A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology

Thorsten Stoeck,1 Hans-Werner Breiner,1 Sabine Filker,1 Veronika Ostermaier,2 Barbara Kammerlander2,3 and Bettina Sonntag2,*
1Department of Ecology, Faculty of Biology, University of Kaiserslautern, Kaiserslautern, Germany.
2Research Institute for Limnology, Mondsee, University of Innsbruck, Mondsee, Austria.
3Institute of Ecology, University of Innsbruck, Innsbruck, Austria.

Summary

Analyses of high-throughput environmental sequencing data have become the 'gold-standard' to address fundamental questions of microbial diversity, ecology and biogeography. Findings that emerged from sequencing are, e.g. the discovery of the extensive 'rare microbial biosphere' and its potential function as a seed-bank. Even though applied since several years, results from high-throughput environmental sequencing have hardly been validated. We assessed how well pyrosequenced amplicons [the hyper-variable eukaryotic V4 region of the small subunit ribosomal RNA (SSU rRNA) gene] reflected morphotype ciliate plankton. Moreover, we assessed if amplicon sequencing had the potential to detect the annual ciliate plankton stock. In both cases, we identified significant quantitative and qualitative differences. Our study makes evident that taxon abundance distributions inferred from amplicon data are highly biased and do not mirror actual morphotype abundances at all. Potential reasons included cell losses after fixation, cryptic morphotypes, resting stages, insufficient sequence data availability of morphologically described species and the unsatisfying resolution of the V4 SSU rRNA fragment for accurate taxonomic assignments. The latter two underline the necessity of barcoding initiatives for eukaryotic microbes to better and fully exploit environmental amplicon data sets, which then will also allow studying the potential of seed-bank taxa as a buffer for environmental changes.

Introduction

Traditional approaches for the investigation of protistan plankton diversity, including enrichment and clonal cultivation, manual selection of individual cells, live observation through microscopy and fixation and staining procedures were the methods of choice for several hundred years to study the composition, dynamics and distribution of protistan plankton communities (Patterson, 1992 and references therein). More than 10 years ago, the interrogation of taxonomic marker genes obtained from environmental samples amended these traditional approaches in protistan plankton diversity research (López-García et al., 2001; Moon-van der Staay et al., 2001). Since then, a vast number of studies on the molecular diversity of protistan plankton has contributed to our understanding and perception of this group. Important discoveries include: (i) the detection of an ecologically and numerically important plankton group, namely eukaryote picoplankton, which is microscopically difficult to detect and to identify because of small cell sizes (< 3 μm) and insufficient morphological characters (Massana et al., 2006); (ii) the identification of morphologically indistinguishable cryptic species (Nanney et al., 1998; de Vargas et al., 1999; Darling et al., 2004; Stoek et al., 2008); (iii) a genetic heterogeneity within protists that is much higher when compared with the findings of studies on morphological diversity (Foissner et al., 2008); (iv) the detection of novel taxon groups on high taxonomic levels such as the novel class Cariacotrichea within the morphologically well-characterized ciliates (Stoek et al., 2006; Orsi et al., 2012); (v) the existence of life and active eukaryote plankton communities in even the most extreme environments on our planet, which were previously difficult to sample and to analyse by traditional plankton research approaches (López-García et al., 2003; Alexander et al., 2009; Edgcomb et al., 2009); and (vi) the detection and identification of biogeographic and temporal patterns through the high resolution of molecular phytotypes previously obscured by the lower resolution of morphotypes (Bass et al., 2007; Boenigk et al., 2007).
The basic idea of a molecular approach included the extraction of genomic DNA or RNA from environmental samples, the polymerase chain reaction (PCR) amplification of taxonomic marker genes, predominantly the small subunit of the ribosomal RNA gene (SSU rRNA) with specific oligonucleotide primer sets, the isolation and multiplication of individual PCR fragments through cloning into plasmid vectors and transformation in competent bacterial cells, and the subsequent isolation, purification and Sanger sequencing of the obtained target genes. These sequences were then placed into a phylogenetic context and analysed statistically. This strategy evolved to the ‘gold standard’ in protistan diversity research (Caron and Countway, 2009; Dawson and Hagen, 2009; Stoeck and Epstein, 2009; Stoeck et al., 2010; Caron et al., 2012; Bittner et al., 2013). One of the hypothesized functions of the rare biosphere assumed that, for example, the high genomic potential of the numerous but rare seed-bank taxa (relative abundance of a species < 0.1%) allowed a microbial community to react to environmental changes (Sogin et al., 2006; Caron and Countway, 2009; Dawson and Hagen, 2009; Stoeck and Epstein, 2009).

