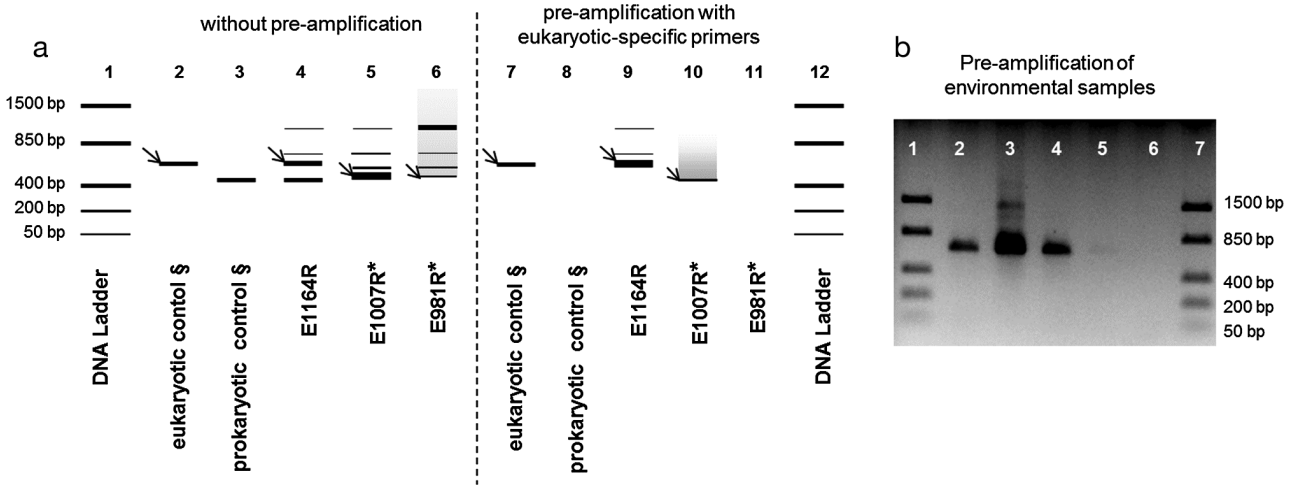

Fig. 2. Schematic presentation of amplification results obtained with the newly designed primers. The results are shown as (a) a schematic electrophoreses gel of the mock community amplified with the forward primer E561F and different reverse primers (lanes 2–6) using the cDNA template. PCRs without (left) and with (right, lanes 9–11) pre-amplification of the templates using eukaryotic-specific primers are shown. The expected bands of the eukaryotic PCR products obtained using the different V4 primers are indicated with arrows. The controls were amplified/re-amplified using E1164R (§); eukaryotic and prokaryotic controls were included (IOW109, *Ancyromonas sigmoides* and strain GD1, *Sulfurimonas gotlandica*, respectively). The marked (*) primers did not generate PCR products with the Black Sea templates. (b) Agarose gel of pre-amplified environmental DNA samples. Lane 2: suboxic Black Sea sample (130 m); lane 3: sulfidic Black Sea sample (155 m); lane 4: Baltic Sea sample (143 m); lane 5: Baltic Sea sample (95 m); lane 6: re-amplified blank control. Longer unspecific bands can occur (lanes 4 and 7 in panel a; lane 3 in panel b). Lanes 1 and 12 (panel a), and lanes 1 and 7 (panel b) show the FastRulerTM DNA ladder Low Range (Fermentas Life Sciences)

