High resolution analysis of local and global protistan soil community structures using morphological and modern molecular techniques

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Anne Kathrin Domonell

aus Bergisch-Gladbach

Köln 2013

Berichterstatter:	Prof. Dr. Hartmut Arndt
(Gutachter)	Prof. Dr. Michael Bonkowski

Tag der mündlichen Prüfung: 18.10.2013

Danksagung

Zuallererst möchte ich mich bei meinem Betreuer Prof. Hartmut Arndt bedanken, der mir die Möglichkeit zur Promotion gegeben hat und mir während meiner Dissertation mit wissenschaftlichen Diskussionen und Anregungen zur Seite stand. Außerdem danke ich ihm, dass er mir die Möglichkeit gab an einer Forschungsreise nach Grönland und zahlreichen Tagungen teilzunehmen.

Prof. Michael Bonkowski danke ich für die Zweitbegutachtung meiner Dissertation und seine Hilfe bei bodenökologischen und statistischen Fragen.

Prof. Frank Schäbitz danke ich für die Übernahme des Vorsitzes meiner Disputation.

Der deutschen Forschungsgemeinschaft (DFG) und Hartmut Arndt möchte ich für die Finanzierung meiner Stelle und die im Rahmen dieser Arbeit benötigten Sachmittel danken.

Großen Dank möchte ich Frank Nitsche aussprechen, der mir in den schwierigen Zeiten der molekularbiologischen und bioinformatischen Analysen stets engagiert zur Seite stand. Seine fachlichen Ratschläge und seine Geduld waren mir überaus hilfreich.

Dank auch an die gesamte Arbeitsgruppe Arndt, vor allem Bärbel Jendral, Rosita Bieg und Anja Schwerwaß für die Hilfe bei adminstrativen und technischen Problemen und die schöne gemeinsame Zeit.

Des Weiteren möchte ich die gute Zusammenarbeit mit Marian Brabender in Bezug auf die Cercomonaden-Beschreibungen nicht unerwähnt lassen.

Meinen lieben Kollegen Alex, Jana, Mar, Kathi und Timm, danke ich für ihre Freundschaft und moralische Unterstützung. Danke für die schöne gemeinsame Zeit und die anregenden fachlichen und privaten Gespräche. Es war gut euch dabei gehabt zu haben.

Besonderer Dank gilt meinen Eltern und meinem Bruder, die mir während und besonders zum Ende der Dissertation eine unersetzbare Stütze waren. Ich danke euch für diesen Rückhalt und euer Vertrauen in mich.

…

„Alles Wissen und alle Vermehrung unseres Wissens endet nicht mit einem Schlusspunkt, sondern mit Fragezeichen.“

Hermann Hesse

…

Table of Contents

Zusammenfassung

Protisten sind einzellige Eukaryoten, die verschiedenste aquatische und terrestrische Lebensräume besiedeln. In der vorliegenden Arbeit wurden vor allem Protistengemeinschaften in terrestrischen Habitaten untersucht. Als Hauptkonsumenten von Bakterien im Boden tragen sie dort maßgeblich zu Mineralisierungsprozessen und bedeutsamen Ökosystemprozessen bei. Protistengemeinschaften verschiedener Grünland- und Waldböden mit unterschiedlicher Landnutzung wurden mittels morphologischer und molekularer Methoden untersucht. Die morphologischen Untersuchungen konnten dabei hohe Abundanzen an Protozoen (heterotrophe Protisten), beeinflusst durch die Landnutzung, in den untersuchten Böden aufdecken. Zudem waren erste Einblicke in die Zusammensetzung der Protozoengemeinschaften der Böden möglich, die vor allem durch Flagellaten und nackte Amöben dominiert wurden. Cercomonaden, eine der diversesten und abundantesten Flagellaten der Studie, wurden mittels weiterer morphologischer Bestimmungen und phylogenetischer Analysen untersucht. Die Ergebnisse der morphologischen Studien begründeten eine weitere detaillierte Analyse der Protisten im Boden und stellten zugleich die Basis für weitere molekulare Analysen dar. Die morphologisch untersuchten Böden wurden mittels next generation sequencing (NGS) der V4-Region der SSU rDNA vertiefend analysiert. Durch das eingeführte Verfahren zur Analyse der NGS-Daten, konnten sehr detaillierte und umfangreiche Untersuchungen der Protistengemeinschaft durchgeführt werden. Zum ersten Mal wurden so Bodengemeinschaften bis auf das Artniveau hin untersucht. Die Kombination von morphologischen und molekularen Methoden zur Untersuchung von Gemeinschaften von Flagellaten ist einzigartig. Verschiedene isolierte und neubeschriebene Arten wurden mittels NGS wieder aufgefunden und deuten darauf hin, dass einige abundante Arten bisher möglicherweise übersehen wurden. Vertreter der SAR Gruppe (Stramenopiles, Alveolata und Rhizaria) dominierten die mittels NGS analysierten Protistengemeinschaften im Boden. Ein Landnutzungseffekt zeigte sich vor allem im Vorkommen bestimmter Thekaamöben und Ciliaten. Methodische Untersuchungen der NGS-Technik zeigten, dass sich Schlussfolgerungen über das Vorkommen von Arten ziehen

lassen, Aussagen über Abundanzen jedoch nur vorsichtig getroffen werden sollten.

Weitere Proben aus der Arktis, Antarktis und der USA wurden untersucht und zeigten eine größere Abhängigkeit der Protistengemeinschaft hinsichtlich des Habitats in dem sie leben, als hinsichtlich der geographischen Region, aus der sie stammen. Die Untersuchungen von Luft-, Schnee-, Eis- und Bodenproben eines anthropogen wenig beeinflussten Ortes in Grönland unterstützten, zusammen mit den unterschiedlichen globalen Proben, die Annahme einer unbegrenzten Luftverbreitung von Protisten.

Abstract

The unicellular protists inhabit all types of aquatic and terrestrial environments. This thesis is focused on terrestrial protistan communities where they contribute significantly to mineralization processes and key ecosystem functions, as they are major consumers of bacterial production. The protistan soil communities of grassland and forest sites with a different land use were morphologically and molecularly investigated in this study. The morphological investigations revealed high protozoan abundances influenced by the intensity of land use. First insights to the community composition, dominated by flagellates and amoebae could be made, giving the basis for further molecular investigations. Cercomonads were the most abundant and diverse flagellates and were analyzed in more detail, resulting in new description of species. Next generation sequencing (NGS) of the V4 SSU rDNA on the same soil sites was applied. The development of an analysis pipeline for NGS data made in this study, led to a high resolution of the protistan community. This resulted in the first detailed study down to species level of total protistan community structures in soil. Several isolated and newly described species could be recovered by NGS in various samples, indicating that several very common protozoan species in soil have been overlooked. A combination of morphological and molecular studies of soil protistan flagellate communities is unique. Members of the SAR group dominated the protistan communities and land use effects by changes of testate amoebae and ciliates could be registered. Further investigations of Arctic, Antarctic and USA samples, confirmed protistan communities to be dependent on the habitat, but independent on the geographic region. The idea of unrestricted global distribution by air was supported by several genotypes reported from a variety of globally distant sites. The assumed pathways by air were supported by air, ice, snow and soil samples of a low anthropogenic influenced site in Greenland.

General Introduction

Protists are unicellular organisms, compromising autotrophic, heterotrophic and mixotrophic eukaryotes (Hausmann et al. 2003). The organisms are small sized, mainly between 2 and 200µm, and may occur in high abundances up to 10^4-10^7 active individuals per gram dry soil and litter (e.g. Adl and Gupta 2006). Their ability to form cysts make them highly adaptive to environmental changes, allowing them to outlive various harsh conditions (e.g. Rivera et al. 1992). Protists are globally distributed and inhabit all types of aquatic and terrestrial environments (e.g.Ekelund and Ronn 1994; Finlay et al. 2000; Sherr and Sherr 2002).

This thesis is focused on terrestrial protistan communities where they contribute significantly to mineralization processes and key ecosystem functions (Bonkowski and Clarholm 2012; Darbyshire et al. 1974; Finlay et al. 2000; Ronn et al. 2012). Protists are major consumers of bacterial production and link primary producers with higher trophic levels (Bonkowski 2004; Bonkowski and Brandt 2002; Clarholm 1985).

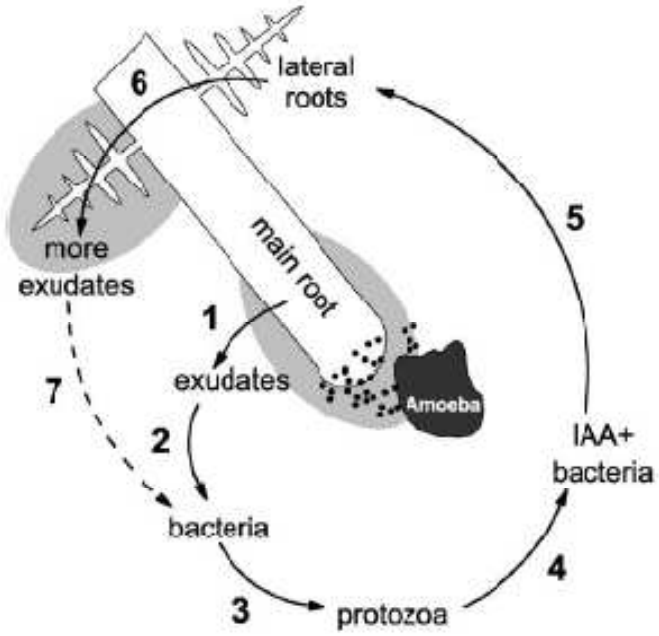


Figure 1. Connection between bacterial grazing and root growth. Exudates are released by the roots (1) and stimulate bacterial growth (2) and protozoan grazing (3). The selective bacterial grazing favor IAA-producing bacteria (producing plant growth promoting hormones) (4) and subsequently increase roots growth (5) and by this the rates of released exudates (6) (from Bonkowski and Brandt 2002).

As illustrated in Figure 1, their grazing has been shown to stimulate plant growth and bacterial metabolism, respectively (Adl and Gupta 2006; Bonkowski 2004; Bonkowski and Brandt 2002; Kreuzer et al. 2004). The selective grazing of protozoans on bacteria has a strong influence on the bacterial community composition and by this on the bacterial-plant interactions (Glücksman et al. 2010; Rosenberg et al. 2009). Additional effects on bacterial functional gene expression due to protistan grazing are indicated (Rosenberg et al. 2009). Plant growth is mainly limited by N availability (Mokhele et al. 2012) and microbial growth mainly by C availability (Jones et al. 2009; Paterson 2003). The availability of both is enhanced by protozoan grazing (e.g. Bonkowski 2004; Bonkowski and Brandt 2002; Clarholm 1985). The importance of the interaction between abuscular Mycorrhiza and grazing protozoans for plant growth is indicated by an increased uptake of N via abuscular Mycorrhiza and an increased C release from the plants to the soil (Herdler et al. 2008; Koller et al. 2013; Wamberg et al. 2003). Detailed analysis of protistan soil communities are required to enlarge the knowledge of protists and the understanding of the microbial interactions in soil. Other studies based on next generation sequencing (NGS) concentrated on certain protistan taxa or identified the protistan community only to a very rough taxonomic level (Bates et al. 2013; Kamono et al. 2013; Triado-Margarit and Casamayor 2012; Urich et al. 2011). The aim of the thesis was a high resolution analysis of protistan soil community structures, based on morphological (Chapters 1 and 2) and molecular analysis (Chapters 4 and 5). NGS was applied for comprehensive analysis of protistan communities down to species level.

In Chapter 1, the protozoan (heterotrophic protists) soil communities of different grassland and forest sites of Thuringia were investigated by cultivation and morphological investigation. The analyses were carried out in the framework of the Biodiversity Exploratories. Among the community structures, the influence of habitat type and land use intensity on the protozoan community structure was revealed. The ability of protists to colonize and reproduce in new habitats is mainly defined by the given conditions, like soil moisture, temperature and food abundance (Adl and Coleman 2005). The dependency on those conditions, their contribution on nutrient cycling processes (Bonkowski 2004) and their short generation time, high abundances and consumption (Foissner 1999b) make them

very suitable as indicators of abiotic and biotic changes. Thereby environmental changes are reflected by differences in the protistan abundance, community structure and diversity, as certain species are adapted to different environmental conditions, where they are metabolically active and reproduce. Traditionally ciliates and testate amoebae were widely used as soil bioindicators, mainly due to their larger body size and easier identification (Foissner 1999b). The present study was intended to increase the general understanding how other protistan groups such as small forms of flagellates, naked and testate filose amoebae and small ciliates contribute to microbial soil communities. Cercozoans and especially cercomonads, were the most diverse and abundant flagellates in the soil samples, which was supported by other studies (Bass et al. 2009; Ekelund and Patterson 1997; Foissner 1991). Their metabolic cell shape, the possession of flagella and the ability to form cysts within a short period of time, make them very suitable to survive extreme conditions. The identification of cercomonads is difficult because of their highly variable cell shape. Nevertheless, the biflagellated gliding amoeboflagellates are up to now mainly determined based on their morphology (e.g. Mylnikov and Karpov, 2004). The development of molecular techniques enabled further work on the taxonomic position of cercomonads. A combination of morphological investigations, including detailed ultrastructural studies, with phylogenetic analysis is required to determine these highly metabolic flagellate and was applied in the studies of Chapter 2.

The morphological analyses of Chapters 1 and 2 gave first ideas regarding the protozoan structure in different soil communities and additionally provide the basis for further detailed analysis. A preliminary knowledge about the species and their abundances representing the community of a certain habitat, are necessary for a successful study design of culture independent techniques. Morphological studies on protists are generally restricted by the low abundances of certain species, encysted resting stages and methodological biases based on selective determination and enrichment (Weisse 2008). The use of high-throughput methods can evade these morphological determination restrictions and additionally provide larger datasets, which enables a highly resolved analysis of e.g. species richness and diversity (Buee et al. 2009; Gottel et al. 2011; Jumpponen and Jones 2009; Lecroq et al. 2011; Nacke et al. 2011; Stoeck et al.

2010; Stoeck et al. 2009; Triadó-Margarit and Casamayor 2013). In Chapter 4, NGS was applied on the preliminary investigated sites of grassland and forest, to improve the influence of land use intensity and site type on protistan soil, giving the first detailed study of protistan communities in soil. The higher resolution of protistan species richness, due to the large NGS dataset, has the potential to enlarge and refine the assumptions made by the morphological investigations made in Chapter 1.

As NGS is applied more frequently and the data yielded by this method accumulate, new questions arise on the application fields as well as the data reliability of these methods (Edgcomb et al. 2011; Lee et al. 2012; Medinger et al. 2010; Weber and Pawlowski 2013). Sequencing errors of individual sequencing reads can result in misidentification and an overestimation of diversity (Huse et al. 2010; Kunin et al. 2010). Therefore, a prerequisite of the proper interpretation of NGS data from field studies were methodological studies. In Chapter 3 the influence of protozoan community composition, based on the species diversity and species richness on the NGS outcome of a given artificially composed community was investigated. Furthermore the influence of PCR amplification was considered in this study. The influence of NGS on the recovery of artificial constructed protozoan communities was applied to verify the results based on NGS made in Chapter 4 and 5.

The understanding of protistan biodiversity and their biogeography is a current issue of ecological studies (e.g. Nolte et al. 2010; Scheckenbach et al. 2010; Stoeck et al. 2010). Different aspects support the assumption of microbial potential of global and long distance dispersal, i. the ability to form cysts (e.g. Rivera et al. 1992), ii. the low weight and small body/cyst size facilitating distribution via air by wind as single cysts or associated to dust particles and other organisms (Rivera et al. 1992; Rogerson and Detwiler 1999; Wilkinson 2001) and iii. the large population sizes increasing the chance for individuals dispersing (Wilkinson et al. 2012). Thus, geographical limitations of dispersal are absent in microorganism in contrary to larger organisms, which have to overcome vertical and horizontal boundaries, such as mountains and seas.

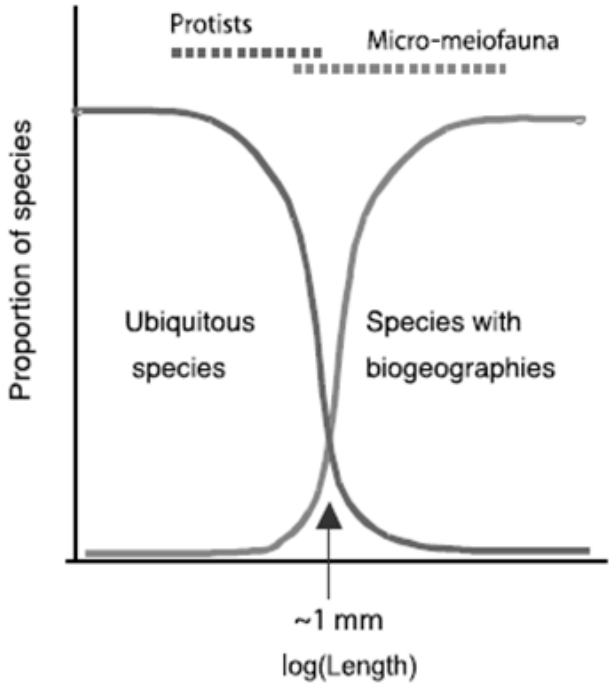


Figure 2. Hypothetical model of the ubiquity-biogeography transition, which is believed to be located in the species size range 1 to 10 mm, as large mammal species obviously have biogeographies (from Finaly 2002).