In this study, we investigated a protistan morphotype community structure in a mountain lake (Piburgersee, Austria) directly compared with pyrosequenced ciliate amplicon data. As the effort to identify, quantify and finally record the diverse microbial morphotypes within a specific ecosystem is enormous, we focused on the morphologically relatively well distinguishable ciliate plankton. In general, ciliates are major plankton components and play crucial roles in microbial food webs. For example, they belong to the most important consumers of algal biomass in lakes (Weisse and Müller, 1998), contribute to the top-down control of specific bacterial taxon groups (Šimek et al., 2002), and they transfer energy and organic matter from the microbial food web to higher trophic levels (Weisse and Müller, 1998).

Here, we analysed the morphotypes and pyrosequenced phylotypes of ciliates from the same samples taken during a unique sampling event in winter 2011. We presumed that the microscopically detected morphotypes represented the numerically abundant taxa and accordingly expected most of them to be reflected quantitatively in the abundant fraction of the amplicon data set. Furthermore, considering the seed-bank hypothesis for low-abundant phylotypes (see above), it was reasonable to assume that the highly sensitive pyrosequencing (if sampled to saturation) would reveal the near-complete ciliate plankton inventory. Therefore, we expected to find ciliate taxa in the seed-bank phylotypes that were microscopically undetectable in the winter sampling but recorded throughout an annual sampling at the same site.

Results

Abiotic parameters and chlorophyll a

On the day of sampling, the clear ice cover was approximately 40 cm thick and covered by ca 15 cm of snow. We measured the most important abiotic parameters and chlorophyll a to show the pronounced differences between the three selected sampling depths (Table 1). Temperature was lowest (1.2°C) directly under the ice
cover followed by more or less stable values down to the greatest depths (21 m). Oxygen and chlorophyll a concentrations decreased from surface to bottom waters, while ammonium increased. Highest values for total phosphorus were detected directly under the ice with a decrease in 9 m and again an increase in the bottom waters. The other parameters remained more or less stable throughout the water column.

V4 amplicon analyses

After data cleaning, we obtained 82,493 ciliate V4 tags from the surface water sample (0 m), 57,117 ciliate tags from the 9 m sample and 87,445 tags from the bottom water sample (21 m). When clustering all tags into operational taxonomic units (OTUs), OTU numbers collapsed exponentially with decreasing cluster levels from 100% to 97%, followed by a gentle linear decline from 97% to 90% (Fig. 1). OTUs clustering at 97% sequence similarity (OTU97) resulted in 1386, 1464 and 1657 OTU97 for water samples from 0 m, 9 m and 21 m respectively. The rank-abundance distribution (Fig. S1) showed relatively few singletons and doubletons (OTUs97 with one and two tags respectively) indicating near-saturated sampling profiles. Moreover, the rarefaction analyses (Fig. 2) confirmed that the sequencing effort was sufficiently deep to record the near-complete OTU97-inventory. This was a substantial prerequisite to address the questions raised in this study and for solid comparative analyses of molecular samples to each other as well as to the morphological data sets.

In total, 69.4% (n = 157,662) of all sequences detected in March 2011 showed a sequence similarity of >95% to their closest Basic Local Alignment Search Tool (BLAST) hit in the ciliate V4 18S rDNA database (Fig. 3). These sequences were taxonomically assigned to ciliate genera, and the remaining tags obtained a ‘candidatus status’ on the genus level (Table S1). In total, the assigned sequences represented 47 different ciliate genera with distinctively different relative proportions (Fig. 4A). The highest amplicon abundances were assigned to Pseudomonilicaryon (43.0%), Orthoamphisiella (18.4%) and Askenasia (12.6%). Rimostrombidium, Urocentrum, Halteria, Didinium, Enchelyodon and Spirostomum accounted for 1–10% whereas most of the remaining genera (n = 38) were represented by less than 1% of all assigned tags.

Comparing morphological data with sequencing data

From morphologic analyses, we detected 30 ciliate genera (Fig. 4B) including 48 species (Table S2). In total,
in the March sampling we observed 2.4 ± 1.2 ind. ml$^{-1}$ directly under the ice cover, 3.2 ± 0.3 ind. ml$^{-1}$ in 9 m and 5.4 ± 2.4 ind. ml$^{-1}$ in 21 m respectively. Significant differences were found between pyrosequencing and morphological data from the March sampling (Fisher’s exact test: $P < 0.01$) as only 11 morphogenera out of the 47 genera detected in the sequence survey were in accordance (Fig. 4A and B). Amplicon sequencing revealed that 36 genera remained undetected from morphological analyses. Similarly, 19 genera were detected microscopically but escaped the molecular survey. We note that eight of these genera (red bars in Fig. 4B) had no representative sequences in publicly accessible databases, and five were detected in the sequencing survey as candidatus genera (blue bars in Fig. 4B). Thus, only six genera that were detected morphologically and had representative V4

Fig. 2. Rarefaction curves of amplicon data sets from the three different depths screened for ciliate phylotype diversity in Piburgersee in March 2011. All three data sets showed near-saturation profiles, suggesting that the nearly complete ciliate phylotype inventory (OTUs called at 97% sequence similarity) was obtained. Data were rarified to 82 140 sequences per sample.