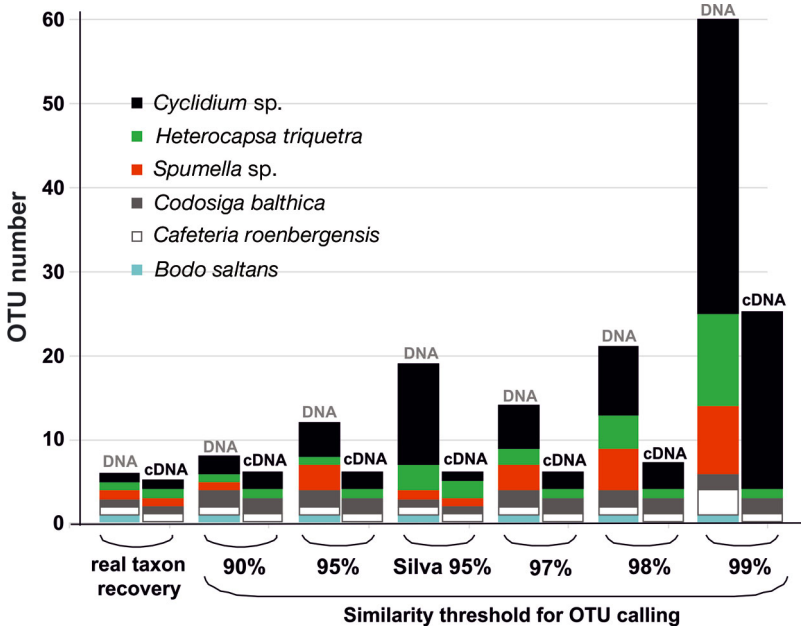

Fig. 3. Operational taxonomic unit (OTU) assessment using the mock community (containing 6 taxa) and the DNA and cDNA templates, with the application of different cut-offs and after a quality check and a denoising step (Acacia). All taxa detected in the raw data sets before the quality control are shown in the left columns (taxon recovery). *Bodo saltans* was not sequenced with the cDNA data set, whereas *Spumella* sp. was lost during sequence processing. For comparison, the Silva NGS pipeline was included; the cut-off was 95%

cation, only 16% protist reads but 67% prokaryotic (including 50% Archaea) and 17% metazoan reads in the Black Sea samples were recovered. In contrast, sequencing of the pre-amplified samples resulted in 81% protistan reads after quality control (Fig. S2 in the Supplement).

Using the new primers and pre-amplification, a total of 115 637 reads were obtained for the 6 samples, 19 187 (±4763, SD) reads on average per sample. After quality checking and denoising, there remained 33 117 Black Sea reads (44% of initial reads) and 33 509 mock community reads (83% of initial reads) that were clustered into OTUs (see below). For the environmental samples, a relatively large number of metazoan reads (mean 32%) and also some archaeplastid reads (mean 0.1%) were removed.

Based on the results from the mock community and the new primer system, we set a conservative cut-off

level of 95% in the OTU-calling of the environmental Black Sea samples. Following the procedure described above using both DNA and cDNA templates, we identified a total of 92 OTUs for the Black Sea redoxcline. Together with the 57 clone library OTUs identified in a previous study from the same samples (based on 98% sequence identity; Wylezich & Jürgens 2011), 106 different Black Sea OTUs were detected, 49 of which were only found in this study, 12 were specifically found in the clone libraries, and 45 were present in the HTS and clone library based study.

A comparison (presence/absence matrix) of OTUs resulting from the HTS approach (this study) and clone library OTUs (Wylezich & Jürgens 2011) is shown in Fig. 4 (for read abundances, see Table S1). In general, the cDNA samples (HTS and clone library) were more similar to each other than to the DNA samples. With both approaches, we found similar patterns regarding ciliates represented by sequences related to *Pleuronema* (OTUs 7 and 8) and *Strombidium* (OTUs 18–20 and 70) that were mainly dominant in suboxic water layers. Sequences affiliated with anaerobic plagiopylids (OTUs 14–16 and 68) and *Cyclidium* (OTUs 1–4 and 67) were detected only in the sulfidic zone. Among flagellates, a remarkable richness of bicosoecid taxa (OTUs 28–33, 37, 71, and 75–78, see Table S2 in the Supplement for read abundance) was confirmed for the Black Sea oxygen-depleted zone. Of the 22 dinoflagellate OTUs, most were dominating in the suboxic layer, whereas only 4 could be assigned to those in the sulfidic sample. The clone library and HTS data sets were compared on the basis of the Shannon diversity index and the OTU richness value, which both confirmed the pattern of higher values in the sulfidic sample (Fig. 5). Furthermore, both richness and diversity were lower using DNA rather than cDNA as the template.