The dependency of organism size to unrestricted dispersal is indicated in Figure 2, where 1mm borders organism's ubiquity. Recent investigations even suggest a size limit of as low as 20µm for global dispersal (Wilkinson et al., 2012). In the study of Chapter 5 aims at applying molecular methods to give first insights into protist dispersal between air, ice, snow and soil using a low anthropogenic influenced model site in Greenland and to compare protist species richness of globally distant sites.

Chapter 1 – Community structure of cultivable protists in different grassland and forest soils of Thuringia

Introduction

Protozoa are ubiquitous and essential components in terrestrial ecosystems. They are major consumers of bacterial production and link primary producers with higher trophic levels (Clarholm 1985; Bonkowski 2004). Protozoan grazing has been shown to stimulate plant growth and bacterial metabolism, respectively (Adl and Gupta 2006; Bonkowski 2004; Bonkowski and Brandt 2002; Ekelund and Ronn 1994; Kreuzer et al. 2004). Therefore grazing of bacteria by soil protozoans plays an important role in mineralization and cycling of nutrients in soil food webs (Finlay et al. 2000).

Up to now, data on the taxonomic composition of protozoans in soil are very scarce. A comprehensive study by Finlay et al. (2000) at the Sourhope experimental grassland site showed average densities of $45*10^3$ flagellates g^{-1} dry weight, $17*10^3$ naked amoebae g^{-1} dry weight, $3.5*10^3$ testate amoebae g^{-1} dry weight and $5*10^3$ ciliates g^{-1} dry weight. Although flagellates and naked amoebae appear to be the most abundant protozoans in soil and constitute an important part of soil food webs, testate amoebae and ciliates traditionally have gained more attention in taxonomic studies due to their larger size, but mostly to their distinct morphology that made species identification easier (Foissner 1987a, 1987b; Foissner et al. 1990). For biodiversity studies, more detailed analyses including the abundances of all protozoan groups are needed. There are only very few data available which combine quantitative and qualitative studies especially regarding the abundant nanoprotists in the size range of 1-20 µm (Arndt et al. 2000). This is a major problem for understanding of the role of protozoans in respect to the flux of matter in soil.

The aim of the study was to contribute to the above mentioned gap of knowledge regarding the diversity and abundance of nanosized protozoans in soil. The protozoan communities of eight different sampling sites including four grassland and four forest soils with variable land use were compared. Higher abundances at intensively used sites and a significant difference in community structure of grassland and forest soils were expected. The abundances of flagellates and naked amoebae were determined with a modification of the liquid aliquot method (Butler and Rogerson 1995) and combined this with a rough classification of these

nanoprotists into morphotypes via light microscopy. The function of nanofauna soil communities should be derived from the obtained knowledge of the community structure and published data on the feeding ecology of the recorded protozoan taxa.

Material and Methods

Study site and soil sampling

The study area was located in central Germany (N 51°13'0'', E 10°27'0''), in the Hainich-Dün region (Thuringia, Germany Fig. 1), a limestone plateau (300-400 m a.s.l.) with an annual mean temperature of 6.5-7.5 °C and annual mean precipitation of 750-800 mm (Fischer et al. 2010). The experimental plots were extremes of a land use gradient in established land use systems as part of an interdisciplinary project, the Biodiversity Exploratories (http://www.biodiversity-exploratories.de).

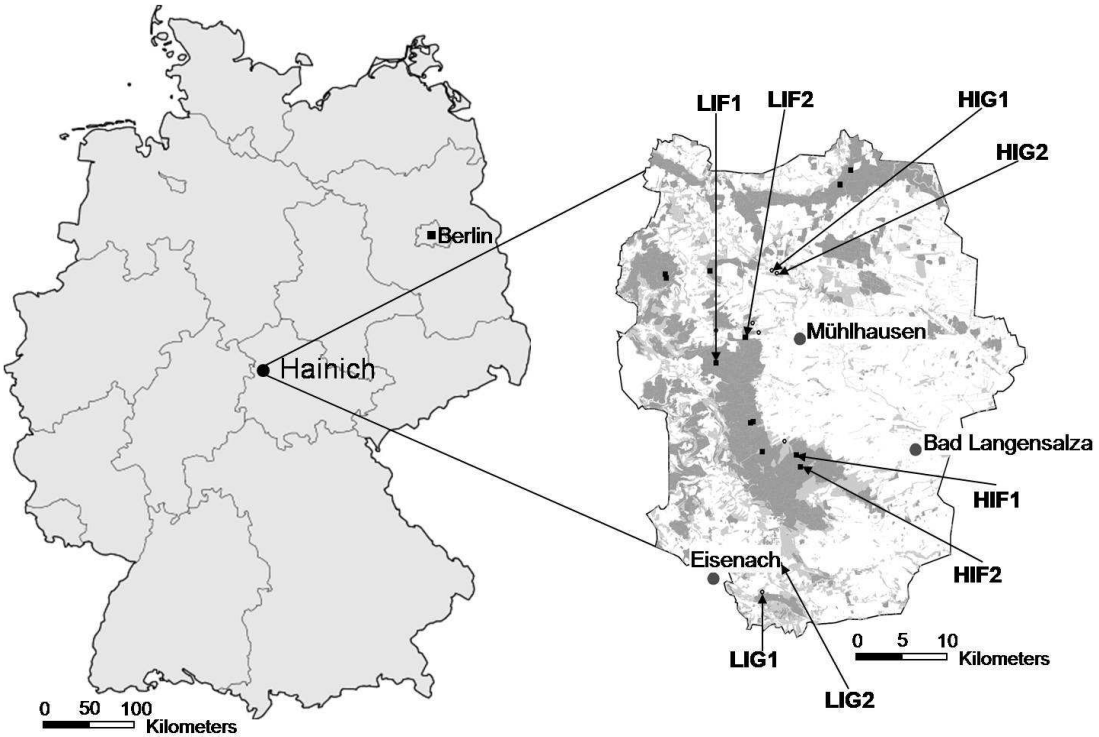


Figure 1. Study area with sampling sites of biodiversity project across Germany, with a closer look to National Park of Hainich-Dün. Explored grassland areas of Hainich-Dün were marked with circles and forest areas were marked with squares. (http://www.biodiversity-exploratories.de).

A comparison of the protozoan community structure at different plots (four grassland sites and four forest sites) chosen according to their degree of land use (for details see Table 1) was carried out. Grassland sampling sites comprised two fertilized low intensively used managed grasslands, mown three times per year (referred as LIG 1 and LIG 2) and two unfertilized pastures grazed by cattle with intensive land use (referred as HIG 1 and HIG 2). The forest sampling sites included two low intensity unmanaged beech forests at late succession (referred as LIF 1 and LIF 2) and two intensive managed spruce age class forests at early succession (referred as HIF 1 and HIF 2).

Table 1. List of sampling sites in grassland (LIG and HIG) and forests (LIF and HIF). Dataset obtained from the Biodiversity Exploratories (www.biodiversity-exploratories.de).

Plot	Soil type	Land use	Soil water content [%]	RW	HW
LIG 1	Brown soil	Fertilized meadow (mown 3 times)	37.5	4388100	5649700
LIG 2	Pelosol	Fertilized meadow (mown 3 times)	23.0	4390100	5652600
HIG 1	Pseudogley	Unfertilized cattle pasture	37.8	4389180	5683280
HIG 2	Pseudogley	Unfertilized cattle pasture	44.9	4389700	5683010
LIF 1	Parabrown soil	Beech Plenter forest, very late succession	37.0	4396800	5692270
LIF 2	Parabrown soil	Unfarmed beech forest, very late succession	39.0	4388100	5664300
HIF 1	Pseudogley	Sprunce forest, early succession	33.1	4382900	5673600
HIF 2	Parabrown soil	Sprunce forest, early succession	31.1	4386200	5676270

The samples were collected during the 14th and 30th of April in 2008 in the Hainich-Dün National Park (Fig. 1). Grassland plots and forest plots were 20m x 20m in size. At each sampling plot nine subsamples were taken. For each subsample a metal frame (20cm x 20cm) was driven into the soil. After removing the root-layer of the top 5cm of grassland samples and the litter-layer of the forest samples, the deadwood and roots with a diameter > 2cm were rejected. The A-level of the nine subsamples was collected and thoroughly mixed to one sample to reduce small-scale heterogeneity. The samples were stored at field moisture content at 4℃ until

the analysis started. Soil water content was determined by subtraction of soil dry weight from soil fresh weight, after drying at 105°C for 24h.

Determination of cultivable protozoans

For quantification of cultivable protozoans the liquid aliquot method (LAM) according to Butler and Rogerson (1995)was employed which is a modification of the classical MPN method (Darbyshire et al. 1974). The LAM method bears the advantage that the inoculum per culture vessel is so small that maximal one cultivable protist is contained. Thus, potentially selective further dilutions are not necessary and the establishment of clonal cultures for later identification is easier. Pre-experiments were carried out to apply adequate dilution rates. One g of homogenized soil sample was suspended for 10 min by shaking in 200 ml Volvic© mineral water (Danone Waters Deutschland GmbH, Frankfurt, Germany) to dilute and separate the organisms from the soil. 5 µl aliquots of this suspension were filled into wells of six 24-well plates (Sarstedt, Nümbrecht, Germany), each well contained 1.5 ml Volvic© water and one sterilized amaranth (*Amaranthus caudatus*) grain to stimulate bacterial growth. The plates were checked weekly for a period of four weeks using an inverted microscope (Zeiss Axiovert S100) with phase contrast optics. For better identification and later sequencing, some protists were transferred into 50ml tissue culture flasks (Sarstedt, Nümbrecht, Germany) filled with 30ml WC-medium, a freshwater medium of double-distilled water supplied with inorganic salts and vitamins (Guillard and Lorenzen 1972) and one sterilized quinoa (*Chenopodium quinoa*) grain. The 24-well plates and culture flasks were stored at 14℃ in the dark. To maintain sterile conditions, the whole work was conducted at a clean bench.

Morphotypes of protozoans were determined according to keys and figures published by Lee et al. (2000) and Smirnov and Goodkov (1999). Systematics followed Cavalier-Smith and Chao (2003) and Howe et al. (2009). Protozoan species were determined to the morphotype level or summed up to morphotype-groups, based on their similar and variable morphology as follows: (i) glissomonads: including species of bodomorphids and thamatomonads; (ii) cercomonad morphotype-group A: bulb shaped cercomonads with moderate

metabolic cell shape in a size range of 12-18µm possessing a long anterior flagellum; (iii) cercomonad morphotype-group B: cercomonads with a metabolic cell shape in the size of 5-11µm; (iv) cercomonad morphotype-group C: cercomonads with a very metabolic cell shape in the size range of 15-20µm. All aboved mentioned cercomonads as well as other cercomonads and glissomonads were summarized under the term cercozoans; (v) euglenozoans: bodonids and euglenids.

Total protozoan numbers were calculated from the cumulative appearance of protists in the 24-well plates and corrected by a modified Poisson distribution.

Statistical analyses

All statistical tests were performed using the SAS® software (SAS Institute Inc., Cary, NY, USA, 1993). Data were analyzed by analysis of variance (ANOVA). Data of percentage contributions were arcsin-transformed to approximate homogeneity of variance.

A cluster analysis (Bray Curtis similarity) in respect to the abundance and community structure was performed with the program Primer 6 (Clarke and Gorley 2006).

The diversity of the cultivable protozoan morphotypes was calculated by the Shannon-Weaver index (Shannon and Weaver 1949):

$$H = -\sum_i p_i \cdot \ln p_i$$

where P_i is the proportion of the *i*-th morphotype.

To calculate the evenness of the cultivable protozoan morphotypes, the Simpson index (D) was calculated (Simpson 1949):

$$D = 1 - \sum_{i=1}^{S} \frac{n_i(n_i - 1)}{n(n-1)}$$

where n_i is the number of individuals of one morphospecies and n is the total number of individuals.

Results

Comparison of protozoan community structure

For an overall comparison, all four subsamples of a respective landscape type (grassland/forest, extensive/intensive land use) were summarized in Figure 2 to simplify the comparison and to clarify general differences between these different land types. The percentage contribution of the protozoan groups did not differ significantly between the site types (ANOVA, P>0.05, see Appx. Tab. 1.1). The comparison of the protozoan community concerning different land use intensities revealed higher numbers of amoebae and significantly lower numbers of cercozoans ($F_{[3,7]}$= 17.03; P=0.015 for land use) in intensively used sites. A significant influence of the intensity of land use on any other protozoan group could not be confirmed (ANOVA, P>0.05, see Appx. Tab. 1.1).

Variability of the protozoan community across different grassland and forest sites

On average, forest sites showed slightly lower abundances (49.7 $*10^3$ ± 8.9 ind. g^{-1} dry weight) compared to grassland sites (69.3 $*10^3$ ± 48.5 ind. g^{-1} dry weight, see Table 2). The relative contribution of each protozoan group was similar at grassland and forest sites. Flagellates and amoebae each had a similar share of about 50% of the abundance at all site types. Total numbers of protozoans differed significantly between extensive and intensive grassland sites ($F_{[3,7]}$= 9.39; P= 0.038 for the interaction site type x land use), mainly caused by variability of amoebae. Total numbers of amoebae increased 5.4 fold from extensively to intensively managed grasslands ($F_{[4,7]}$= 12.12; P= 0.025 for the interaction site type x land use), but no significant effect of land use on amoebae in forests could be detected. Monotactic ($F_{[3,7]}$= 4.68; P=0.097) and fan-shaped ($F_{[3,7]}$= 10.52; P=0.032) amoebae were mainly responsible for the increase in abundance of protozoans in intensively managed grassland. Among flagellates, cercomonads dominated followed by euglenozoans (mainly bodonids) and stramenopiles. Apusomonads and ciliates were of minor importance.

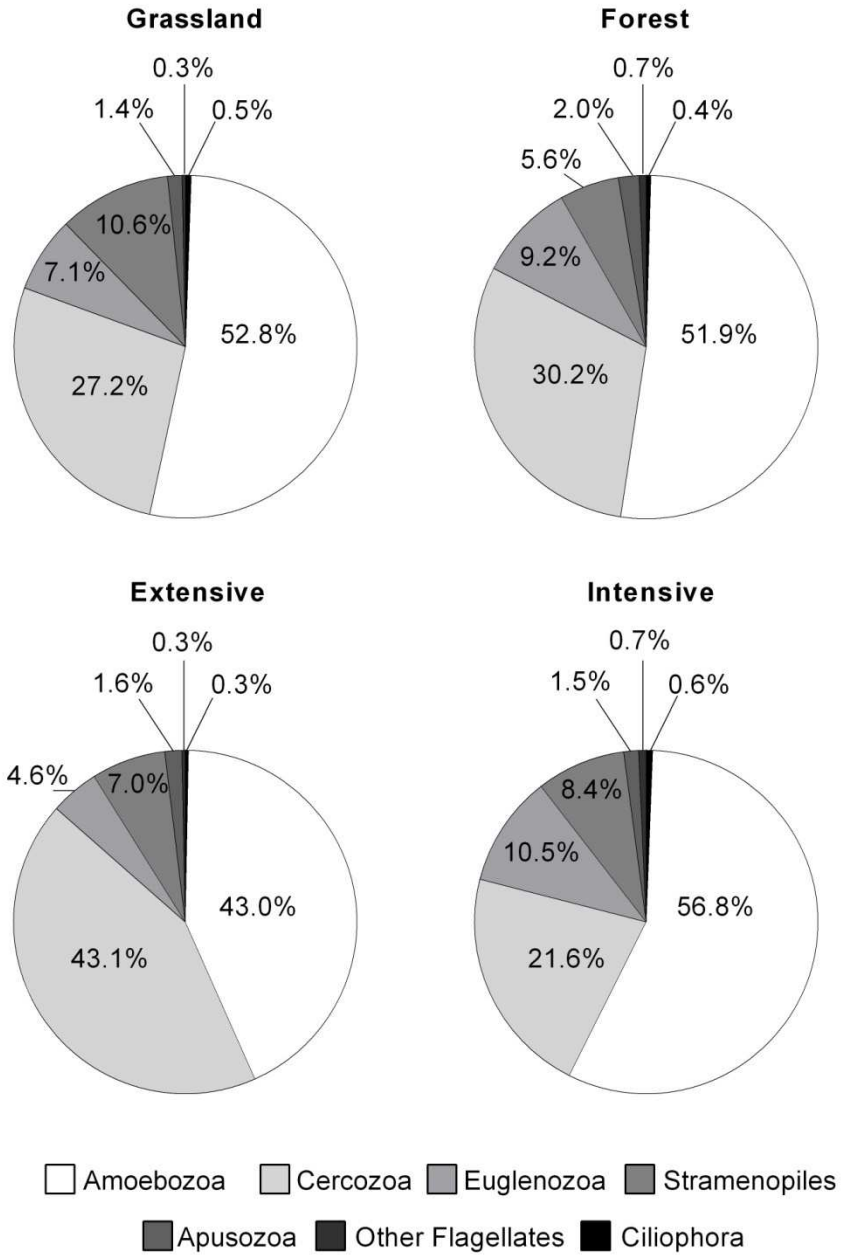


Figure 2. Comparison of the composition of cultivable protozoans at grassland and forest sites and at extensive and intensive used sites (Percentage contribution in brackets).