Fig. 3. Similarity distribution (percentage of identity) between all high-quality amplicon reads ($n = 227 065$) and SSU rDNA V4 regions of database Sanger sequences of described ciliates as obtained from NCBI’s nr database collection. Note that most amplicons had a sequence similarity of at least 98% to database entries. In total, 69.4% of all high-quality sequences were at least 95% similar to deposited sequence entries of ciliate morphospecies.

© 2013 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd, Environmental Microbiology, 16, 430–444
Fig. 4. Relative proportion of identified genera in samples analysed in this study. (A) Pyrosequenced amplicon data collected in March 2011. Only amplicons were considered that were at least 95% similar to deposited database sequences from ciliate morphospecies. (B) Morphotype data collected in March 2011. Sample aliquots were obtained from the same bulk water samples that were used for pyrosequencing. (C) Morphotype data collected over the course of 1 year (see Fig. 5) at the same sampling location and depths where samples for Fig. 4A and B were obtained. Blue bars in (B) and (C) are genera that were also detected with pyrosequenced amplicons, but sequence similarity to closest relative sequence in reference database was < 95% (‘amplicon candidatus genera’). Red bars represent genera not available in public databases. Green bars represent genera available as sequence entries in the reference database but escaped amplicon detection.
Assignment with JAGUC (Nebel et al., 2011) data sets compared with other widely used taxonomic strategies (Nebel et al., 2011). Because of its high sensitivity and a higher taxonomic resolution power of hypervariable fragments in taxonomic marker genes (Pawlowski et al., 2012), it was not unexpected that sequencing discovered many more taxa than detected through the microscopic analyses. However, taxon composition and abundances obtained from sequencing data were hardly congruent with morphotype data. Qualitatively, 20 morphotype genera (63%) remained undetected in the sequence analyses. We attribute this to the following main reasons:

(i) The V4-database assembled from publicly available data included sequences from only 308 ciliate morphospecies belonging to 247 genera. This contrasts with the number of ca 4500 globally described free-living ciliate species (Foissner et al., 2008) included into ca 1500 genera (Aescht, 2001). Our study makes evident that the value of a molecular sequence data set can only be as strong as the underlying database. The importance of cataloguing sequence data from described protist species in databases as suggested, for example, in the protistan barcoding initiative (Pawlowski et al., 2012) is mandatory to make more use of environmental sequence data sets and to obtain less biased pictures of natural protist communities.

(ii) The accuracy of taxon assignment at the genus level seemed to be biased at least in some cases, especially when it comes to the hypervariable V4 region of the SSU rDNA. This becomes obvious in five detected morphogenera in our study (Fig. 4B, blue bars). We found V4 amplicons that could tentatively be assigned to these genera, but their low sequence similarity made the taxon assignment uncertain. This category included also falsely assigned sequences, which may have obscured the presence of specific taxa in the sample. For example, morphological analyses of the ciliate inventory in Piburgersee revealed that the euplanktonic Pelagodileptus trachelioideus was a common species found in the lake throughout the year (B. Sonntag, unpublished). Surprisingly, via sequencing, Pelagodileptus was not detected; instead Pseudomonilicaryon was the quantitatively most important genus accounting for the largest DNA pool. Yet the latter morphologically conspicuous genus has never been recorded before microscopically from Piburgersee, neither from live nor from preserved plankton samples (Fig. 4B and C; B. Sonntag, pers. obs.). Morphotaxonomically, both genera are assigned to the Moniliacaryon branch of the family Dileptidae represented by two clades including Pseudomonilicaryon and Moniliacaryon on the one hand and Paradileptus and Pelagodileptus on the other (Vďačný et al., 2011). A detailed discussion about the morphological and molecular assignment of these genera can be found in Vďačný and colleagues (2011). However, as all these genera resembled conspicuous medium-
large-sized (around 100–800 μm in length) ciliate species, it seems unlikely that such individuals could have been overlooked over decades. Further, we found that the V4 fragments of *Pseudomonilicaryon* and *Pelagodileptus* were identical. Thus, the computational selection of *Pseudomonilicaryon* as closest related database entry to the corresponding V4 amplicon was arbitrary. Several genes, including the SSU rDNA (Lynn and Strueder-Kypke, 2006), the cytochrome oxidase subunit 1 (Greczek-Stachura et al., 2012), the intergenic transcribed spacer region (Gentekaki and Lynn, 2009) and protein-coding genes were evaluated and discussed as potential barcodes for ciliates (Barth et al., 2008), but none of them could satisfy the criteria to serve as an adequate genetic species marker (Pawlowski et al., 2012). However, at least for tintinnid ciliates, the ITS of the SSU rDNA seemed to provide a good marker even for species identification (Bachy et al., 2012; 2013). The quest for protistan barcodes can be defined as a major and important task for a better interpretation of environmental sequence data sets. It may be legitimate to question the necessity of putting a Linnean taxonomy onto sequence OTUs (Bittner et al., 2010). However, because biological properties are coded in a taxon name (Schuh and Brower, 2000; Sonntag et al., 2008), we believe that this is mandatory for a better understanding and interpretation of gene data sets in ecology.