DISCUSSION

Performance of the new V4/V5 primer system

The new V4/V5 primers (E561F and E1164R) were designed to amplify a broader range of eukaryotes compared to known V4 primers (Fig. 1). A remaining challenge was that these primers amplified parts of the rRNAs from all 3 domains of life (*Bacteria*, *Archaea*, and Eukaryota), although this was not predicted from *in silico* analyses. The prokaryotic portion in the environmental samples might have been preferentially amplified due to the shorter fragment lengths of the amplified region in prokaryotes than in eukaryotes (see Fig. 2). The use of these primers without prior steps to exclude or reduce prokaryotes from the sample (e.g. size fractionation of water samples or pre-amplification using eukaryotic-specific primers) would result in many prokaryote sequences at the disadvantage of protist reads. This was proven in our study by sequencing both the pre-amplified and not pre-amplified samples (Fig. S1). Pre-amplification — in some cases with a large number of pre-amplification cycles — has been included in other HTS studies using V4 primers (Bråte et al. 2010) with eukaryote- or ciliate-specific primers (Engel et al. 2012, Stoeck et al. 2014). Here, we tried to keep the number of pre-amplification cycles as low as possible, and the optimal amplification of the target region and target group (protists) was obtained with the relatively low number of 16 pre-PCR cycles.

From the Black Sea samples, the new primer set allowed the recovery of several ciliate groups including the formerly missing mesodiniids, as well as some dinoflagellates, apicomplexans, apusomonads, different stramenopiles (bicosoecids, diatoms, chrysophytes, several marine stramenopile clades, bolidophytes, and haptophytes), excavates (jakobids, different euglenozoan groups, and schizopyrenids), rhizarians, amoebozoans, and opisthokont protists (choanoflagellates). In a further study, in which Baltic Sea redoxcline samples were also analyzed using the new primer system, nucleariids, centrohelids, kathablepharids, goniomonads, and telonemids were additionally detected (C. Wylezich unpubl. results). In particular, the different excavate taxa and mesodiniids discriminated by other V4 primer systems were found to be abundant and important taxa in the Black Sea oxygen-deficient zone (Wylezich & Jürgens 2011) and comparable habitats, such as the central basins of the Baltic Sea (Weber et al. 2014), as determined by clone libraries and microscopy counts. Consequently, these taxa can now be included with the new primer set in HTS surveys.

Several newly designed 18S rRNA gene primers of different variable regions, applicable to HTS approaches have been proposed previously (Hugerth et al. 2014). One of the primer pairs, 563f and 1132r, corresponds to the same regions of the 18S rRNA gene as the primer pair E561F and E1164R. The environmental samples of Hugerth et al. (2014), sequenced with primers for the same fragment but without pre-amplification, also produced shorter PCR products compared to those in our study. In the cited study, amplicons of this fragment were not