The two grassland sites with a more intensive land use (HIG 1 and 2) had highest protozoan abundances (88.3 and 127.3 *103 ind. g^{-1} dry weight) both, in comparison to the other grassland sites and to the forest sites (Table 2). The four forest sites had similar abundances ranging between 41.0 and 61.2 *103 ind. g^{-1} dry weight. The extensively used grassland sites (LIG 1 and 2) had slightly lower abundances of 44.3 and 17.2 *103 ind. g^{-1} dry weight (Fig. 3a). While the protozoan community was dominated by flagellates (35 - 65%) and amoebae (32 -

65%) at each sampling site, ciliates were only rarely recorded (0.0 – 1.6 %). Though the general community structure of protozoans was very similar in grassland and forest sites, a higher resolution of morphotypes revealed a high variability between sites (Fig. 3b). Only cercomonad morphotype A was significantly correlated with site type and land use intensity, all other protozoan morphotypes had no significant relationship neither to the site type nor to the land use intensity (see Appx. Tab. 1.3).

The most abundant groups of cultivable soil Protozoa were cercozoans, euglenozoans (especially bodonids) and various amoebae (Fig. 3b). Within the groups of amoebae and cercozoans just slight differences between grassland and forest sites could be detected. At all sites, the amoebae community was dominated by fan-shaped taxa (mostly vannellids) followed by other naked amoebae morphotypes. Testate amoebae were of minor importance with respect to their abundance. Representatives of the *Neobodo designis*–complex were the most important euglenids at all sites followed by representatives of the *Bodo saltans*-complex which tended to have higher abundances at forest sites. Among cercozoans, glissomonads contributed the highest proportion, followed by other cercomonad morphotypes (Fig. 3b).

The cluster analysis indicated a significant difference between the grassland sites LIG 2 and HIG 2 and both sites compared to the other grassland and forest sites ($P<0.05$). All other sites were not significantly different from each other. There were no significant differences between the diversities of grassland and forest sites (Shannon-Weaver $P>0.05$, Simpson $P>0.05$).

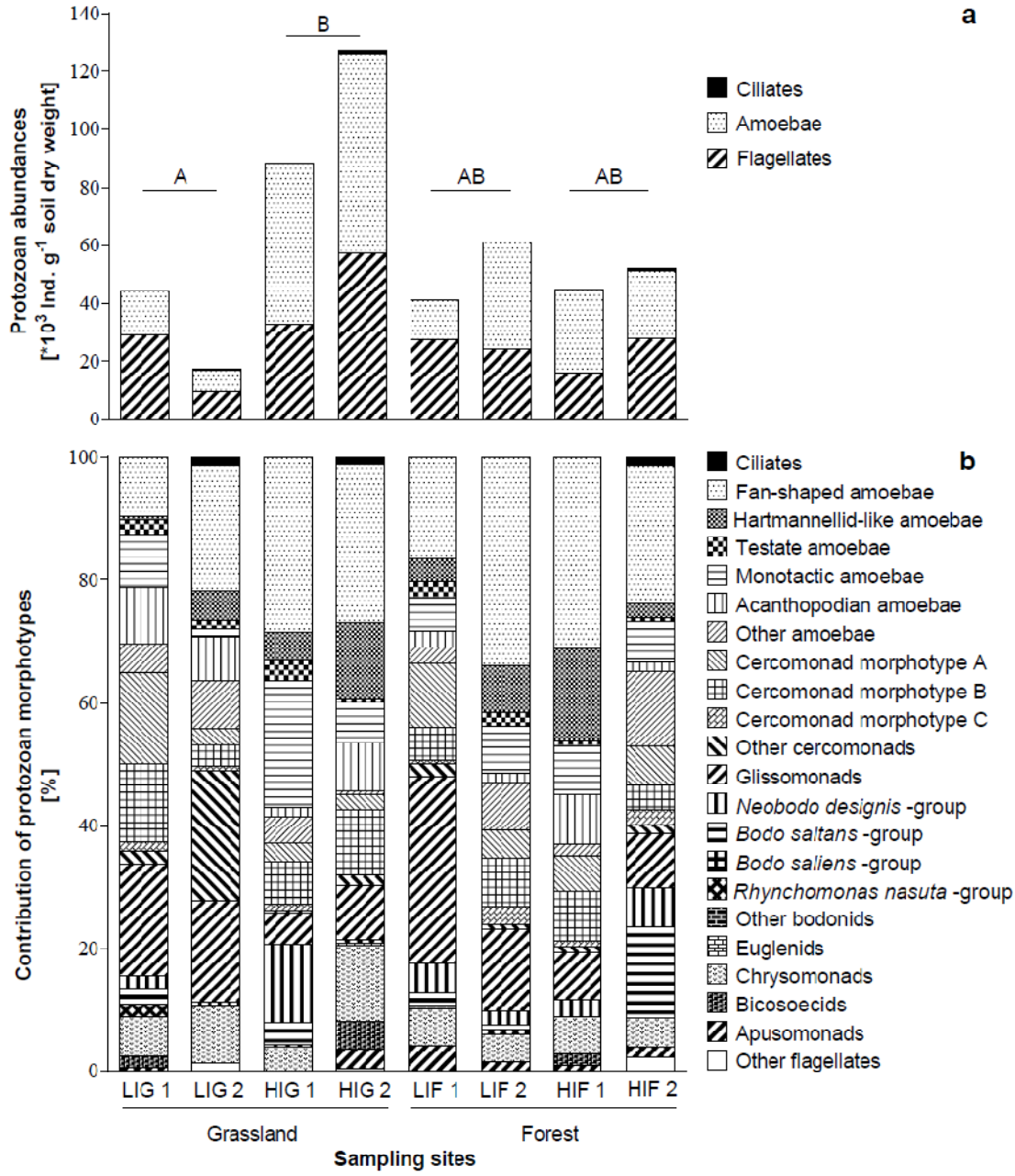


Figure 3. Protozoan community structure at grassland sites (LIG and HIG) and forest sites (LIF and HIF) in Thuringia, Germany. a) Abundances of cultivable protozoans of grassland and forest sites with different land use. Different letters indicate a significant difference (Turkey's Studentized Range Test (HSD) at p=0.05). b) Percentage contribution of different protozoan groups. The "other"-groups contained forms which could not be assigned to one of the specific morphotypes.

Table 2. Abundance estimates [ind./g soil dry weight] of cultivable protozoans of grassland (LIG and HIG) and forest (LIF and HIF) sites with different land use.

Morphotype	LIG1	LIG2	HIG1	HIG2	LIF1	LIF2	HIF1	HIF2
Bicosecids	902	0	0	5854	0	0	860	0
Chrysomonads	2784	1597	3497	15792	2417	2791	2624	2471
Rhynchomonas nasuta -group	902	0	257	630	0	455	0	0
Bodo saliens -group	0	117	0	0	0	0	0	0
Bodo saltans -group	1131	0	3218	0	857	455	0	7753
Neobodo designis -group	902	0	11330	0	1964	1384	1301	3318
Other bodonids	0	0	0	0	212	0	0	0
Euglenids	0	0	0	630	0	0	0	0
Apusomonads	222	0	0	3863	1740	915	428	810
Glissomonads	8149	2840	4499	11249	12487	8179	3475	4556
Other cercomonads	902	3651	257	1916	857	455	428	810
Cercomonad morphotype C	673	117	1043	0	212	1849	428	1226
Cercomonad morphotype B	5569	597	6130	13568	2190	4718	3523	2052
Cercomonad morphotype A	6644	476	2660	3208	4301	2791	2624	3318
Other flagellates	0	236	0	630	0	0	0	1226
Acanthopodian amoebae	4020	1215	1307	9982	1077	915	3523	810
Monotactic amoebae	3769	236	18211	8583	2190	4718	3523	3318
Testate amoebae	1131	236	2938	630	1077	1384	428	403
Hartmannellid-like amoebae	222	842	4068	15791	1517	4718	6780	1226
Fan-shaped amoebae	4273	3506	25119	33090	6821	20710	13859	11563
Other amoebae	2064	1341	3782	630	1077	4718	860	6386
Ciliates	0	236	0	1267	0	0	0	810

Discussion

There are less than a dozen studies available which comprise quantitative estimates of all soil protozoan groups (Ekelund et al. 2001; Finlay et al. 2000; Finlay and Fenchel 2001). And even less quantitative studies considered different taxonomic groups in more detail, especially Protozoa in the size range of 1-20 µm. Data was obtained on the most important protozoan groups, including flagellates, small naked amoebae, testate amoebae and ciliates. The most abundant protozoans, flagellates and naked amoebae were determined on the level of some easy recognizable morphotypes. Protozoan abundances ranged between 17.2 and 127.3 * 10^3 ind. g^{-1} dry weight which indicated their important role in the investigated soil systems. Especially, protozoan nanofauna was mostly neglected up to now. Due to their small body size, they have the highest metabolic activity among the protozoans (Fenchel 1974) and should contribute significantly to "animal" respiration (Foissner 1992).

Within the protozoans, heterotrophic flagellates with about 28 * 10^3 ind. g^{-1} dry weight and amoebae with about 31 * 10^3 ind. g^{-1} dry weight shared a similar part of total abundance at all sampling sites. The investigated protozoan abundances were in the range of published data. Finlay et al. (2000) summarized reported soil protozoan abundances, therein heterotrophic flagellates reached abundances of 0.77 – 539 * 10^3 ind. g^{-1} soil dry weight. Likewise for naked amoebae, numbers between 0.2 – 200 * 10^4 ind. g^{-1} soil dry weight have been recorded (Foissner 1999b). It is well known that protozoan abundances tend to vary widely with environmental parameters like rainfall, temperature, grazing pressure, physical disturbance and patchy distribution of bacterial food (e.g. Clarholm 1989; Clarholm et al. 1981; Ekelund and Ronn 1994; Finlay et al. 2000). Working with soil microorganisms always bears the problem that the organisms are masked by soil particles (Ekelund and Patterson 1997) and differentiation of active and inactive forms is difficult (Adl and Gupta 2006; Foissner 1987b). All cultivation based methods bear the disadvantage that uncultivable organisms are neglected (Caron et al. 1989). Thus the calculated abundances have to be considered as lower estimates. In addition, it represents what can actually grow in relatively small volumes and can overcome competition with slower growing species. Nineteen

different morphotypes of flagellates and amoebae were found. The low number of morphotypes likely hides a high hidden genotype diversity (e.g. Brabender et al. 2012; Finlay and Fenchel 2004; Scheckenbach et al. 2006). The communities investigated in our study included typical soil flagellates, such as cercozoans, bodonids and chrysomonads (Foissner 1991). However some other morphotypes found in our study such as apusomonads, euglenids and bicosoecids have been reported as rare taxa in the literature (Ekelund et al. 2001; Griffiths et al. 2001). There are different functional groups among the protozoans observed. Chrysomonads, bicosoecids as well as the *Bodo saltans*-complex are typical interception feeders in water films (22 + 10 % of total flagellate community). They capture particles carried along the flow lines produced by their flagellum (Boenigk and Arndt 2002). This feeding type is critically dependent on the availability of water. This is reflected by their relatively high contribution to the total number of soil flagellates at the sampling sites with the highest soil moisture content (grassland GL 2b, Fig. 3b). Euglenids, apusomonads, the *Rhynchomonas nasuta*-complex and the *Neobodo designis*-complex are representatives of raptorial direct interception feeders accounting for 13 + 10 % of the total flagellate community. Together with the cercozoans (64 + 14 % of total flagellate community), they inhabit the surface of the soil particles and feed on attached bacteria in biofilms. Cercozoans are amoebae like flagellates and typical soil inhabitants (e.g. Bass and Cavalier-Smith 2004; Ekelund and Patterson 1997; Foissner 1991). They dominated the flagellate community in all samples and were analyzed in more detail in a separate study (Brabender et al. 2012). These flagellates may reach into the smallest soil pores with the help of their pseudopodia and can rapidly reach favorable feeding sites with their flagella. Even at the study site with the lowest soil moisture, cercozoans were dominating and reached the highest abundance. Whenever the amoebozoan abundances were low, the cercozoan abundances were high and vice versa, suggesting competition between both groups.

The overall comparison of major taxonomic groups between grassland and forest sites without consideration of land use intensity revealed no significant differences. This implies that plant cover and the overlaying litter had no effect on the species composition of these tiny cultivable bacterivores. However, a higher resolution of

the community structure down to the morphotype level and site type level indicated high variations. The increasing intensity of land use in grassland sites was reflected by a significant increase of protozoan abundance and changes in the dominance structure. The amoebae abundances increased 5.4 fold, mainly due to fan-shaped and monotactic amoebae. This indicated a stimulus of anthropogenic influence to the protozoan community at grassland sites. The number of sites was too low to derive general conclusions yet, though there was a tendency for lower abundances and a higher proportion of cercozoans at grassland sites with low land use intensity. Adl and Gupta (2006) showed a loss of cercozoan species with increasing disturbance by tillage of soil and an increase of protozoan diversity with the time of no tillage. A significant decrease of cercozoans at intensively used sites was also reflected in our study. However, the intensity of land use was not reflected in differences of diversity estimates using the level of taxonomic resolution applied in our study. An influence of the different soil types, which have a different water potential, aeration, nutrient availability and soil compaction, could not be detected. A canonical correspondence analysis did not reveal any close relationship between the taxonomic structure and neither the pH values nor the total bacterial abundance. These observations may be explained by the ability of most soil protozoans to form cysts and exist independently of changing soil conditions. Another possibility could be that the influence of the different factors did affect cryptic species that could not be distinguished by morphology alone or that all species observed were generalists.

To sum up, abundances of cultivable forms reached values of up to $12.7 * 10^4$ ind. g^{-1} dry weight. The pH-values and bacterial abundance were not correlated to protozoan species composition. Grassland and forest sites did not show significant differences concerning the abundance and protozoan group distribution. But a higher resolution down to the morphotype level indicated a high variability between the different sites. The intensity of land use at grassland sites was reflected by a significant increase of protozoan abundances, mainly caused by an increase of amoebae. The flagellate community was dominated by cercozoans, followed by kinetoplastids and chrysomonads. Additionally, also rare forms in soil like bicosoecids and apusomonads could be detected. An effect of anthropogenic influence based on a different degree of land use was difficult to derive from the

present data set. There are indications that e.g. protozoan abundances at cattle grazed grasslands were higher.

Summary

Research on soil microbiota is one of the frontiers in biodiversity research. Bacterivores, such as protozoans, contribute significantly to mineralization processes and key ecosystem functions in soil. Protozoan grazing on rhizobacteria stimulates microbial mineralization processes and thereby indirectly nutrient cycling. Understanding the diversity of protozoans and their fundamental functions is essential for the knowledge of nutrient cycling. Reliable data on protozoan numbers and taxonomic composition in soil are scarce. The occurrence of morphotypes of small cultivable protozoans of eight different grassland and forest sites in Germany was compared. The investigated sites differed both, in type of area and intensity of land use. Additionally, a liquid aliquot method was modified to determine protozoan abundances and community structure on the morphotype level in soil. Protozoan abundances at the investigated sites ranged between 1.7 and 12.7 $*10^4$ ind. g^{-1} dry weight. The dominant groups were various amoebae (49.9 + 12.4 %) and cercozoans (32.4 + 13.2 %), followed by stramenopiles (7.7 + 4.2 %), euglenozoans (7.5 + 7.6 %), apusomonads (1.5 + 1.5 %) and ciliates (0.5 + 0.7 %).

Heterotrophic flagellates and naked amoebae showed similar contributions to total protozoan abundances at all sites. Protozoan abundances were slightly higher and more variable in grassland sites compared to forest sites. Both habitat types showed a surprisingly similar community structure on a rough taxonomic level, but a more diverse pattern at higher taxonomic resolution.

Chapter 2 – Diversity of soil cercomonad species: phylogeny, morphology and description of new species

Introduction

This Chapter focuses on the flagellate order Cercomonadida (= cercomonads). These globally distributed flagellates (Cavalier-Smith and Oates 2012; Karpov et al. 2006) are one of the most abundant flagellates in soil (Bass et al. 2009; Ekelund and Patterson 1997; Foissner 1991). Beside their major contribution to protozoan soil fauna, they inhabit other ecosystems like freshwater, marine and brackish water (Arndt et al. 2000; Auer and Arndt 2001; Bass et al. 2007; Bernard et al. 2000; Ekelund et al. 2004; Hänel 1979; Kiss et al. 2009). In general, flagellates play an important role in soil ecosystems, concerning mineralization processes, plant growth and nutrient uptake of plants (e.g. Bonkowski et al. 2000; Clarholm 1985; and see Chapter 1).