(iii) Technical difficulties may have prevented some genera belonging to the ciliate consortium under study here, from PCR amplification and/or sequencing (Fig. 4B, green bars). Such biases included primer specificities as in the case of *Mesodinium*. The ciliate-specific forward primer used in this study had three mismatches to its target region in the primary structure of the *Mesodinium* SSU rDNA sequence. Interestingly, *Mesodinium* is well known for its highly divergent SSU rDNA sequence distinguishing this genus from other ciliates and even suggesting that this taxon did not belong to the phylum *Ciliophora* (Johnson et al., 2004). Other taxa such as *Cinetochilum* detected microscopically but not genetically in this study have longer inserts in the V4 region, which may result in secondary structures that could prevent this gene fragment from amplification or sequencing (Pinto and Raskin, 2012). Alternatively, the increased length of this specific V4 fragment compared with the V4 fragments of other taxa in the same sample may result in length discrimination during pyrosequencing.

Numerous genera that were detected with pyrosequencing remained unobserved microscopically. Three major reasons might explain this discrepancy. First, for microscopic analyses, protists are commonly fixed immediately after sampling and species-specific cell losses of more than 40% can occur after preservation (Pfister et al., 1999; Sonntag et al., 2000). Second, if cell abundances of specific taxa in the water sample fall below a specific threshold, these cells may still be detectable with a highly sensitive molecular tool as used here, but may not be found in small volumes of water screened by microscopy. Third, resting stages of ciliates that cannot be identified and assigned correctly morphologically may be more easily recorded in a molecular survey (Medinger et al., 2010). However, although ciliate resting cysts can be quite conspicuous (Müller et al., 2002) and hardly be overlooked in plankton samples, they were only occasionally observed in live and preserved samples in Piburgersee in the last decades (B. Sonntag, pers. obs.).

**Taxon abundance distribution**

Quantitatively, we had expected that at least the most abundant morphotype genera would have been mirrored in the molecular ciliate profiles. However, considering the genera that were detected simultaneously by microscopy and sequencing, we could not find a correlation in abundance distribution (Fig. S2). The taxon-assigned amplicon abundances did not reflect the true taxon abundances in the samples considered in this study. This agrees with findings of Medinger and colleagues (2010), who also found incongruities between morphotype and phylotype abundances in the protist community of an oligotrophic freshwater lake, and also with a study by Amend and colleagues (2010) on fungal communities in house dust.

We mainly explain our observation with the highly variable SSU rDNA copy number in different protist taxa. In their study, Zhu and colleagues (2005) demonstrated that the SSU rDNA copy number could vary several thousand fold amongst different protist taxa. The authors gave evidence for a direct correlation between SSU rDNA copy number and cell size. In the following case, we demonstrate that this is a major reason for biased protistan abundance data obtained from pyrosequencing: the most abundant morphotypes identified in March 2011 belonged to the genus *Urotricha* with 905 ± 408 ind. l−1 on average accounting for only a relatively small proportion of amplicons (Fig. 4A and B). In contrast, the average abundance of *P. trachelioides* was as low as 3.4 ± 5.7 ind. l−1 in the quantitative March 2011 samples, although, interestingly, from the number of amplicons observed, *Pelagodileptus* was the most abundant taxon (assuming that the *Pseudomonilicaryon*-assigned amplicons actually belonged to *Pelagodileptus* – see discussion earlier). This discrepancy between low abundance of individuals vs. high amplicon number and vice versa can most probably be explained by the number of DNA copies found in the size of the macronucleus of both genera. Large species of several hundred micrometers in length such as *P. trachelioides* or *Pseudomonilicaryon fraterculum* have a moniform macronucleus consisting of many nodules (for a detailed compilation of the species, see Vďačný and