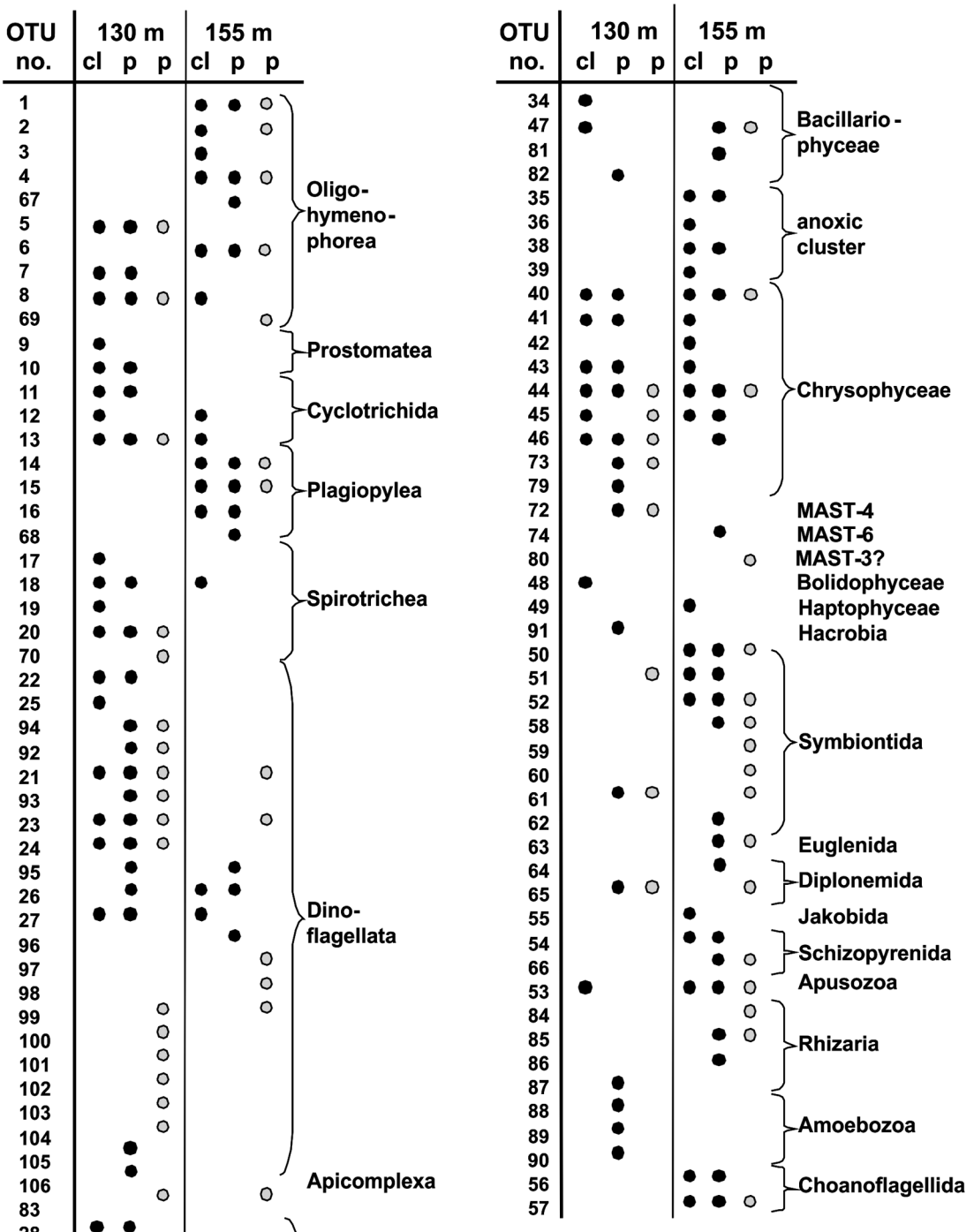

Fig. 4. Presence/absence matrix of the sequencing reads of the 2 Black Sea samples taken from oxygen-deficient waters (suboxic, 130 m; sulfidic, 155 m), resulting in 106 operational taxonomic units (OTUs). Clone library (cl) sequences (57 OTUs; cut-off 98%; Wylezich & Jürgens 2011) and pyrosequencing (p) reads (92 OTUs; cut-off 95%; this study) are compared. Black dots refer to cDNA templates. The high-throughput sequencing (HTS) approach (454 pyrosequencing) was applied to both DNA (grey dots) and cDNA (black dots) templates

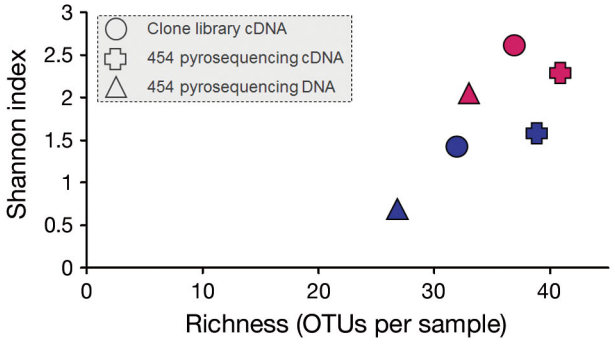

Fig. 5. Comparison of sample sequences obtained using different methods with operational taxonomic unit (OTU) richness and the Shannon diversity index. Sulfidic (155 m) and suboxic (130 m) samples are shown as red and blue symbols, respectively

sequenced, but the short products were assumed to represent the true biological protist diversity indicated by the V4/V5 region. However, based on our results, we assume that the short products of an amplified environmental sample represented mainly prokaryotic amplicons.

In contrast to the 454 GS-FLX platform used in this study, current Illumina sequencing platforms are limited regarding fragment size, which results in a lack of overlap for paired end sequencing when using the proposed primer pair. However, Ion Torrent sequencing, providing about 400–600 bp, or third generation sequencing platforms like PacBio or MinIon produce sufficient length for phylogenetic analysis using this primer set. In addition, the primers designed here are useful in future studies with new sequencing technologies for longer sequence fragments, which also result in a higher and more reliable taxonomic resolution (Höper et al. 2017).