The identification of cercomonads is difficult because of their highly variable cell shape. Nevertheless, the biflagellated gliding amoeboflagellates are up to now mainly determined based on their morphology (e.g. Skuja 1939; Skuja 1948; Skuja 1956; Mylnikov and Karpov 2004). The development of molecular techniques enabled further work on the taxonomic position of cercomonads. A combination of morphological investigations, including detailed ultrastructural studies, with phylogenetic analysis is required to determine these highly metabolic flagellates. The first phylogenetic investigations of *Cercomonas* and *Eocercomonas* (both belonging to cercomonads) were performed by Ekelund et al. (2004) and Karpov et al. (2006). Bass et al. (2009) already suggested a revision of cercomonad taxonomy based on morphological and molecular data. This work was the most comprehensive study concerning cercomonad diversity. Cavalier-Smith and Karpov (2011) published a further taxonomic revision of cercomonads including the ultrastructure of *Paracercomonas*.

According to the analysis of Bass et al. (2009), up to now four different genera belong to the cercomonads, including at least 104 species.

The morphological investigation of the present study of the protozoan communities in different soil systems showed that cercozoans (mainly cercomonads) dominated the flagellate community across all different sites, see Chapter 1. The morphological investigations of the cercomonads prove to be difficult based on their metabolic cell shape in addition with their changing morphology in connection

with their particular life stage (e.g. dividing, encystment). Hence, cercomonads could only be determined to morphotypes and not to species level. The high contribution of cercomonads to the protozoan soil community and their interesting cell morphology lead us to investigate cercomonads based on morphological and molecular analysis in more detail. In collaboration with other scientists, 23 cercomonad clonal cultures were isolated from soil samples (according to Chapter 1) and analyzed. Within the analysis, ten new species and two new genera could be described.

In the following, I describe two new species. The other species descriptions can be found in the publication of Brabender et al. (2012). For the morphological investigations and the species description, a combination of three different microscopic methods was applied consisting of (i) water immersion phase contrast microscopy, (ii) inverted microscopy with a special oblique illumination and (iii) high resolution DIC microscopy. In all cases videomicrographic records were made. For the molecular and phylogentic analysis, the 18S rDNA gene sequence was used, as it is a distinct marker of species identity and has most entries in the database. The highly variable V4 region of the 18S rDNA was also used as marker sequence for Next-Generation-Sequencing of protozoan community structures (see Chapters 3, 4 and 5). Therefore the sequences of the newly described species could be included in the analysis of the Next-Generation-Sequencing.

Material and Methods

Study area and soil sampling

The sampling took place in the frame work of the interdisciplinary project, the Biodiversity Exploratories (http://www.biodiversity-exploratories.de). The samples were taken in April 2009 in the Hainich-Dün National Park (Thuringia, Germany), as one of the three investigation areas of the project. Soil samples were taken from respectively four grassland and four forest sites. The sampling procedure followed the protocol described in Chapter 1. The samples were stored at field moisture content at 4℃ until analysis started in D ecember 2009.

Cultivation and isolation

For the cultivation of the protozoans and the establishment of monoclonal cultures a modification of the liquid aliquot method (Butler and Rogerson 1995) was applied. 1 g of homogenized soil was suspended in a 75 cm2 tissue culture flask (Sarstedt, Nümbrecht, Germany) with 200 ml of Volvic© water and shaken gently by hand for ten minutes. 5 µl aliquots of the soil suspension were filled into the wells of six prepared 24-well plates (Sarstedt, Nümbrecht, Germany). Each well contained 1.5 ml Volvic© water (Danone Waters Deutschland GmbH, Frankfurt, Germany) and one sterilized amaranth grain (*Amaranthus caudatus*) to stimulate bacterial growth. After the inoculation of twelve wells, the soil solution was resuspended by shaking for a few seconds to avoid soil settlement. Finally, the plates were lined with Parafilm©. The plates were checked for appearing organisms weekly for a period of four weeks using an inverted microscope (Zeiss Axiovert S100) with phase contrast optics (63x 0.75 NA obj.).

By appearance of a single species in the well, 100 µl of the suspension was transferred at a clean bench to 30 ml WC-Medium (Guillard and Lorenzen 1972) in a 50 ml tissue culture flask (Sarstedt, Nümbrecht, Germany) including one sterilized quinoa grain (*Chenopodium quinoa*) as carbon source for bacterial growth. Two replicate monoclonal cultures were established and stored at about 14℃ in the dark. The cultures were checked every second month and inoculated into new culture flasks when organism numbers decreased.

Sequencing and phylogenetic analyses

Aliquot-PCR was used for further molecular analysis. Therefore, 15 µl aliquots of the monoclonal cultures were transferred to Eppendorf tubes (Eppendorf AG, Hamburg, Germany) and frozen at -20℃ for three hours before amplification PCR reactions were performed in a 50 µl reaction mixture containing both primers at a concentration of 0.1 µM each, all deoxynucleoside triphosphates at a concentration of 200 µM, reaction buffer (including 2 mM $MgCl_2$) and 1 U DreamTaq DNApolymerase (Applied Biosystems). The general eukaryotic primerset 18SFor (AAC CTG GTT GAT CCT GCC AGT) and 18SRev (TGA TCC TTC CGC AGG TTC ACC TAC) was used for amplification. The sequencing of

rDNA was done using Big Dye-Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Weiterstadt, Germany) in accordance with the manufacturer's instructions. The forward strand was sequenced by using the general eukaryotic primers EukA (AAC CTG GTT GAT CCT GCC AGT), 590F (CGG TAA TTC CAG CTC CAA TAG C) and 1280F (TGC ATG GCC GTT CTT AGT TGG TG). The reverse strand was sequenced by using the general eukaryotic primers 600R (GCT ATT GGA GCT GGA ATT ACC G), 1300R (CAC CAA CTA AGA ACG GCC ATG C) and EukB (TGA TCC TTC TGC AGG TTC ACC TAC). The obtained sequence parts of the 18SrDNA sequence were aligned and manually corrected with BioEdit (Hall 1999).

Further alignments of the different sequences were performed by using ClustalX (Thompson et al. 1997) and corrections were made manually with BioEdit 7.0.5.3 (Hall 1999). The GTRIG model was used for maximum likelihood analysis, which was determined by MrAIC (Nylander 2004). The maximum likelihood analysis was computed by PhyML (Guindon and Gascuel 2003) using 1000 replicates for the bootstrap analysis. MrBayes 3.1 (Ronquist and Huelsenbeck 2003) was applied for the Bayesian analysis. The analysis ran for 5,000,000 generations using GTR+G (4 discrete categories) + invariant sites with the covarion model and autocorrelation. The parameter values were estimated from the data by MrAIC. Trees were sampled every 100 generations. The tree search used two MCMCMC runs each with four chains (default heat parameters). The likelihood values of the two MCMCMC searches were compared to check whether they had converged. Every tree sampled before the likelihood plot reached a plateau was discarded (burn-in = 125,000). A consensus of the remaining trees was generated to show posterior probabilities of the branching pattern.

Videomicroscopy and image analysis

The used cultures aged from two days up to two months to give an overview of the typical forms and stages which appear during the time of cultivation. The cell shape, flagellar lengths, movement, behaviour and different life cycle stages of the species could be investigated efficiently by phase contrast and oblique

illumination, while high resolution differential interference contrast (DIC) was a powerful tool for the investigation of some ultrastructural characters of the cells.

Phase contrast microscopy

An upright Zeiss Axiophot 2 microscope with a 0.9 NA condenser, and a 100x 1.0 NA water immersion Zeiss Plan-Apochromat objective for phase contrast microscopy was used. A hole was cut into the top of the culture flasks to enable the lowering of the objective through the hole into the culture medium to get an undisturbed view at the organisms. The high definition video records were made with a Sony HDD Handycam AVCHD.

Oblique illumination microscopy

Beneath the phase contrast microscopy, a special oblique illumination microscopy for an undisturbed view on the organisms in the culture flasks at any life stage was applied. An inverted Zeiss Axiovert S100 microscope with a large working distance 0.55 NA condenser and a 63x 0.75 NA Zeiss Fluoroplan objective corrected for plastic was used. The microscope was equipped with multiple oblique illumination to enhance contrast and resolution. With this system a good relief contrast and a resolution of about 350 nm was achieved. Videos were recorded using a black/white analogous Hamamatsu C6489 camera. For noise reduction and contrast enhancement the signal was handled by an Allen Video Enhanced Contrast (AVEC) system (Hamamatsu, Argus-20). The Canopus ADVC-300 A/D converter with subsequent 3D noise reduction and further contrast enhancement was used for digitalisation of the videos.

High resolution differential interference contrast (DIC) microscopy

To investigate the cell structures in living cells an upright Zeiss Axiophot 2 microscope with a 1.4 NA condenser and a 100x 1.4 NA HI Zeiss Plan Apochromat objective was used. Small subsamples of dense monoclonal cultures were transferred onto a slide and covered by a coverslip. Oil was used between

objective and coverslip as well as between condenser and slide. A 445/50 BP interference filter was used to decrease the wavelength of illumination. In combination with the AVEC-DIC system, the differential contrast enhancement and the wavelet sharpening, the resolution of the system was about 150 nm. The high resolution enabled detection and recording of very fine structures of the organisms.

The same video recording system was used as in the oblique illumination system. The high resolution DIC videos are deposited in the video library of the University of Cologne, Department of General Ecology.

Video and image analysis

The recorded videos were analysed by watching them frame by frame. The images of the frames were enhances by noise filtration, differential contrast enhancement and Wavelet transformation sharpening (VirtualDub, ImageJ, CellD, AviStack softwares).

Results

The inoculation of the diluted soil samples into the 24-well plates resulted in 233 monoclonal cultures, from which 23 were cercomonads. These 23 cultures represented 17 species, from which ten were new, including two new genera. Here I present the results of the two new described species *Cercomonas jendrali* and *Paracercomonas bassi* and additionally a short overview about the two new genera. An overview about all ten newly described species is given in Figure 3.

Morphological data on new strains

The more general morphological characteristics of both new species are summed up in Appendix Table 2.1 and are not repeated in the detailed description of the species. These measurements included: cell length, cell with, length of anterior flagellum, gliding speed and diameter of cyst. The observations concerning

movement, behaviour, life cycle and typical characteristics of the organism are listed in Appendix Table 2.2. Displays of the phase contrast microscopy for visualisation of the movement and the behaviour of the organisms are deposited. The underlying characteristics of the description followed the nomenclature established by Bass et al. (2009). Some descriptions used in Table 2 require further explanation for better understanding of meaning. (i) Probing movement of the anterior flagellum = non-directed movement of the flagellum (ii) In the description of the life cycle the term 'unattached' = lack of attachment of the cell body (not the entire organism) to the substratum (iii) unattached, non-gliding, wobbling cells are attached to the substratum with the distal end of the posterior flagellum and wobble on it.

Species concept

Species differentiation was morphologically based on the nomenclature established by Bass et al. (2009) and genetically based on 18SrDNA barcoding of species also applied by Bass et al. (2009). Organisms were separated to species on basis of unambiguous nucleotide changes in the barcoding sequence. The classification of genera was justified by significant phylogenetic distance in the maximum likelihood 18SrDNA tree. In addition with morphological differences between the two groups of organisms, a new genus was established.

Morphological descriptions of new species and comparison with similar species

Cercomonas jendrali (Domonell, Brabender, Kiss, Nitsche & Arndt. 2012), Figure 1, type strain: HFCC 905 (Hainich- Dün; forest soil; 2009), type sequence: 18S rDNA: Hm536150, type video material: Ca15-Ca19.

Diagnosis: *Cercomonas* with a body length of 8.9-25.1 µm and a body width of 4.7-15 µm. Anterior flagellum has a length of 1.4-1.9x the body length, the posterior flagellum is 1.1-1.6x the body length. Gliding cells are elongated oval or oblong, or often irregularly amoeboid. Cells are very metabolic, but change their shape slowly. They form finger-like and bulbous pseudopodia at the posterior end and lamellar pseudopodia at the left cell edge. Unattached, non-gliding, wobbling

forms are present. The nuclear envelope of the large vesicular nucleus is thin, but several microbodies attach to it. The thick nucleoplasm region contains many small granules; the large central nucleolus is spherical. The nucleus is closely connected to the basal bodies. Several contractile vacuoles are situated posteriorly, anteriorly and laterally. There are 10-20 larger crystals are evenly distributed in the cytoplasm. Small spherical extrusomes are situated beneath the plasma membrane. The species is able to form irregular angular cysts with swellings and depressions. The cyst wall is thick and creased, it may be double-layered; the cystoplasm attaches closely to it.

Description: Gliding cells are generally elongated oval or oblong, but often irregularly amoeboid because of the pseudopodia. The body length is 8.9-26.1 µm (20.2 µm on average) and the body width is 4.7-15 µm (11 µm on average). Rounded cells are 7-11 µm long and 7.3-9.2 µm wide. The mean anterior flagellum length is 31.7 µm (1.4-1.9 times the body length, 1.6 on average). The posterior flagellum has a mean length of 26.4 µm (1.1-1.6 times the body length, 1.3 on average). The cell glides at a low speed (mean 0.9 µm/s). The entire anterior flagellum is mostly probing, the distal end of it beats with a frequency of 1.3 Hz. Cells are very metabolic and regularly form finger-like and bulbous pseudopodia, mainly at the posterior part of the cell (Fig. 1C). Lamellar pseudopodia are often formed; they are produced exclusively at the left cell edge (Fig. 1A, C). The movement is only seldom directed for a longer distance. The cells glide in a flowing manner with many changes in direction. Unattached, non-gliding, wobbling forms are found. These are rounded and flap on the tip of the posterior flagellum by the beating of the posterior flagellum. The flagellar insertion is apical (Fig. 1E, I). The posterior flagellum attaches to the cell body along the entire body length (Fig. 1E). Cells have a large spherical, vesicular nucleus (Fig. 1D, F). The nuclear envelope is thin, but in several parts small granules attach to it, giving it a thicker appearance (Fig. 1F; presumably microbodies, like in *C. karpovi*, see Karpov et al. 2006). The nucleoplasm region is thick and contains many slightly refractile granules (Fig. 1 D, F). The large central nucleolus is spherical and may have an inner lacuna (Fig. 1D). The nucleus is closely connected to the basal bodies (Fig. 1I). Several contractile vacuoles are situated anteriorly, laterally (Fig. 1D) and posteriorly. There are about 10-20 larger crystals evenly distributed in the

cytoplasm (Fig. 1H). Small spherical extrusomes are situated beneath the plasma membrane (Fig. 1E). This species is capable of forming large cysts (mean diameter 13.1 µm). Cysts are angular and have rounded protuberances and depressions on the surface; the cyst wall is relatively thick and creased (Fig. 1B) and may be double-layered. The cystoplasm closely attaches to the cyst wall. Etym. dedicated to Bärbel Jendral for her valuable contribution and support during the cercomonad studies.

Distinction from similar species: Similar species in the same size range, described with gene sequence data (species described in Bass et al. 2009): Clade A1b *Cercomonas dactyloptera* (sensu Bass et al. 2009) is very slow, often remaining in one location; it produces a filose cytoplasmic tail and bulbous pseudopodia all over the cell. *Cercomonas lata* occasion-ally produces a filose cytoplasmic tail and filose and bulbous pseudopodia all over the cell. It contains many more strongly refractile cytoplasmic granules. *Cercomonas clavideferens* produces spike-like pseudopodia anteriorly, which move into posterior direction. *Cercomonas parincurva* produces finger-like, bulbous, lamellar and filose pseudopodia, mostly on left and posterior sides and rarely on the right side; many types of pseudopodia are often present in one cell. Clade A1a *Cercomonas lenta* is similar, but produces cytoplasmic tails, filose pseudopodia and anterior bulbous pseudopodia. *Cercomonas paravarians* frequently has lamellar cytoplasmic tails and posterior lamellipodia on both sides. Many clade A1b2 species are quite similar to *C. jendrali*: Species described without gene sequence data: *Cercomonas varians* has a long filose cytoplasmic tail and an anterior contractile vacuole (Skuja 1948). *Cercomonas jendrali* can be distinguished from most other *Cercomonas* species by the production of posterior bulbous and finger-like pseudopodia, the left lamellar pseudopodia and the lack of filopodia or filose cytoplasmic tails.

Figure 1. General morphology and pseudopodium formation of *Cercomonas jendrali*
A: Cell shape and pseudopodia. B: Cysts. C: Gliding and pseudopodium forming specimens. D-I: Fine structure of the cells. D, F: Structure of the nucleus. E: Flagellar insertion with the basal bodies, and extrusomes beneath the plasma membrane. G: Contractile vacuole. H: Crystals in the cytoplasm (white arrows). I: Connection of the nucleus to the basal bodies. A: phase contrast, B-C: oblique illumination, D-I: high resolution DIC. Scale: A: indefinable, B-C: 10 µm, D-I: 5µm.