© 2013 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd, Environmental Microbiology, 16, 430–444
Foissner, 2012). Their average macronuclear volume per individual can be calculated as follows: assuming 15 macronuclear nodules for *P. trachelioides*, each of a volume of 1600 μm³, the total volume would be 24 000 μm³. For *P. fraterculum*, 26 nodules were found on average with approximately 200 μm³ per nodule accounting for 5200 μm³ for the whole macronucleus. In contrast, urotrichs usually bear only one macronuclear nodule (assuming 10 x 5 μm resulting in 130 μm³ per macronucleus). From these exemplary calculations, we clearly see the discrepancy to quantify and directly relate amplicon numbers to ciliate abundance in a habitat. Therefore, we have to be aware of misinterpretation of protistan environmental amplicon data and ecological analyses of α- and β-diversity. These include, for example, indices such as Morisita–Horn, Bray–Curtis and the abundance-based Shannon index, as well as richness estimators including Abundance-based Coverage Estimator (ACE) and Chao (Chao and Shen, 2003–2005; Colwell, 2009). Because some indices are strongly influenced by rank-abundance distribution and are also sensitive to the few highly abundant taxa, they are of very little value for environmental molecular sequence data (Haegeman et al., 2013). Therefore, at the current state of the art, we suggest to evaluate such pyrosequencing data sets from presence/absence data only.

**Seed-bank taxa as a buffer for environmental changes?**

Our sequencing effort for the planktonic ciliate community in Piburgersee was considerably higher (ca three orders of magnitude) when compared with previous Sanger sequencing of environmental clone libraries (e.g., Edgcomb et al., 2002; Massana et al., 2004; Zuenendorf et al., 2006; Countway et al., 2007; Stock et al., 2012). Yet the vast majority of OTUs (97%) included very few amplicons (< 0.1% of all amplicons). This agrees with previous environmental pyrosequencing surveys of protistan diversity. In the literature, these numerous low-abundant OTUs, typically perceptible as long-tail distributions in OTU rank-abundance curves, are interpreted as the rare biosphere or so-called seed-bank taxa (Sogin et al., 2006; Pedrós-Alió, 2007; Dawson and Hagen, 2009; Stoeck et al., 2010). Even though a part of this numerically pronounced rare biosphere is the result of data production and processing (Kunin et al., 2010; Behnke et al., 2011), approved strategies were developed and applied in this study (s. methods) to minimize its artificial inflation. There is much speculation about the ecological function of these seed-bank taxa, and one suggestion is that they may buffer environmental changes (Sogin et al., 2006; Pedrós-Alió, 2007; Dawson and Hagen, 2009; Stoeck and Epstein, 2009). Seasonal changes in an aquatic ecosystem are an example of environmental changes, and as expected, the ciliate morphotype community in Piburgersee showed a pronounced seasonal pattern (Fig. 5). This phenomenon has been reported frequently from morphological (Müller et al., 1991; Salbrechter and Arndt, 1994; Carrias et al., 1998; Sonntag et al., 2006) as well as from molecular studies (Behnke et al., 2010; Bielewicz et al., 2011), and has been attributed mainly to food availability and abundance, oxygen concentration, temperature or predation. In principal, knowledge of the ciliate morphotype inventory, their seasonal patterns and the existence of seed-bank OTUs are ideal model conditions to assess the function of the seed-bank ciliates as buffer for environmental changes.

If this was the case, we would expect the following: assuming that the annual morphotype sampling revealed the (near) complete morphotype inventory of planktonic ciliates in Piburgersee, we would assume that a unique pyrosequenced sample should have revealed at least most of these taxa. However, in total 39 out of 59 genera recorded in the annual morphotype sampling (from 16 sampling events) escaped the unique pyrosequencing sampling event (bars in shaded area of Fig. 4C). Thirteen of these genera had the theoretical potential for molecular discovery because their SSU rDNA V4 sequences are available in the ciliate sequence database we used for taxonomic assignments (green bars in Fig. 4C). One reason for their non-detection besides the factors discussed earlier might have been their patchy distribution in the plankton (Montagnes et al., 1999). Another possibility may have been the formation of cysts and subsequent disappearance from the water column due to sedimentation (Müller et al., 2002). For euplanktonic ciliates, encystment has so far only been identified for few species, for example, for *Pelagostrombidium* spp. (Müller, 1996) or *Cytolophosis* sp. (B. Sonntag, pers. obs.).

However, we found a staggering number of 26 additional genera in the annual morphotype sampling but not in the molecular survey (red and blue bars in Fig. 4C). These genera were either not represented in sequence databases or the obtained V4 amplicons could not be assigned confidently to the genus level. One example from our study is *Ctedoctema acanthocryptum*, which had relatively high abundances in the warmer season, i.e. 12 000 ind. l⁻¹ on average in July mainly in 0–9 m depths but was not detected in observations over the whole water column from January to April (B. Sonntag, unpublished). In the NCBI database, no sequence had been deposited so far for the genus *Ctedoctema*. Examples for falsely assigned genera are discussed earlier.