Abundance and richness in the model protist community

In general, the cells of larger protists, especially ciliate and dinoflagellate taxa, are highly overrepresented in HTS results of mock communities (Table S1; see also Geisen et al. 2015) or when compared with morphological data (Medinger et al. 2010), because 18S rRNA gene copy-numbers of these groups are very high (e.g. Gong et al. 2013). Moreover, read numbers of flagellates do not represent the initial cell abundance in a defined mock community sample (Table S1, compare also Egge et al. 2013, Geisen et al. 2015). Nonetheless, it should be mentioned that the degree of bias in either direction might be highly variable, and this study only illustrates the variability among the used taxa and certain growth conditions and cannot be extrapolated to other taxa or to the same taxa in different compositions.

Although our results indicate the parallel amplification of amplicons of different lengths, longer sequences are usually biased in the PCR amplification as well as the subsequent sequencing procedure. Hence, this study strongly supports the conclusion that PCR-based HTS data cannot be used in quantitative assessments of complex natural protist assemblages (see also Medinger et al. 2010, Bachy et al. 2013, Stoeck et al. 2014). Thus, until more precise methods are developed and thoroughly tested, read abundances should be carefully validated within the study of interest using cell counts and quantitative methods, such as cell hybridization, and otherwise treated mainly as qualitative data.

By applying the commonly used OTU clustering cut-offs (97–99% similarity), we detected much higher OTU numbers than the number of species actually present in the mock community, especially when using DNA (Fig. 3). Although it seems to distort the relative sequence abundance information to an extent to which some species may fall beyond the limit of detection of HTS (Egge et al. 2015), the use of cDNA templates might improve read quality or sequence consistency. Genomes often contain numerous rRNA gene loci but only a few are transcribed. As a result, a reduced diversity of 18S rRNA sequences might be expected when RNA is analyzed, compared to DNA. Nonetheless, the excess OTUs were surprising, since the mock community consisted of clonal cultures, which theoretically should have produced only 1 OTU per species. The overestimation of OTUs in the mock community reads might also have been due to our choice of the less computation-intensive denoising program Acacia instead of AmpliconNoise (Bachy et al. 2013, Egge et al. 2013). However, Bragg et al. (2012) recommended Acacia as an alternative to AmpliconNoise because it introduces fewer errors, especially non-homopolymer substitutions, into the data set. An alternative approach is the Silva NGS pipeline (Glöckner et al. 2017), in which quality control is achieved by aligning the sequences with those in a large, manually curated database. Here, sequences with high dissimilarity to those in the database are rejected (Klindworth et al. 2013). Although we used the Silva NGS pipeline for the mock community, with a similarity threshold of 95%, this still resulted in an artificially high number of OTUs, particularly for the ciliates.

Characterization of a Black Sea redoxcline protistan community using the new primer system

Based on the OTU assessment of the mock community (using Acacia and JAguc), a 95% cut-off was applied to cluster our Black Sea HTS reads into OTUs and to evaluate the new primer system with environmental samples. Undoubtedly, this conservative cut-off prevented the detection of many closely related taxa. Bachy et al. (2013) found that by applying a 95% similarity threshold for OTU separation their HTS data (from 454 pyrosequencing) generally matched the morphologically estimated species numbers in tintinnids (ciliates). However, they still imposed a 99% cut-off since this was the threshold determined by Sanger sequencing studies. Egge et al. (2013) could separate closely related haptophyte taxa using 99% OTU clustering. In general, the handling of sequence clusters as OTUs is an auxiliary concept used for environmental sequence data of microbial eukaryotes. The application of a particular threshold for OTU separation to analyze a complex community of protists with different rates of genetic evolution is taxon-dependent and a compromise between the species and the genus level (compare Caron et al. 2009).

We used different similarity cut-offs in the HTS study compared to the clone library based study. This was done since OTUs from the Sanger sequenced clone library represent a manually corrected 18S rRNA gene sequence (Wylezich & Jürgens 2011), which we consider highly reliable. In 454 pyrosequencing, by contrast, a comparable read correction is not possible. Quality checks can be performed using different algorithms, but for HTS reads there is no means of correction comparable to the chromatograms commonly used for Sanger sequences.