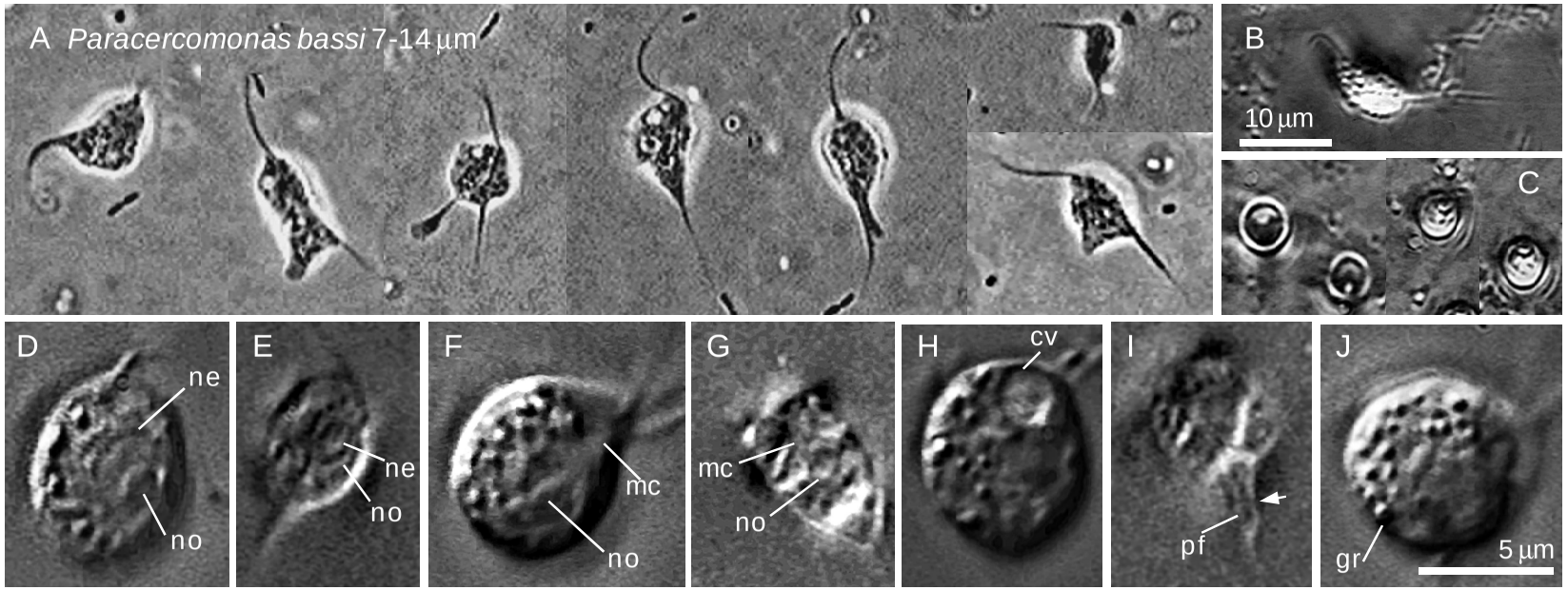


Figure 2. General morphology and pseudopodium formation of *Paracercomonas bassi*.
A-B: Cell shape and pseudopodium formation. C: Resting stage. D-J: Fine structure of the cells. D: Contractile vacuole. E: Extrusomes. F-H: Structure of the nucleus. I: Flagellar insertion with the basal bodies. J: Connection of the nucleus to the basal bodies. K-T: Paracercomonas bassi. A-B: Cell shape and pseudopodium formation. C: Cysts. D-J: Fine structure of the cells. D-E: Structure of the nucleus. D: Eccentric nucleolus. E: Central nucleolus. F-G: Microtubular cone connecting the nucleus to the basal bodies. H: Contractile vacuole. I: Lamellar posterior pseudopodium (white arrow). J: Strongly refractile cytoplasmic granules. A: phase contrast, B-C: oblique illumination, D-J: high resolution DIC. Scale: A indefinable, B-C: 10 µm, D-J: 5 µm.

Paracercomonas bassi (Domonell, Brabender, Kiss, Nitsche & Arndt 2012), Figure 2 B-I, type strain: HFCC 918 (Hainich-Dün; grassland soil; 2009), type sequence: 18S rDNA: Hm536163, type video material: Ca38-Ca46.

Diagnosis: *Paracercomonas* with a body length of 7.4- 14.5 µm and a body width of 4.9-8.7 µm. The anterior flagellum is 0.9-1.6x the body length, the posterior flagellum is 1-2.2x the body length. The drop-shaped, fusiform or triangular cells are very metabolic and change their shape relatively quickly. The cells form finger-like and bulbous pseudopodia at the posterior cell part and small lamellipodia on the left lateral side. Spatulate pseudopodia are sometimes produced laterally and posteriorly. Unattached wobbling forms are present. The nuclear envelope of the large vesicular nucleus is quite thin, the nucleoplasm is homogeneous, the spherical nucleolus is central or eccentric. The nuclear envelope is connected to the basal bodies by a long microtubular cone. The single contractile vacuole is anterio-lateral. About 10-25 larger strongly refractile granules are present in the cytoplasm. No extrusomes were found. The species is able to form regular spherical cysts with smooth walls. The cystoplasm closely attaches to the cyst wall.

Description: *Paracercomonas* with a body length of 7.4-14.5 µm (10.8 µm on average) and a body width of 4.9-8.7 µm (6.8 µm on aver- age). Rounded cells are 6-9 µm long and 3.6-5.3 µm wide. Their mean anterior flagellum length is 13 µm (0.9-1.6 times the body length, 1.2 on average). The posterior flagellum has a mean length of 16.7 µm (1-2.2 times the body length, 1.6 on average). The cells glide slowly with a mean speed of 1.3 µm/s and nervously beat with the distal part of its anterior flagellum with a frequency of 1.4 Hz. The anterior flagellum is straight and relatively stiff; most beating is restricted to its distal part. It is acronematic (Fig. 2A). Sporadically, the entire anterior flagellum flaps more than 90 degrees, often reaching the cell side. The drop-shaped, fusiform or triangular cells are very metabolic and change their shape relatively quickly. They form finger-like and bulbous pseudopodia mainly at the posterior part of the cell (Fig. 2A-B); lamellar-like projections are also produced on the left lateral side (Fig. 2A). A filose plasmatic tail can be formed while gliding (Fig. 2A, I). Spatulate pseudopodia are sometimes produced laterally and posteriorly (Fig. 2A). There are straight stretches during gliding, but the cell often changes its direction.

Unattached, non-gliding forms are found wobbling on their posterior flagellum. The flagellar insertion is apical (Fig. 2E-F). The posterior flagellum attaches to the cell body along the entire cell body length (Fig. 2I). The anterior or central (Fig. 2G) nucleus is vesicular. It is covered by a thin and smooth nuclear envelope (Fig. 2E). The nucleoplasm is very homogeneous with some denser patches (Fig. 2D-E) and forms a thick layer. The homogeneous, spherical nucleolus may have a central (Fig. 2E, G) or eccentric position (Fig. 2D). The nucleus is connected to the basal bodies by a long microtubular cone (Fig. 2F-G). The single contractile vacuole is situated in the anterior or lateral cell parts (Fig. 2A, H). In the cytoplasm there are 10-25 larger strongly refractile granules (Fig. 2J). No extrusomes were found. This species is capable of forming cysts (mean diameter 8.2 µm; Fig. 2C). Cysts are regular spherical with a smooth wall. The cystoplasm closely attaches to the cyst wall. A large refractile eccentric body is situated in the cystoplasm. Etym. dedicated to David Bass as a pioneer in the research of the diversity, phylogeny and taxonomy of Cercozoa.

Distinction from similar species: Similar species in the same size range, described with gene sequence data (species described in Bass et al. 2009): Most metabolic *Paracercomonas* species, which are similar in shape and the production of finger-like, bulbous and lamellar pseudopodia (*Paracercomonas vonderheydeni*, *P. minima*, *P. producta*, *P. oxoniensis*, *P. crassicauda*, *P. elongata* and *P. paralaciniaegerens*) differ in the ability of filose pseudopodium production. *Paracercomonas astra* differs in the absence of spatulate pseudopodia, and in the frequent anterior production of finger-like pseudopodia. *Paracercomonas saepenatans* has a much longer anterior flagellum. Many metabolic *Cercomonas* clade A1a species are rather similar in general appearance. Most similar species, like *Cercomonas nebulosa*, *C. mutans*, *C. volcana*, *C. longicauda*, *C. hederae*, *C. mtoleri*, *C. kiaerdammane*, *C. ambigua*, *C. parambigua*, *C. paraglobosa* and *C. fastigata*, have longer anterior flagellum. *Cercomonas phylloplana* produces lamellipodia at the posterior cell part and at both sides. *Cercomonas laeva* produces filose and filose branching pseudopodia. *Cercomonas karpovi*, *C. kolskia*, *C. elliptica* and *C. rotunda* are more compact and roundish. No *Cercomonas* species are reported to produce spatulate pseudopodia. Similar, metabolic *Eocercomonas* species (like *E. tribula*, *E. echina*) are smaller and produce filopodia. Species described without gene sequence data: *Cercomonas*

levis is able to produce lamellar pseudopodia posteriorly, but lateral lamellipodia are not reported and the anterior flagellum beats with many sinusoid waves (Skuja 1939). *Cercomonas varians* also produces large posterior filose and lamellar projections and contains cytoplasmic granules, but left lateral lamellipodia are not reported, the anterior flagellum beats with several sinusoid waves and the posterior flagellum does not always attach to the cell body (Skuja 1948). *Cercomonas. agilis* (recently combined to *Agitata agilis*, Howe et al. 2011) similarly produces posterior lamellipodia and finger-like pseudopodia, but its gliding form is much thinner and spindle-shaped (Lemmermann 1914). *Cercomonas amoebinus* produces extended lamellipodia all around the cell, however, its gliding form is more elongated fusiform and its anterior flagellum is much longer (Mylnikov 1985). *Cercomonas plasmodialis* produces large left lateral lamellipodia, but its anterior flagellum is longer and its posterior flagellum does not always attach to the cell body (Zhukov 1993). *Cercomonas pseudocrassicauda* produces few posterior filopodia, but has large lamellipodia on both lateral sides and even at the anterior parts (Tikhonenkov 2007). The short anterior flagellum, the lack of filopodium production (except for the cytoplasmic tail), the left lateral lamellipodia and the lateral-posterior spatulate pseudopodia distinguish *Paracercomonas bassi* from most cercomonad species with a similar appearance.

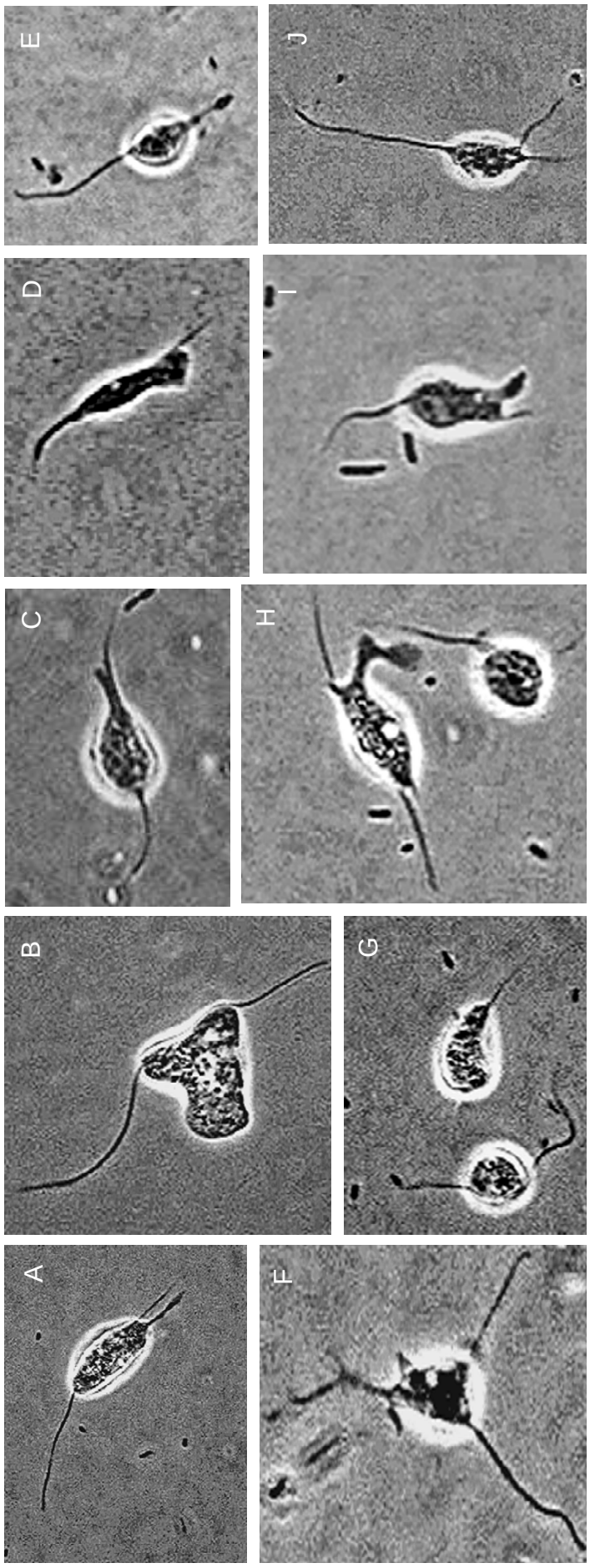


Figure 3. Overview about the morphology of the ten newly described species.
A: *Cercomonas directa* (14-22 µm). B: *Cercomonas jendrali* (9-25 µm). C: *Paracercomonas bassi* (7-14 µm). D: *Paracercomonas proboscata* (6-11 µm). E: *Eocercomonas exploratorii* (5-12 µm). F: *Eocercomonas perecta* (9-12 µm). G: *Cercomonas pellucida* (9-17 µm). H: *Metabolomonas insania* (9-13 µm). I: *Paracercomonas kruegeri* (8-11 µm). J: *Nucleocercomonas praelonga* (10-13 µm).

Two new genera could be described during the work on cercomonad species diversity (see also Brabender et al. 2012). These genera are *Nucleocercomonas* and *Metabolomonas* and will briefly described here.

Representatives of the genus *Nucleocercomonas* are biflagellated, basically non-gliding Cercozoa with a spherical or oval cell body. The anterior flagellum is several times longer than the cell body, and is longer than the posterior one. The most frequent life cycle stages are cells, which do not attach to the surface with the cell body, but only with the distal part of the posterior flagellum. Finger-like, bulbous, filose and lamellar pseudopodia are produced into the water body or along the attachment surface. Diverse life forms are present in this genus including unattached non-gliding amoeboid (attached to the substratum with the distal end while the cell body is raised up), attached amoeboid, free swimming, cyst and plasmodial stages. The description of the genus *Nucleocercomonas* included the description of the type species *Nucleocercomonas praelonga*. The lack of gliding in the most frequent life stage, the very long anterior flagellum and the unique nucleus structure distinguishes this genus from all other cercomonad genera.

The other newly described genus is *Metabolomonas*. Representatives of this genus are biflagellated, fast gliding, amoeboid cercomonads with a highly variable shape. The cells are extremely metabolic and change their shape very quickly. They produce finger-like, bulbous, filose, lamellar or smaller anastomosing reticulate pseudopodia very quickly anywhere on the cell. Gliding, non-gliding amoeboid, unattached non-gliding wobbling, swimming, resting and plasmodial stages may be present as life stage in this genus. The type species *Metabolomonas insania* was also described. The main characteristic of this genus is the high beating frequency of the anterior flagellum in combination with the fast gliding speed. The very metabolic cell and quickly changing cell shape, the fast production of pseudopodia and the intensive cytoplasmic streaming and the large vesicular nucleus are also typical characteristics of *Metabolomonas*.

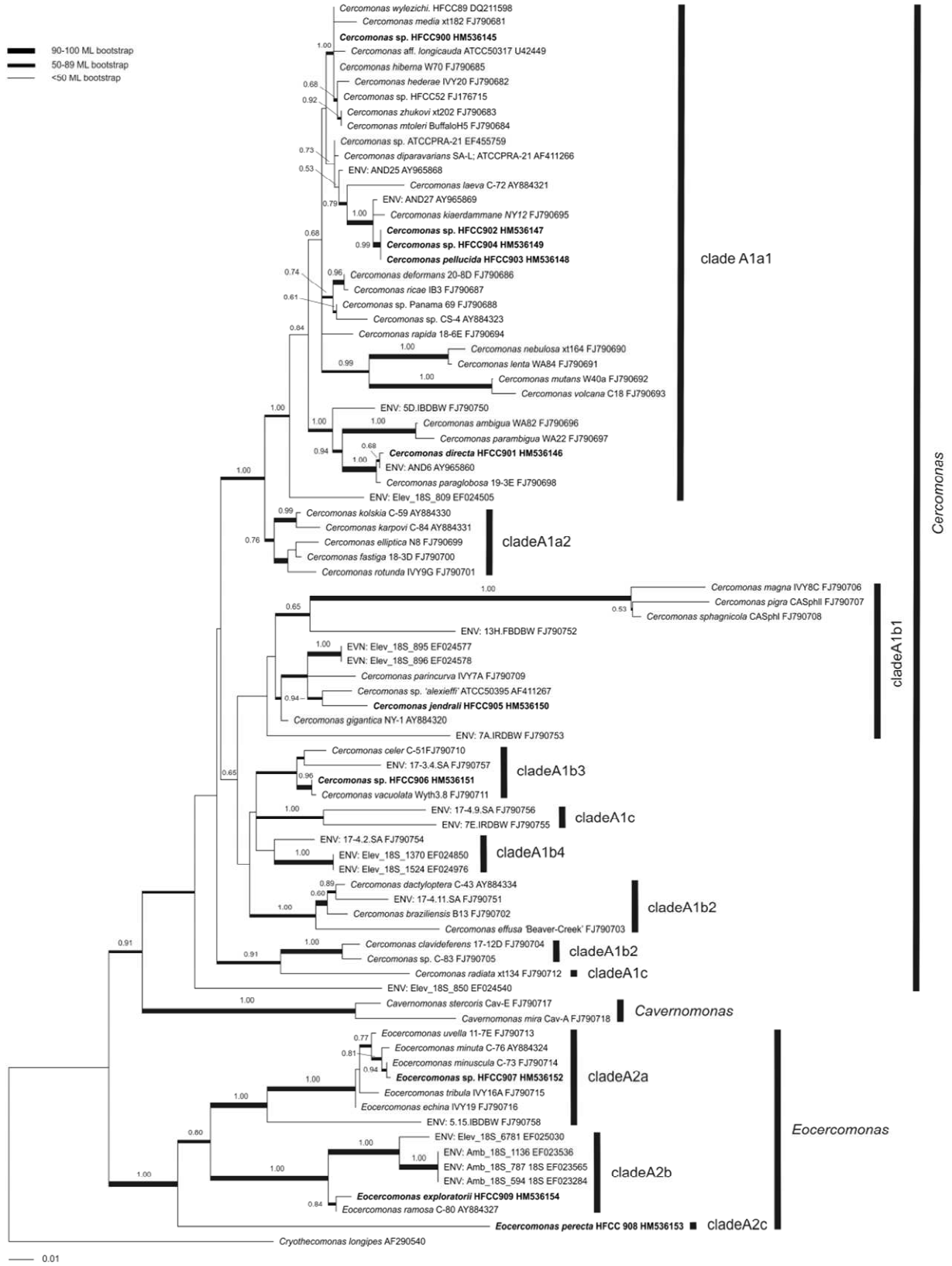


Figure 4. Maximum Likelihood (ML) 18S rDNA phylogenetic tree based on 83 sequences of clade A cercomonads from GenBank and this study, created from 1069 aligned nucleotide positions. ML bootstrap values are shown by differences in the thickness of branches. All Bayesian posterior probabilities (PP) are denoted at the branches when exceeding 0.50. The species *Cryothecomonas longipes* was used as an outgroup.