**Conclusions**

Our study pinpoints two major difficulties in molecular protistan ecology research: insufficient database coverage of the described taxa (in this case: ciliates) and the unreli-
able use of the V4 SSU rDNA fragment as barcoding marker. Another lesson learned from our study is that the environmental buffer hypothesis for the seed-bank taxa is very difficult to test through the comparison of morphological and molecular data sets. One alternative would be a frequent deep-sequencing survey with numerous sampling events at different seasons. A high genomic potential to buffer environmental (seasonal) changes would be indicated if the basic OTU composition would remain the same but with changes in amplicon abundances in the different OTUs through seasonal cycles. However, because of relatively high expenses and a largely obscured taxonomy of OTUs, the ecological value of the data sets would be very limited with a poor cost–benefit effect. A better approach would be to intensify barcoding efforts for protists (Pawlowski et al., 2012). The identification of a barcode marker for individual protistan taxon groups, the collection and identification of taxa in specific habitats and the deposition of their barcodes in databases would then allow more powerful ecological studies with molecular tools such as conducted in this study. Such studies are fundamental as they are the baselines to address elementary ecological questions such as whether new abundant taxa occupied the same ecological niches of organisms that at times ‘disappeared’ into the rare biosphere. Even though taxonomic assignment of short sequence reads from Next Generation Sequencing (NGS) strategies is a major impediment, also other issues should be considered. These include, but are not restricted to, the true nature of rare amplicons. Such rare amplicons may reflect seed-bank taxa, sequence and/or PCR artefacts or taxa with extremely low ribosomal RNA operon copies. In-depth studies investigating the relation between abundance and number of rRNA copies in specific ciliate genera may be a promising approach to this subject.

**Experimental procedures**

**Sampling site and sampling**

Piburgersee is an oligo-mesotrophic lake situated in the Austrian Central Alps (47°11′N, 10°53′E). The lake has an area of 13.4 ha, a maximum depth of 24.6 m and is ice covered from December through April. For details on limnochemistry and chlorophyll a see Tolotti and Thies (2002) and Salcher and colleagues (2008).

We sampled the lake on 1 March 2011 by mechanically pecking two holes at a distance of ten meters into the ice cover at the deepest point of the lake. For measurements of abiotic parameters [temperature, oxygen concentration, pH, conductivity, nitrate, ammonium, dissolved nitrogen, total and dissolved phosphorus, sulfate, chloride and dissolved organic carbon (DOC)] and chlorophyll a water was collected
with a 5 l Schindler–Patalas sampler from one of the two ice holes. For ciliate sampling, three depths were chosen based on previous observations showing different assemblages in accordance to the prevailing abiotic and biotic characteristics at this time of the year (B. Sonntag, unpublished). Sampling depths were (i) directly under the ice cover with highest chlorophyll concentrations providing food for algivorous ciliates, (ii) at 9 m depth where typically low nutrients were available for the ciliates and (iii) at 21 m depth where oxygen was strongly depleted at this time of the year, and bacterivorous ciliates prevailed (B. Sonntag, unpublished; Salcher et al., 2010).

For ciliate morphology and quantification, 200 ml subsamples were taken in triplicates and preserved immediately with freshly prepared Bouin’s fixative according to subsamples were taken in triplicates and preserved immediately and thoroughly rinsed 1 l plastic bottles and transported at temperatures to the laboratory of the Institute of Ecology and Viennese University of Innsbruck). For methodological details, see Tolotti and Thies (2002) and Salcher and colleagues (2008).

For DNA extraction, 5 l of raw water were drawn onto 0.65 μm Durapore membranes (Millipore, EMD Millipore Corporation, Billerica, MA, USA) using a peristaltic pump (flow rate ca 50 ml min⁻¹). Filters were immediately put into cryovials with RNA later (Qiagen, Qiagen GmbH, Hilden, Germany) and stored as recommended by the manufacturer until further processing.

For a more detailed comparison of the morphological data set obtained in this study to the actually prevailing ciliate assemblage in Piburgersee, we used a detailed compilation of the ciliates in this study, both living and specifically silver-stained individuals were considered for the morphological investigation. First, the living ciliates that were kept in a refrigerator at 4°C were identified within 2 days after lake sampling. Then, the Bouin’s-fixed ciliate samples were filtered and stained by applying a quantitative protargol stain (Skibbe, 1994; Pfister et al., 1999), revealing the characters necessary for identification such as the nuclei or the specific ciliary patterns. All ciliates were identified to the species level by the use of an Olympus BX50 microscope (Olympus, Vienna, Austria) under differential interference contrast (living cells) or brightfield (preserved cells). The protargol-stained ciliates were also quantified from the obtained permanent slides. Identification followed the keys of Foissner and colleagues (1991; 1992; 1994; 1995; 1999) and references therein.