In general, the 2 approaches — HTS sequencing and the clone library (Wylezich & Jürgens 2011) — produced similar patterns. Sequences related to *Pleuronema* (OTUs 7 and 8) and *Strombidium* (OTUs 18–20, 70) were mainly dominant in suboxic water layers, whereas sequences affiliated with anaerobic plagiopylids (OTUs 14–16, 68; for read abundances see Table S2) and *Cyclidium* sp. (OTUs 1–4, 67) were detected only in the sulfidic sample. In the case of flagellates, a remarkable richness of bicosoecid taxa (OTUs 28–33, 37, 71, 75–78) was confirmed for oxygen-depleted zones in the Black Sea, in contrast to other pelagic redoxclines (compare Table 2 of Wylezich & Jürgens 2011). Excavate taxa were concentrated in sulfidic waters, whereas dinoflagellates, with the highest OTU richness, were dominant in suboxic conditions. Only 4 of the 22 dinoflagellate OTUs were detected in the sulfidic water layer. This result is in concordance not only with our earlier data from the Black Sea but also with the study of Weber et al. (2014) for the Baltic Sea redoxcline in which the strongest shifts in composition occurred at the transition to anoxic and sulfidic water layers.

The sulfidic water sample of the Black Sea contained a higher diversity and richness than the suboxic water sample, as demonstrated by clone libraries, DGGE fingerprinting (Wylezich & Jürgens 2011), and HTS (this study), based on both templates (DNA, cDNA; Fig. 5). Thus, the upper sulfidic water layer likely represents a diversity hot spot within oxygen-depleted waters. The absence of predation pressure by small metazoans in these waters might be one reason for this phenomenon. Weber et al. (2014) concluded that low oxygen content (1–2 µM) is responsible for the very dramatic changes in the oxic and anoxic protist communities of the Baltic Sea. The specialized taxa found in this transition zone and in the upper sulfidic zone are well-adapted to low-oxygen or oxygen-free and low-sulfide conditions. Additionally, protist taxa with broad tolerance ranges in terms of oxygen and hydrogen sulfide seem to overlap with specialists, giving rise to high levels of diversity and species richness.

Altogether, the applicability of the newly designed primers and PCR protocol proves to be suitable to assess protist diversity of oxygen-deficient systems. The primers have a high coverage for protist taxa including the often neglected kinetoplastids. The V4/V5 fragment length differences of prokaryotes (450–500 bp) and eukaryotes (>600 bp) are very helpful in separating protists from prokaryotes using preparative gel electrophoresis.

Concluding remarks

In this study, we designed and tested new eukaryotic primers which recovered a larger spectrum of protist groups from oxygen-deficient water columns than previously used primers for HTS approaches. Like other fingerprinting methods, HTS studies are undoubtedly powerful tools to compare different samples and to obtain preliminary information about the presence of taxa in previously unstudied environments. With the new primer pair, we succeeded in increasing the diversity of taxonomic protist groups as shown for the suboxic and anoxic zones of the Black Sea. For other habitats, it should also be possible to gain a more complete picture of the protistan

diversity with the new primers. However, confirming previous studies, our results also revealed that the abundance information obtained with this approach is highly biased and can thus easily lead to misinterpretations of both the real cell abundance and diversity estimates. Therefore, we recommend lower cut-off values (95–97%) for HTS data than for Sanger sequencing data for more realistic estimates of richness and Shannon diversity.

Acknowledgements. The assistance of the captain and crew of the RV 'Meteor' is gratefully acknowledged. We are indebted to Katja Becker and Bärbel Buuk for excellent technical support, as well as to Alexander P. Mylnikov (Borok, Russia) and Hanne Halliger (Sylt, Germany) for isolating and providing cultures. Sincere thanks go to Sabine Glaubitz for providing the DNA of strain GD1 and to Felix Weber for helpful advice regarding the experimental design. This work was funded by a grant from the DFG (JU 367/11–1).