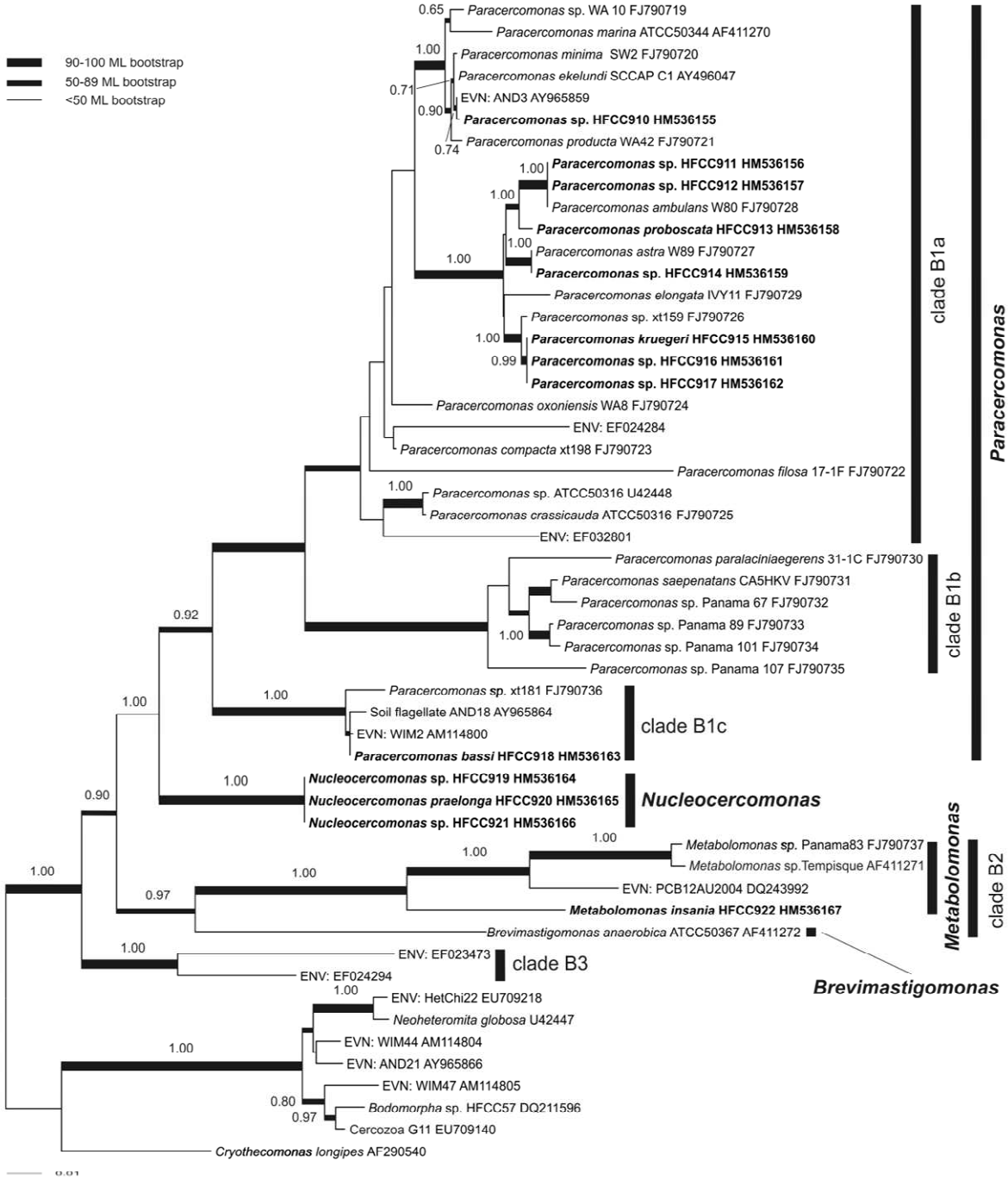


Figure 5. Maximum Likelihood (ML) 18S rDNA phylogenetic tree based on 45 sequences of clade B cercomonads from GenBank and this study, created from 1539 aligned nucleotide positions. ML bootstrap values are shown by differences in the thickness of branches. All Bayesian posterior probabilities (PP) are denoted at the branches when exceeding 0.50. Several glissomonad sequences from GenBank were selected as outgroup.

Phylogenetic analysis

The 18SrDNA of all isolated monoclonal cercomonad cultures were sequenced. The results of the phylogenetic analysis are presented in Figures 4 and 5. The cercomonads were separated in two clades A (Fig. 4) and clade B (Fig. 5) according to Bass and Cavalier-Smith (2004), Cavalier-Smith and Chao (2003) and Karpov et al. (2006), and confirmed with the results of our study. The newly described species and genera are marked in bold in the phylogenetic tree. For a higher resolution, the two clades were separately analyzed. Ten of our 23 sequences belong to clade A and ten species sequences belong to clade B. *Cercomonas jendrali* fell into sub-clade A1b1 as one of seven new species belonging to *Cercomonas* (clade A1). The other three new species of this clade were belonging to *Eocercomonas* (clade A2). Its closest cultured relative is the probably misidentified strain *Cercomonas* sp. "*alexieffi*" ATCC50395 (Bass et al. 2009), while the closest described species is C. parincurva. As one of the nine new species which belong to *Paracercomonas* (clade B1), *Paracercomonas bassi* fell as a single one into clade B1c. This new species is the first morphologically well-described species of this clade. The other sequences were belonging to the new genus *Nucleocercomonas* (three new species) and *Metabolomonas* (one new species).

Discussion

Cercomonads form an important part of soil microbiota, due to their high abundances (Ekelund and Patterson 1997; Foissner 1991) and their contribution to remineralization of inorganic nutrients (Bonkowski et al. 2000). Their typical morphology is based on a very metabolic cell shape, which made the identification of certain species very difficult. The phylogenetic position of the organisms is still uncertain and needs revision (Bass et al. 2009). The morphological and molecular analysis of the newly described species and genera enhanced the knowledge about cercomonad phylogeny. The analysis of our newly described species and genera increased the resolution of cercomonad systematics. The included newly described species often fell into formally weakly resolved clusters which increased the reliability.

The phylogenetic position of the newly described species was additionally related to the morphology of the next related species to quantify the information of the phylogenetic analysis.

Cercomonas jendrali spec. nov. appeared in clade A1b1, belonging to the genus *Cercomonas* subclade cladeA1 (established by Bass et al. 2009). The next described species relatives of *Cercomonas jendrali* were *C. sp. alexieffi*, *C. gigantic* and *C. parincurva*. Representatives of *Cercomonas*, diagnosis in Karpov et al. (2006), after type species *Cercomonas longicauda* (Dujardin 1841; neotypified: Karpov et al. (2006); neotype CCAP 1910/2), are variably sized gliding biflagellates, often with long flagella, including the largest known cercomonads; with mostly rounded or finger-like pseudopodia (Bass et al. 2009). Members of clade A1b are significantly larger than those in clade A1a (Bass et al. 2009). *Cercomonas jendrali* is the largest species described in our study. It has a flattened body shape, long flagella and often produces very large lamellar, finger-like pseudopodia. Its closest relative *C. parincurva* is considerably similar regarding most morphological characters.

Paracercomonas bassi nov. spec. was positioned in clade B1c. The cluster B1c is phylogenetically far away from the other *Paracercomonas* clusters B1a and B1b. Nevertheless, a clear differentiation of *Paracercomonas bassi* to other species of clade B1 is impossible. The characters found in Paracercomonas bassi can also be found in the other clades among some other species. E.g., the lack of filopodium production is not frequent among *Paracercomonas* species, just as spatulate pseudopodia, but both occurs also among B1a and B1b species. Any other differences of *P. bass*i to other *Paracercomonas* species are small and unremarkable. The relatively closely related strain AND 18 (Lara et al. 2007) has large refractile granule in the body cells and cysts, but only one large granule could be found in the cysts of *P. bassi*.

The new genera were established based on the high phylogenetic distance and the distinct morphology of the included species. The morphology of *Nucleocercomonas* completely differs from any other cercomonad species or genus. Some life forms seem very uncommon or even new among flagellates and have never been observed in cercozoa before. The most frequent life form of *Nucleocercomonas*, the amoeboid, only with the end of the posterior flagellum attached cell body, stretching pseudopodia into the water body have never been found in attached flagellates before. The types of pseudopodia, high metabolicity and rare gliding activity in the *Nucleocercomonas* clade justify the affiliation to the Cercozoa.

The clade of the new genus *Metabolomonas* is phylogenetically even farther away from the other *Paracercomonas* clades than *Nucleocercomonas*, but morphologically more similar to other species of *Paracercomonas*. Typical unique morphological characteristics of *Metabolomonas* are the extremely high metabolic activity, the fast movement and the large smooth vesicular nucleus. Some other characteristics are present in *Metabolomonas* which can also rarely be found among *Paracercomonas*, like the spatulate pseudopodia and the anastomosing reticulopodia.

Metabolomonas build clade B2 together with *Brevimastigomonas,* which was newly erected, based on the high phylogenetic distance between both genera. The typical morphological characteristic of the type species *Brevimastigomonas anaerobica* is its very short anterior flagellum. Together with its slow movement and its moderate metabolic activity, *Brevimastigomonas* is clearly different from *Metabolomonas*.

The results of the phylogenetic analysis indicated that a close phylogenetic relationship does not correspond automatically to a similar morphology. The high phylogenetic distance of different genera does not correspond to continuous morphological differences. Closely related species can morphologically more dissimilar than distant related ones and the other way around. Therefore a determination of *Cercomonas* species based only on morphological characteristics is impossible.

The 10 newly described species, including two new genera, were based on only 23 monoclonal cercomonad cultures. On the other hand, only four cercomonad cultures could be identified to species level based on the identical 18SrDNA sequences. This indicates how poorly the diversity of cercomonads is explored and underlines the importance of further research on the cercomonads.

Summary

Cercomonads are globally distributed flagellates and the most abundant flagellates in soil ecosystems. In addition, they form an important part of the microbial food web in soil and aquatic sediments. Based on their ecological importance and their interesting cell morphology, a detailed study on cercomonads was applied. Therefore, the general morphology, behaviour, life cycle and 18S rDNA phylogeny of isolated cercomonad cultures from a German grassland soil was investigated. In collaboration with other scientists, ten new species and two new genera could be described from 23 isolated strains. The new species belonged to the genera *Cercomonas*, *Eocercomonas* and *Paracercomonas*. The two new novel clade B genera were *Nucleocercomonas* nov. gen. and *Metabolomonas* nov. gen.. These genera were erected based on the large phylogenetic distance to the other cercomonad genera and their distinct morphology. The two new species which were described in more detail were *Cercomonas jendrali* spec. nov. and *Paracercomonas bassi* nov. spec., located in subclade A1b1 and clade B1c. Representatives of clade A1b are significantly larger than those of clade A1a. *Cercomonas jendrali* is the largest species described in our study. Although, *Paracercomonas bass* was located in the cluster B1c, which is far away from the other *Paracercomonas* cluster a clear differentiation between the other species of clade B1 is impossible.

In general, the phylogenetic and morphological analysis indicated a high variation in the morphology even between closely related organisms. Thus, the morphology of cercomonad species often does not correspond with their phylogenetic position. A clear species determination based only on morphological characteristics is impossible for cercomonad species, requiring further phylogenetic analysis.

Chapter 3 – Evaluation of next generation sequencing to identify in-vitro simulated flagellate community structures

Introduction

Investigations of protozoan communities can essentially increase our understanding of various basic ecosystem functions. Historically, most studies of protozoan diversity relied on isolation, followed by cultivation and species identification based on morphological and molecular determination (Foissner et al. 2003; Lara et al. 2007; Treonis et al. 2010). The resulting monoclonal cultures are essential to describe new species and label sequences in databases on a distinct level (Bass et al. 2009; Brabender et al. 2012; Howe et al. 2011). Only these sequences, which include a description of origin and species name, can provide the basis for highly resolved molecular analysis. However, isolating and cultivating is a time-consuming and moreover very selective process. It is assumed that only 1% of the soil microorganisms are cultivable (Amann et al. 1995; Smit et al. 2001; Staley and Konopka 1985; Torsvik et al. 1996). Therefore, morphological investigations, relying on these traditional methods, might seriously underestimate species richness (Sorgin et al. 2006). Moreover, the time-consuming cultivation work may result in only rough taxonomic resolutions when no molecular biological methods are applied.

The development of next generation sequencing (NGS) techniques, like Illumina and Roche 454, revolutionized protozoan diversity studies during the last years. Up to now, a single sequencing run of 454 can yield more than one million sequence reads with a mean length of 400 nucleotides (nt). This huge amount of data enables a highly resolved analysis of e.g. species richness and diversity (Buee et al. 2009; Gottel et al. 2011; Jumpponen and Jones 2009; Lecroq et al. 2011; Nacke et al. 2011; Pawlowski et al. 2011; Stoeck et al. 2010; Stoeck et al. 2009). As NGS is applied more frequently and the data yielded by this method accumulate, new questions arise on the application fields as well as the data reliability of these methods (Edgcomb et al. 2011; Lee et al. 2012; Medinger et al. 2010; Weber and Pawlowski 2013). Sequencing errors of individual sequencing reads can result in misidentification and an overestimation of diversity (Huse et al. 2010; Kunin et al. 2010). To avoid this misinterpretation, a strict quality filtering has to be included into the sequence analyses.

Here, different aspects of NGS by Roche 454 sequencing were investigated. The recovery of sequences, which were previously obtained from Sanger sequencing of

isolated and cultivated flagellate species of different soil samples was examined. Therefore, the soil samples were investigated on one hand by classical morphological approaches completed by Sanger sequencing and on the other hand by NGS. The incomplete recovery of sequences by NGS leads to the question, how accurately communities can be discovered by NGS. Thus, an experiment was designed, composed of defined, different artificial protozoan communities based on DNA extracts. On the one hand small communities based on three species, with variable contributions of each species and on the other hand larger communities with equal contributions were composed. Both artificial communities were designed to test if the abundance of a certain organism influences the reliability of the method, or if the species richness of a certain community has the stronger influence. In addition to the influence of composition and number of species, the effect of preceding PCR amplification on the outcome of community structure analysis was investigated.

Material and Methods

Study Design

As a basis for the study isolated cercomonad species which were obtained from environmental soil samples (see Chapters 1 and 2; Brabender et al. 2012; Domonell et al. 2013) were sequenced. The same environmental soil samples were also sequenced by NGS. The NGS dataset was then checked for the recovery of the previously obtained Sanger sequences. The obtained results showed the community structure to be very different, depending on the method used. To unscramble possible reasons, a simple three-species community was composed. The DNA of the species was mixed in different concentrations to create an artificial community consisting of rare and dominant species like in natural communities. The obtained results indicated a high reliability of NGS for small communities. In addition to the small communities we enlarged the communities to twelve species and used equal amounts of DNA to test the influence of community diversity on the reliability of NGS.