**Morphological analyses**

For a comprehensive and precise identification and quantification of the ciliates in this study, both living and specifically silver-stained individuals were considered for the morphological investigation. First, the living ciliates that were kept in a refrigerator at 4°C were identified within 2 days after lake sampling. Then, the Bouin’s-fixed ciliate samples were filtered and stained by applying a quantitative protargol stain (Skibbe, 1994; Pfister et al., 1999), revealing the characters necessary for identification such as the nuclei or the specific ciliary patterns. All ciliates were identified to the species level by the use of an Olympus BX50 microscope (Olympus, Vienna, Austria) under differential interference contrast (living cells) or brightfield (preserved cells). The protargol-stained ciliates were also quantified from the obtained permanent slides. Identification followed the keys of Foissner and colleagues (1991; 1992; 1994; 1995; 1999) and references therein.

**DNA isolation, SSU rDNA amplification and sequencing**

DNA was isolated directly from the Durapore membranes using Qiagen’s AllPrep kit according to the manufacturer’s instructions. For each site and depth, three subsamples (filters) each were extracted and pooled. From these extracts, we first amplified a ca 700 bp-long fragment of the SSU rDNA (including the hypervariable V4 region) using a ciliate-specific primer mix (Table 2). The PCR reaction included 50–100 ng of template DNA in a 50 μl reaction, 1 U of Phusion High-Fidelity DNA polymerase (Finzymes, New England Biolabs, Ipswich, MA, USA), 1x Phusion HF Buffer (New England Biolabs, Ipswich, MA, USA), 200 μM of each deoxynucleotide triphosphate and 0.5 μM of each oligonucleotide primer. The PCR protocol consisted of an initial denaturation (30 s at 98°C) followed by 30 identical amplification cycles (denaturation at 98°C for 10 s, annealing at 59°C for 10 s and extension at 72°C for 30 s) and a final extension at 72°C for 10 min. In a subsequent nested PCR reaction using the purified (Qiagen’s MinElute kit) PCR products from the first reaction as template, the V4 region of the SSU rDNA was amplified by using eukaryote V4-specific primers (Table 2) following the protocol of Stoeck and colleagues (2010). The V4 forward primers were tagged with four base pair-long identifiers (Mul-

| Table 2. Primer sets used in this study for the specific amplification of ciliate V4 SSU rRNA fragments using a two-step (nested) PCR reaction. |
| --- | --- | --- |
| **Primer** | **Sequence** | **Reference** |
| Cil F | 5’-TGTTAGTGATATGGACWACCA-3′ | (Lara et al., 2007) |
| Cil R1 | 5’-CCTGAGTGTCTGAATGCGCA-3′ | (Lara et al., 2007) |
| Cil R2 | 5’-TCATGCTGCTTTGTATCCC-3′ | (Lara et al., 2007) |
| Cil R3 | 5’-TCATGCTGCTTTGTATCCC-3′ | (Lara et al., 2007) |

**Abiotic parameters and chlorophyll a**

Temperature was read off directly from a thermometer attached inside the water sampler. All other parameters were measured in the laboratory at the Institute of Ecology (University of Innsbruck). For methodological details, see Tolotti and Thies (2002) and Salcher and colleagues (2008).
tplex Identifier, MID) at the 5′ end. To purify the PCR products, the target bands of a size of ca 450–500 bp were excised from the agarose gels and purified using Qiagen’s gel extraction kit. By choosing two distant sampling locations in the lake, we accounted for potential local patchiness of the protistan plankton. Replicate DNA extractions (three per sample) and PCR reactions (five per DNA template) were taken into account to minimize potential methodological bias. Prior to emulsion PCR and sequencing, the quantity and integrity of the PCR products were checked with a DNA 7500 Assay (Agilent Bioanalyzer 2100, Agilent Technologies, Waldbronn, Germany). Quantification, single pool amplicon titration, emulsion and pyrosequencing (1/4 plate for three samples) were conducted at Engencore (Columbia, SC, USA) using the standard pyrosequencing protocol (titanium chemistry) on a Genome Sequencer FLX system (Roche, Roche Applied Science, Mannheim, Germany). All V4 sequences (tags) were sequenced from the forward primer (5′ end).

**Amplicon data processing and taxonomic assignments**

Homopolymer denoising of raw sequences was conducted with Acacia (Bragg et al., 2012). Denoised sequences were then further processed with the QIIME software package (Caporaso et al., 2010). Only tags meeting the following criteria were considered for further analyses: (i) containing bases A, C, G or T, (ii) containing the complete and correct forward primer and (iii) having a minimum length of 300 bp after primer removal. Chimeras were identified and removed by using UCHIME (Team, 2006; Edgar et al., 2011). The phylotypes were clustered with Uclust (Edgar, 2010) at different sequence similarities (100–90%). Rank-abundance and length distribution graphs of the tags were constructed in R (Team, 2006).