LITERATURE CITED

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLOS ONE 4:e6372

Anderson R, Winter C, Jürgens K (2012) Protist grazing and viral lysis as prokaryotic mortality factors at Baltic Sea oxic–anoxic interfaces. Mar Ecol Prog Ser 467:1–14

Bachy C, Dolan JR, López-García P, Deschamps P, 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, 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

Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW (2012) Fast accurate error-correction of amplicon pyrosequences using Acacia. Nat Methods 9:425–426

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

Caron DA, Countway PD, Savai P, Gast RJ and others (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microbiol 75: 5797–5808

Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 45 pyrosequencing. ISME J 4:1053–1059

Díez B, Pedrós-Alió C, Marsh TL, Massana R (2001) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. Appl Environ Microbiol 67:2942–2951

Dunthorn M, Klier J, Bunge J, Stoeck T (2012) Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. J Eukaryot Microbiol 59:185–187

Edgcomb V, Orsi W, Bunge J, Jeon S and others (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. ISME J 5:1344–1356

Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B (2013) 454-pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. PLOS ONE 8:e74371

Egge ES, Johannessen TV, Andersen T, Eikrem W and others (2015) Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. Mol Ecol 24:3026–3042

Engel M, Behnke A, Klier J, Buschbaum C, Volkenborn N, Stoeck T (2012) Effects of the bioturbating lugworm *Arenicola marina* on the structure of benthic protistan communities. Mar Ecol Prog Ser 471:87–99

Geisen S, Laros I, Vizcaíno A, Bonkowski M, de Groot GA (2015) Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. Mol Ecol 24:4556–4569

Glöckner FO, Yilmaz P, Quast C, Gerken J and others (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. J Biotechnol 261: 169–176

Gong J, Dong J, Liu X, Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. Protist 164:369–379

Guillard RRL, Ryther JH (1962) Studies of marine planktonic diatoms I. *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran. Can J Microbiol 8:229–239

Guillou L, Viprey M, Chambouvet A, Welsh RM and others (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). Environ Microbiol 10:3349–3365

Guillou L, Bachar D, Audic S, Bass D and others (2013) The protist ribosomal reference database (PR^2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res 41:D597–D604

Höper D, Wylezich C, Beer M (2017) Loeffler 4.0 — diagnostic metagenomics. In: Beer M, Höper D (eds) In Loeffler's footsteps — viral genomics in the era of high-throughput sequencing. Advances in viral research, Vol 99. Academic Press, Cambridge, MA, p 17–37

Hugerth LW, Muller EEL, Hu YOO, Lebrun LAM and others (2014) Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. PLOS ONE 9:e95567

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwälder B, Schlötterer C, Boenigk J (2010) Diversity in a hidden

world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol 19:32–40

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71:491–499

Nebel ME, Wild S, Holzhauser M, Hüttenberger L, Reitzig R, Sperber M, Stoeck T (2011) JAguc — a software package for environmental diversity analyses. J Bioinform Comput Biol 9:749–773

Nolte V, Pandey RV, Jost S, Medinger R, Ottenwälder B, Boenigk J, Schlötterer C (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. Mol Ecol 19: 2908–2915

Pawlowski J, Christen R, Lecroq B, Bachar D, Shahbazkia HR, Amaral-Zettler L, Guillou L (2011) Eukaryotic richness in the abyss: insights from pyrotag sequencing. PLOS ONE 6:e18169

Stoeck T, Behnke A, Christen R, Amaral-Zettler L and others (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biol 7:72

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic waters. Mol Ecol 19:21–31

Stoeck T, Breiner HW, Filker S, Ostermaier V, Kammerlander B, Sonntag B (2014) A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Environ Microbiol 16:430–444

von der Heyden S, Cavalier-Smith T (2005) Culturing and environmental DNA sequencing uncover hidden kinetoplastid biodiversity and a major marine clade within ancestrally freshwater *Neobodo designis*. Int J Syst Evol Microbiol 55:2605–2621

Weber F, Anderson R, Foissner W, Mylnikov AP, Jürgens K (2014) Morphological and molecular approaches reveal highly stratified protist communities along Baltic Sea pelagic redox gradients. Aquat Microb Ecol 73:1–16

Wylezich C, Jürgens K (2011) Protist diversity in suboxic and sulfidic waters of the Black Sea. Environ Microbiol 13: 2939–2956

Wylezich C, Karpov SA, Mylnikov AP, Anderson R, Jürgens K (2012) Ecologically relevant choanoflagellates collected from hypoxic water masses of the Baltic Sea have untypical mitochondrial cristae. BMC Microbiol 12:271

Editorial responsibility: Robert Sanders, Philaldelphia, Pennsylvania, USA

Submitted: June 21, 2017; Accepted: November 23, 2017
Proofs received from author(s): February 5, 2018