Experimental Setup

After analyzing the probability of recovering sequences of different cercomonad species by NGS, different artificial communities were established. These communities were based on a mixture of DNA extracts of different protozoan species to test the effect of species dominance and species richness on the reliability of NGS. Extracts of monoclonal cultures of *Paracercomonas producta*, *Diaphanoeca grandis* and *Spumella sp*. were used for the artificial three-species experiment. Six "artificial communities" were established with different contributions of the DNA extract of each of the three species to the total amount of DNA. Three artificial communities were mixed to represent one of the three species at 96.2% (14.4ng/µl) to total community and correspondingly the other two species each contributed with 1.9% (0.3ng/µl) to total community. Three artificial communities were additionally prepared with 10-times lower concentrations of the underrepresented species. Consequently, one species contributed at 99.6% (14.96ng/µl) and the other two species contributed at 0.2% (0.03ng/µl) to the total community.

To test the influence of the species richness on the reliability of NGS regarding the correct detection of the community structures, the number of species in the artificial communities were increased to twelve species belonging to three different taxonomic groups of flagellates (Choanoflagellida, Chromulinales and Cercomonadida): *Diaphanoeca grandis, Spumella sp*., *Spumella obliqua, Heteromita globosa* type 1, *Heteromita globosa* type 2, *Bodomorpha sp.*, *Paracercomonas astra, Paracercomonas producta, Paracercomonas ambulans, Paracercomonas sp.*, *Cercomonas sp.* type 1 and *Cercomonas sp.* type 2, DNA of these organisms was mixed in equal concentrations of 10ng/µl each.

The artificial three-species and twelve-species communities were established on the basis of mixed DNA extracts of the respective species. Two different PCR approaches were applied (Fig. 1). For the first PCR approach, referred to as PCR product-mix, DNA of the monoclonal cultures was individually amplified and afterwards combined. The concentrations of the PCR products were measured and mixed in the different concentrations as described above. The mix of the individually amplified PCR products was then used as template for the NGS reaction. For the second approach, referred to as DNA-mix, DNA concentrations of each monoclonal culture were measured and pooled to one sample at different concentrations

beforehand, as described above. The pooled mix of DNA was used as template for PCR amplification and the resulting PCR product of the pooled DNA was used as template for NGS.

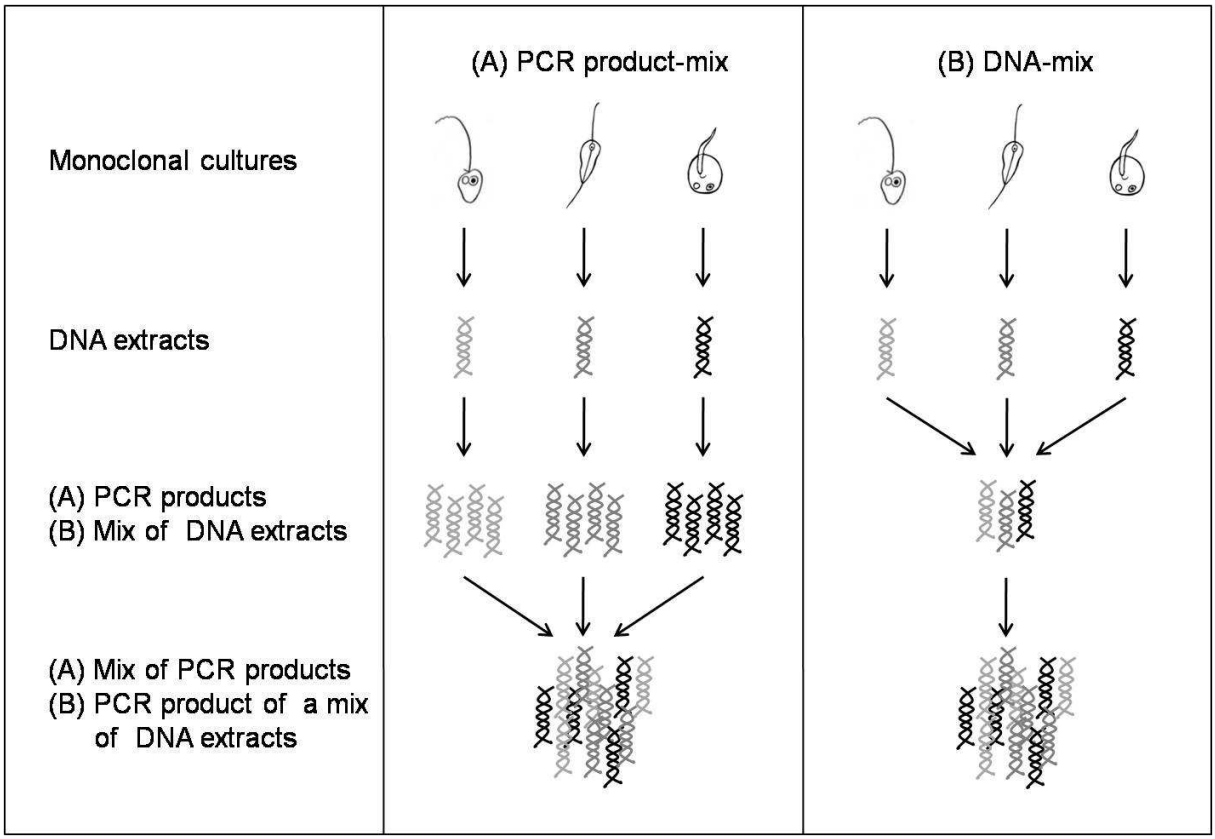


Figure 1. Schematic overview of the two applied PCR approaches. (A) The DNA of monoclonal cultures was isolated and individually amplified. The concentrations of the obtained PCR products were measured and the PCR products were then mixed in defined concentrations to construct the artificial communities which were referred to as PCR product-mix. (B) DNA extracts of each monoclonal culture were measured and combined in defined concentrations. The pooled mix of DNA was then amplified and served as artificial community referred to as DNA-mix. PCR products of the communities of (A) and (B) were used as template for NGS.

DNA Isolation and PCR Treatment

The DNA of monoclonal cultures was extracted by using the cetylmethylammonium bromide (CTAB) method (Clark 1992). The DNA quality was checked by agarose gel electrophoresis (2%). All concentrations of DNA and PCR products were measured using Nanodrop 1000 (Thermo Fisher Scientific, Germany). DNA amplifications were performed in a 50 µl reaction mixture containing two primers at a concentration of 0.1 µM each, all deoxynucleoside triphosphates (dNTP) at a concentration of 200 µM each, reaction buffer, 2 mM $MgCl_2$, 1,5 U Pfu DNA Polymerase (Fermentas, Germany) and 1-5µl DNA as template. Two different primer sets were used for

amplification. For the first amplification, the universal eukaryotic primer set 18SFor (5'-AACCTGGTTGATCCTGCCAGT-3') and 18SRev (5' TGATCCTTCCGCAGGTTCACCTAC-3') for nearly full-length SSU rDNA amplification was used. To enrich the PCR product, a second amplification was performed by using the primer set 42F (5'-CTCAARGAYTAAGCCATGCA-3') and 1450Ri (5'- ATCACAGACCTGTTATTGCC-3'). The integrity of the PCR products was checked by agarose gel electrophoresis (2%), and the PCR yield was measured after purification of the PCR product. All PCR products which were used as template for NGS were amplified as triplets. DNA extracts were sequenced by Sanger sequencing and used as a reference database for the analysis.

Next Generation Sequencing

The PCR products were used as a template for NGS of the variable V4 region of the 18S rDNA. Thus, the universal-eukaryotic primer set 590For (5'-CGGTAATTCCAGCTCCAATAGC-3') and 1300Rev (5'-CACCAACTAAGAACGGCCATGC-3') was chosen for NGS which amplified an about 710nt fragment of the SSU rDNA including the variable V4 region. HPLC purified PCR primers contained at the 5'-end the Roche 454 pyrosequencing adaptors (Adapter A: CCATCTCATCCCTGCGTGTCTCCGAC at the forward primer, Adapter B: CCTATCCCCTGTGTGCCTTGGCAGTC at the reverse primer), key (TCAG at forward and reverse primer) and MID-identifiers at the forward primer, according to the Roche protocol. PCR was carried out by using a 50 µl reaction mixture containing both primers at a concentration of 0.1 µM each, all dNTPs at a concentration of 200 µM each, reaction buffer (including $MgSO_4$), 1,5 U Pfu DNA Polymerase (Fermentas, Germany) and 5µl PCR product as template. PCR products were checked by agarose gel electrophoresis (2%). Concentrations of final PCR products were measured and 10 ng from each PCR product were combined to one sample. The pooled sample was used for NGS by using a 454 GS-FLX Roche sequencer system (GATC Biotech AG, Germany). All samples were sequenced from the forward primer (5' end).

Data Analyses and Statistics

Before analyzing the data a strict quality filtering was performed and all sequences which did not match the following criteria were removed: (i) perfect match to adapter and forward primer sequence, (ii) minimum sequence length of 200nt, (iii) maximum sequence length of 710nt, (v) pairwise identity of at least 99% and (vi) query coverage of at least 80%. The quality filtering eliminated 46.3% of all sequences.

Sequence reads were separated by their specific MID-sequence and assigned to the different approaches by using Geneious Pro 5.5.6 assembler software (Biomatters Ltd.). The sequences of the isolated organisms which were obtained by Sanger sequencing were used to build the reference database for further analysis. NGS sequences were blasted against this reference database using Geneious Pro 5.5.6 assembler software (Biomatters Ltd.).

Data of observed and expected contributions were arcsin-transformed before analysis. Means of observed contributions were compared with expected values by performing a Friedman's Chi^2 test by using R statistical software (http://www.r-project.org/).

Results

Recovery of Isolated Organism Sequences by NGS

Before analyzing the NGS data, all obtained sequence reads were assigned to their specific origin and underwent the quality filtering. Finally, 14,160 sequence reads remained of 25,549 received sequence reads (44.5%) after the stringent quality filtering. The remaining analyzed sequences had a mean length of 387nt.

The number of cercomonad sequences which remained after quality filtering in one of the four soil samples varied between 6 and 23 in the samples. Thus, between 7% and 44% of the cercomonad species could be recovered in the same sample by NGS based on the previous cultivation approach (Fig. 2). Considering the adjusted dataset of the sample, between 0.06% and 0.23% of the sequences were represented by the originally isolated species.

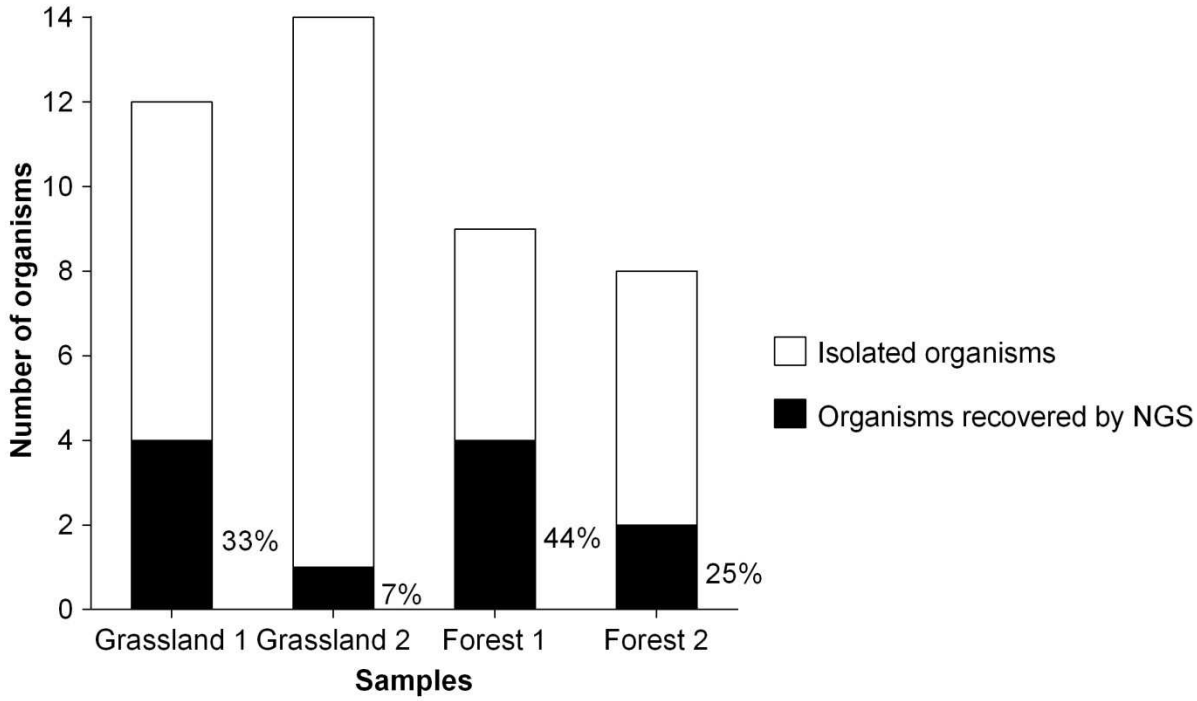


Figure 2. Percentage of isolated species recovered by NGS. Proportion of species recovered by NGS and formerly isolated from the same sample by cultivation methods.

Artificially Composed Communities of Three Species with Varying Relative Contributions

Most of the artificial communities were recovered with the expected contributions of 0.2%, 1.9%, 96.2% and 99.6% after NGS even when a species was present only at very low concentrations (Fig. 3 and Appx. Tab. 3.1). The regression lines of the observed contributions of the sequences did not differ significantly from the regression line of the expected contributions ($F_{[2,42]}$=1.26; P=0.29). This strong correlation between initial contribution and observed contribution indicated the potential to recover dominant and underrepresented taxa. The relative contributions of sequence reads obtained from the mix of PCR products and the mix of DNA were very similar to the expected contributions. The regression line of the sequence reads of the observed DNA-mix approach (R^2=0.914, m= 0.90±0.07) was not as close to the expected regression line as the regression of the observed sequence reads of the PCR product-mix (R^2=0.999, m= 0.97±0.01). This was based on the two outliners of one approach of the DNA-mix. Nevertheless, the differences between both slopes was not significant ($F_{[1,32]}$=1.03, P=0.32).

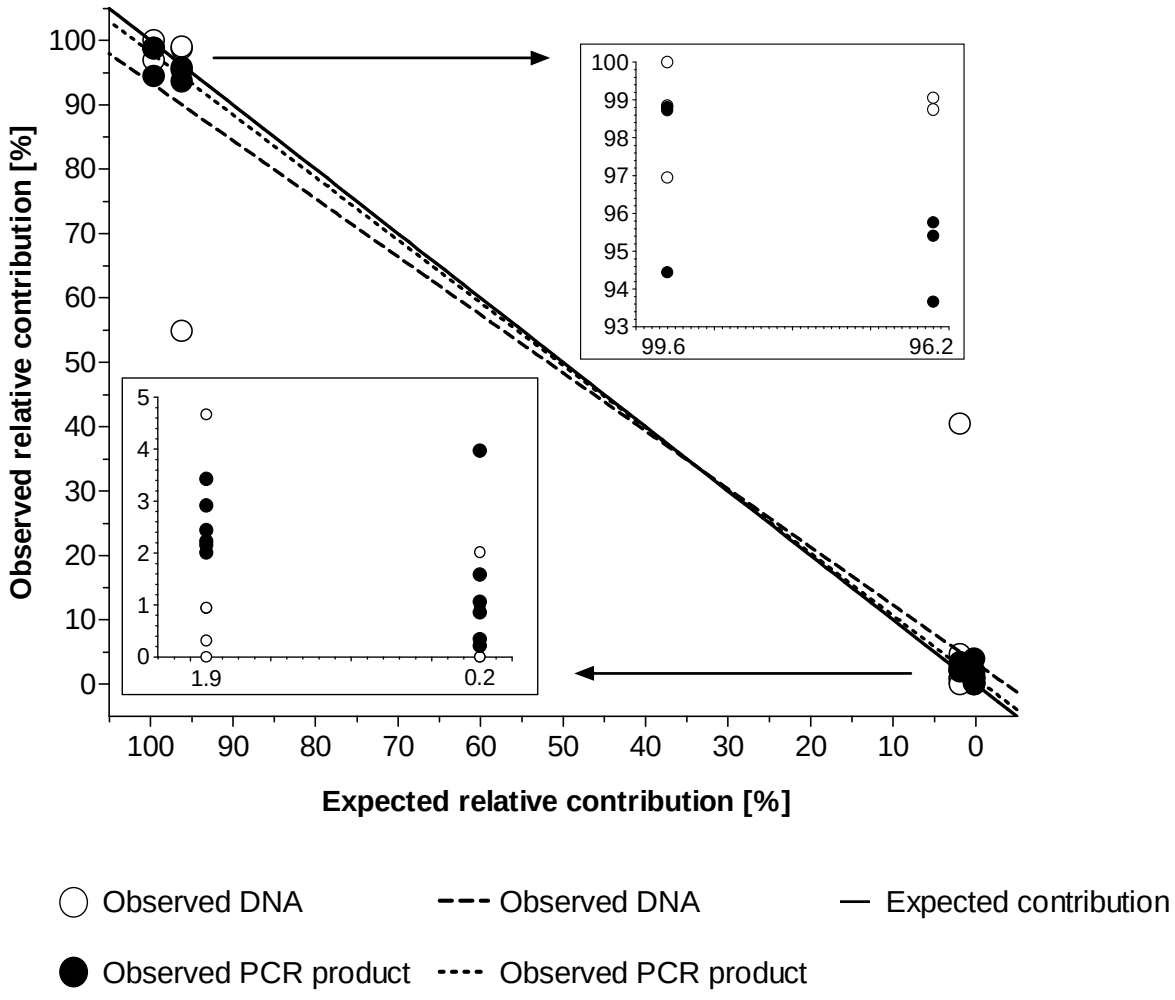


Figure 3. Relative contribution of recovered reads plotted against their expected relative contribution to the total community. Arrows indicate scale up of the respective scatter plot. The solid line represents ideal correlation of observed and expected contributions.