Phylotypes that clustered at 97% sequence similarity were used for taxonomic assignments and diversity statistics (Nebel et al., 2010; Behnke et al., 2011; Dunthorn et al., 2012). For taxonomic assignments and the analysis of a potentially novel diversity, one representative sequence from each phylotype at 97% was extracted (corresponding to the longest sequence in the cluster), and the data were combined into a fasta file. For taxonomic assignments, this file was analysed with the software package JAGUC (Nebel et al., 2011). As reference database, we maintained a manually curated V4-sequence data collection from so far sequenced ciliates (Nebel et al., 2012). Accession numbers are cited in the publication. JAGUC employs BLASTn searches, with algorithm parameters adjusted for short (200–500 bp) reads (-m 7 -r 5 -q -4 -G 8 -E 6 -b 50). For details on the search algorithm and the taxon assignment strategy, see Nebel et al. (2011). As described previously (Stoeck et al., 2010), we extracted all sequences with the closest BLAST hit above 85% sequence similarity into a reference database entry and designated tags below 85% as ‘unassigned’. A reliable genus assignment for ciliates based on 95% sequence similarity in the V4 region of the SSU rDNA has been demonstrated in previous studies (Nebel et al., 2010; Dunthorn et al., 2012). Therefore, to enable comparisons with morphospecies data, relevant tags that showed at least 95% sequence similarity to database entries were considered in downstream statistical diversity analyses for comparisons with morphodata. Tags with similarities between 95% and 85% to reference sequences were also considered but not assigned to specific taxon levels. An abundance table was created from the JAGUC output file (taxon assignments) with the QIIME OTU output file considering clusters built at 97% sequence similarity cluster threshold. This table contained information about the number of phylotypes in each sample, their distribution among the three depths, the number of phylotype tags in each of the three samples, the taxonomic assignment of each phylotype, the name of the representative sequence of each phylotype and the sequence similarity of the representative tag to the closest BLASTn match in the reference database. This file finally served as the basis for α- and β-diversity statistics.

**Statistical analyses**

We used the vegan package in R (Oksanen et al., 2011) to calculate the diversity among the three depths using the Sørensen index, and the Shannon index to compare α-diversity. The Fisher’s exact test (Fisher, 1922) was used to test whether the null hypothesis (the taxon distribution in molecular and morphological data sets is the same) can be accepted. Rarefaction analyses were conducted in QIIME.

**Acknowledgements**

We thank Claudia Grubbauer for help in the laboratory (quantitative protargol stain preparations), Josef Franzoi, Salvador Morales-Gomez and Gry Larsen (Institute of Ecology, University of Innsbruck) who sampled and analysed the abiotic parameters and chlorophyll of Piburgersee. Lucie Bittner (Department of Ecology, University of Kaiserslautern) is acknowledged for writing and providing the scripts for the statistical analyses. This study was funded by the Austrian Science Fund (FWF): projects P21013-B03 (PI B. Sonntag), P16559-B06 (PI R. Sommeruga, the unpublished Data set of the annual ciliate cycle in Piburgersee was collected and compiled by B. Sonntag in the course of this project) and by a grant from the Deutsche Forschungsgemeinschaft (DFG) to Thorsten Stoeck (STO414/3-2). The authors declare that they have no conflicts of interest.

**References**


**Supporting Information**

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

**Fig. S1.** Rank-abundance distribution of amplicon data sets from the three different depths screened for ciliate phylotype diversity in Piburgersee in March 2011. The distribution follows a long-tail distribution, typically for large environmental amplicon data sets.

**Fig. S2.** Log-regression of relative abundances of amplicon ciliate genera and morphotype ciliate genera. Abundance distributions are incongruent and do not correlate, suggesting that amplicon read abundances do not reflect morphotaxon (organismic) abundances.

**Table S1.** ‘Candidatus ciliate genera’ that were detected through V4 SSU rDNA amplicons in data sets from the three different depths screened for ciliate phylotype diversity in Piburgersee in March 2011. Amplicons were assigned to a ‘candidatus genus’ status when the maximum sequence similarity to a reference sequence database of described ciliate morphotypes did not exceed 95%.

**Table S2.** Ciliate genera and number of species (in brackets) detected in Piburgersee on 1 March, 2011 in order of relative frequency (mean abundance in ind. l$^{-1}$ over the three sampling depths) and abundance in the respective depth. $n = 6$ (0 m, 9 m), $n = 3$ (21 m), n.d., not detected.