Artificially Communities of Twelve Species with Equal Contributions

Two artificial twelve-species communities were established for the PCR product-mix approach and the DNA-mix approach. The NGS data set included only sequences of one PCR product-mix approach and the two DNA-mix approaches. No sequence of the second PCR product-mix approach could be recovered by analyzing the NGS sequences, which might be due to technical problems.

Of the 12 original species present in each data set, 11 (92%) were recovered for the PCR product-mix approach, but only 6 (50%) were recovered in both DNA-mix approaches (Fig. 4). Only *Spumella obliqua* failed to be detected in all samples. The frequencies of the sequence reads of the PCR product-mix were not significantly

different and in agreement with the expected contributions of the different species (Chi^2=0.82; P=0.37). Most of the detected species had a similar share in the community (median 9.5%).

The observed community structures after NGS based on equally mixed DNA consisted both of 6 species. Both communities were superimposed by reads of *Spumella sp.* with a contribution of around 97% (Fig. 4a). The contributions of the other species after subtraction of *Spumella sp.* were in the DNA-mix 1 significantly different to the initial contributions (Chi^2=4.45; P=0.03). For DNA-mix 2 no significant differences between the contribution of each of the other species and their initial contributions could be obtained (Chi^2=2.27; P=0.13). A closer look on the observed communities of both DNA-mixes after ignoring *Spumella sp.* indicated a higher variability of the underrepresented taxa in these samples compared to the PCR product-mix sample (Fig. 4b).

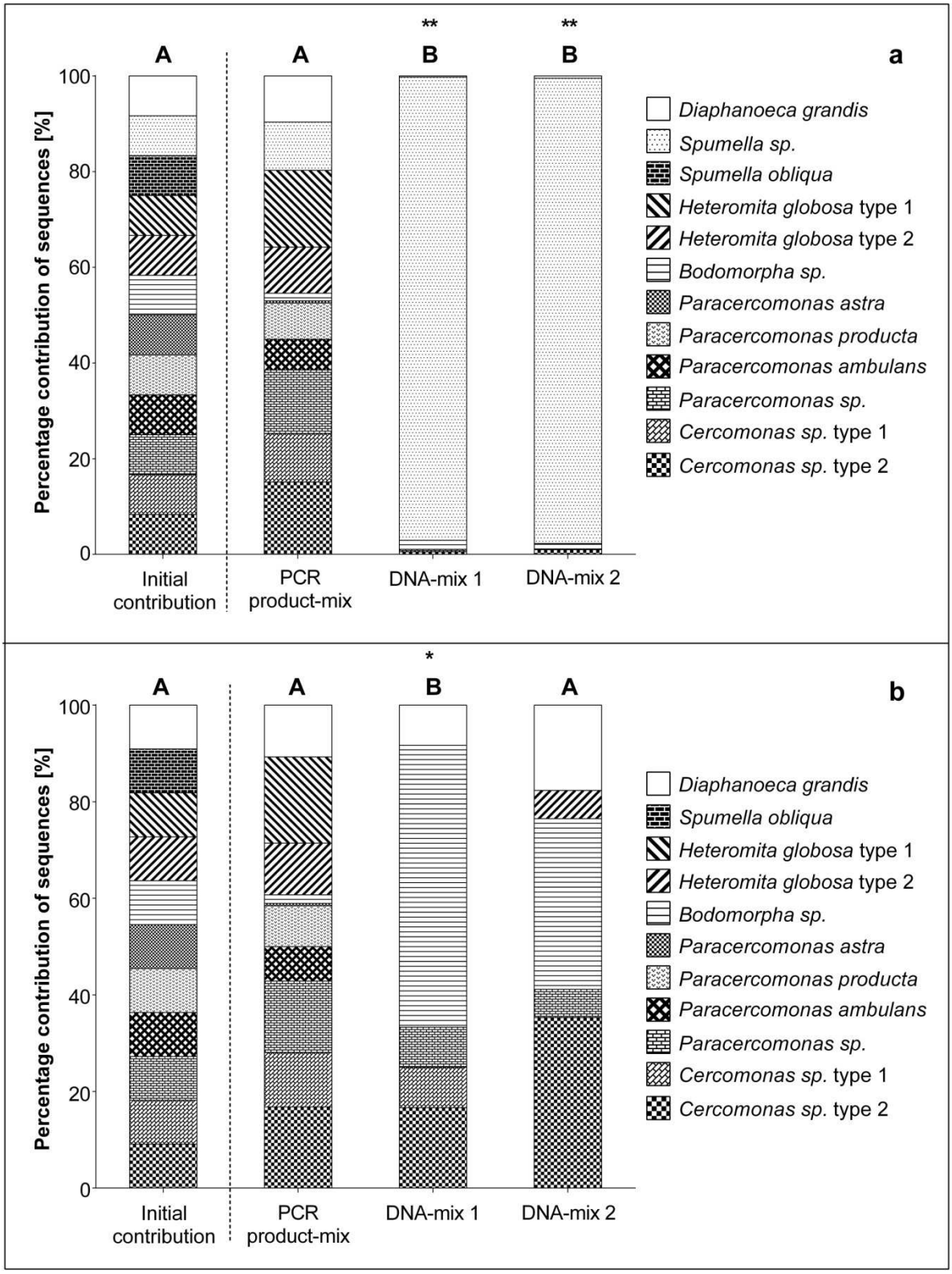


Figure 4. Proportion of sequence reads obtained from the twelve-species communities by NGS. (a) Observed contributions of sequence reads in comparison to the expected initial contributions. (b) Observed contributions of sequence reads after withdrawal of values for *Spumella sp.* in comparison to the expected initial contributions. Arcsin transformed proportions were used for Friedman's Chi^2 test (df=1; α=0.05). Different letters indicate significant differences between the communities, asterisks indicate the different levels of significance: **: p < 0.01; *: p < 0.05.

Discussion

Our results indicated the potential of NGS to analyze protozoan community structures. Organisms which were discovered by isolation and cultivation could also be recovered up to 44% by NGS. This recovery of sequences was investigated for the first time concerning environmental soil samples and flagellate species. The observed contributions of the artificially composed communities were generally in agreement with the initial and expected contributions. Even the underrepresented species with low frequencies were well recovered and in agreement with the expected contributions indicating the potential of NGS to detect even rare organisms. This was true for the three-species communities independent of the used PCR approach.

In contrast to that, a strong bias by initial mixing of PCR products and DNA could be found for the twelve-species communities. Concerning the PCR product-mix, 90% of the individually amplified species were recovered in expected contributions after NGS. The mixing of DNA, followed by amplification of this mix had a strong influence on the outcome of NGS. Neither the number of species, nor the expected contributions were recovered after NGS. This finding support the assumption that a certain bias is constructed by PCR as not all species could be amplified equally good. Amongst the reasons for this unequal amplification pattern can be: (i) a difference in secondary structure of rDNA of different species, e.g. GC content (Strien et al. 2013), (ii) different susceptibility to cell lysis between different species and generally between living cells and cysts (Babaei et al. 2011; Leuko et al. 2008), (iii) different copy numbers of rDNA depending on the individuals' life cycle stage (Gong et al. 2013; Wegener Parfrey et al. 2012) and (iv) species specific copy numbers of rDNA (Gong et al. 2013; Medinger et al. 2010; Zhu et al. 2005). The selected primers were known to amplify the SSU rDNA of all organisms in the artificially composed communities. Thus, a selectivity of the used primers should have been excluded from the experiments. However, the use of a single primer set will always lead to a selective amplification of an environmental sample with an unknown diversity. The selected primers strongly determine which resolution of the community and coverage of different organisms can be expected by analyzing an unknown community structure. A very universal primer set can lead to a broad range of amplified

organisms without covering all organisms. A more specific primer set will lead to highly resolved analysis of only a very small proportion of the whole community. Therefore a combination of primer sets is recommended.

Medinger et al. (2010) compared seasonal fluctuations of protozoan communities based on morphological and respectively molecular determination techniques. Their analysis indicated a strong bias of the used methods. In comparison to morphological determinations, NGS could equally recognize general changes of the protozoan community, but did not reveal a good estimate of species abundances. Even though the differences in their study were mainly due to an over-representation of alveolate sequences, the results indicated the possible pitfalls of NGS based community structures. Also, Weber and Pawlowski (2013) showed a strong influence of rDNA copy number and rRNA expression level of a constructed foraminiferan community on the number of sequence reads by NGS.

Our results based on a soil flagellate community underline these problems of estimating abundances based on NGS data. When the community was enlarged, the initial and observed communities of the DNA-mix were not consistent anymore; though, especially the DNA-mix approach reflected a typical unknown community analyzed by NGS. Possible PCR biases and technical errors cannot be considered in an unknown community and might lead to a misinterpretation of obtained species abundances. Nevertheless, the protozoan species which were used to create the artificially composed communities all belonged to flagellates in a small size range of 5-10µm. Consequently a strong bias of rDNA copy number based on a certain species size (Weber and Pawlowski 2013; Zhu et al. 2005) and taxonomic group should be as small as possible.

In general, the use of NGS provides huge datasets which are intended to characterize different kinds of community structures at a high level of taxonomic resolution. In our study, however, only a small proportion of field samples consisted of sequences formerly obtained from cultivated organisms from the same soil sample. Based on the large number of sequences obtained by NGS, a higher consistency would have been expected. Both methods seem to favor, respectively select certain species and reveal the community structure in a different light. In one case, selectivity is based on the cultivability of species under the conditions applied

and in one case selectivity is due to the selective amplification of organisms by the NGS method.

Our studies indicated that the NGS technique is an appropriate tool to analyze community structures and to compare species richness from different samples. The enlargement of the communities led to a change in the number and the contribution of recovered organisms. This was especially true for both DNA-mixes. The influence of PCR amplification was the main predictor of these differences. Thus, general information about the community structure can be obtained by NGS, but total diversity is difficult to estimate.

Summary

Protozoa are ubiquitous and essential components of nutrient cycling processes in a wide range of different aquatic and soil environments. Therefore the determination of protozoan abundances and species richness play an important role for the understanding of the nutrient cycling processes and ecosystem function. Next generation molecular techniques are becoming increasingly important over the last years, as they enable a high through-put of samples with a high output of data. Yet, it is still unclear how artifacts from the sequencing and analyzing process affect the outcome of the obtained community. The objective of this study was to test the influence of protozoan community composition (including environmental soil samples), based on the species diversity and species richness on the outcome of a given artificially composed community. Furthermore the influence of PCR amplification was considered in this study. Artificial communities of identified species were combined in defined concentrations before and after PCR amplification, resulting in different community approaches for NGS. The species proportion of a small community after NGS was reflected just as initially composed, regardless whether the species was dominating or underrepresented.

For twelve-species communities a strong bias of PCR amplification was observed. The different species were amplified differently well resulting in a change of the initially composed community. The recovery of flagellate sequences from field samples was analyzed. Since next generation sequencing (NGS) is a PCR based tool, errors in sequence reads and outcome can arise from DNA extraction, PCR amplification and data analysis, including the levels of quality filtering and discrimination of operational taxonomic unit (OTU). Based on the results, NGS can lead on assumptions of species richness but do not reveal reliable estimates of total species diversity.

Chapter 4 – Protistan soil communities revealed by NGS: A detailed local study supported by morphological investigations

Introduction

Protists are unicellular eukaryotic organisms (Adl and Gupta, 2006), compromising autotrophic, heterotrophic and mixotrophic organisms. They are globally distributed and inhabit all types of aquatic and terrestrial environments (Ekelund and Ronn 1994; Finlay 2002; Sherr and Sherr 2002). The protistan biogeography is still unexplained and contrasting positions about their distribution are discussed, as a ubiquitous dispersal in suitable habitats (e.g. Finlay and Fenchel 1999) and a more restricted dispersal with endemic species (e.g. Foissner 1999a) is argued. Protists are able to form small cysts (Rivera et al. 1992) and occur in high abundances (Wilkinson et al. 2012). Taken together, these abilities allow an unrestricted air dispersal (Wilkinson 2001). Dispersal by air is characteristic for soil protists, such as flagellates, testate amoebae and small ciliates (Altenburger et al. 2010; Wanner and Xylander 2005). Consequently, protist distribution is mainly restricted to the environmental conditions found at the habitat they migrate to. The ability to colonize and reproduce within new habitats is thereby mainly defined by the given conditions. Thereby, protists are dependent on various conditions, like soil moisture, temperature and food abundance (Adl and Coleman 2005). The dependency on those conditions, their contribution on nutrient cycling processes (Bonkowski 2004) and their short generation time (Foissner 1999b) make them very suitable as indicators of abiotic and biotic changes. Thereby environmental changes are reflected by differences in the protistan abundance, community structure and diversity, as certain species are adapted to different environmental conditions, where they are metabolically active and reproduce.

A significant influence of land use intensity on the protistan abundance was revealed in a preliminary study (Domonell et al. 2013, Chapter 1). Thus, the influence of land use intensity and the site type (grassland and forest) on the protozoan soil community was investigated by cultivation and morphological determination. The results revealed high abundances, significantly different in grassland due to land use intensity, and a high diversity of heterotrophic protists in soil, where flagellates and amoebae contributed equally to total protistan community. On the other hand, the protozoan community structures were independent on the site type, land use intensity and various soil properties.

However, morphological studies on protists are generally restricted by the low abundances of certain species, encysted resting stages and methodological biases based on selective determination and enrichment (Weisse 2008). The use of high-throughput methods can evade these morphological determination restrictions and additionally provide larger datasets, which might enable a high resolution of community structures. Hence, next generation sequencing (NGS) was applied on the preliminary investigated sites to improve the influence of land use intensity and site type on protistan soil. Typical soil organisms as cercozoans, amoebae and ciliates (Bates et al. 2013; Comeau et al. 2013; Domonell et al. 2013; Esteban et al. 2006; Harding et al. 2011; Urich et al. 2011) were expected to dominate the protistan soil community. Furthermore, additional species were expected to be discovered, which would have been undetected when applying culture based methods. This applies especially to ciliates which were only rarely detected in quatitative cultures (see Chapter 1), but generally common in soil and widely used as soil bioindicators (Foissner 1999b). The higher resolution of protistan species richness, due to the large NGS dataset, has the potential to enlarge and refine the assumptions made by the morphological investigations. Furthermore, the detailed analysis might identify (i) protistan communities, which are dependent on the site type, land use intensity and certain soil properties, indicated by species site restriction or (ii) protistan communities, which are independent on site type, land use intensity and soil properties, indicated by similar communities in different soil sites.

The one selected site in the Hainich was used as a model site to apply first detailed analysis of protistan soil communities down to species level by NGS. The expected high resolution of the community structures will provide in-depth data, where other soil associated data (obtained from other studies in the framework of the biodiversity exploratories) can be correlated to. The morphological and further molecular studies (Chapters 1 and 2) revealed first insights into the community composition and structure, which is required for an appropriate setup of culture independent NGS analysis.

Methods

Study area and soil sampling

The soil sampling took place in 2009 at the Hainich-Dün region (Thuringia, Fig. 1), which is located in central Germany (N 51°13'0", E 10°27'0"). The region is characterized by a limestone plateau (300-400 m a.s.l.) with an annual mean temperature of 6.5-7.5 °C and annual mean precipitation of 750-800 mm (Fischer et al. 2010). The experimental plots were chosen by extremes of a land use gradient in established land use systems as part of an interdisciplinary project, the Biodiversity Exploratories (http://www.biodiversity-exploratories.de). Within the present study, protistan community structures of different sampling sites (four grassland sites and four forest sites) were compared. The sites were chosen according to their degree of land use intensity (for details about the sites and their labeling, see Table 1 and Chapter 1) and correspond to the samples from 2008 which were analyzed in Chapter 1. Two replicates of each low (LI) and high (HI) intensively used grassland (G) and forest (F) site were taken. As the replicates had similar land use intensities and were near-by, the replicates were summed up for further analysis. The sampling procedure and soil sample handling followed the protocol of Chapter 1.

Soil properties

Soil water content was determined by subtraction of soil dry weight from soil fresh weight, after drying at 105°C for 24h. The soil organic material was determined by the loss of ignition after burning at 530°C for 12h. The C/N was calculated from dry soil by using a CN-Analyser (peformed by Conny Thielen). The soil pH values were measured in 0.01 mol l^{-1} CaCl2 using a 1:2.5 w/w soil:solution ratio (performed by Fabian Alt). Bacterial cell numbers were measured by fluorescent staining and counting. Cells were fixed with 2% glutardialdehyde directly after soil collection and stained in the lab with SYBRGreenI. The cells were than counted on black PC filters on glass slides with a fluorescent microscope (Zeiss) using a 100x objective and a net-micrometer. The number of bacterial cells was determined by Verena Nägele).

DNA Isolation, PCR amplification and next generation sequencing of 18SrDNA

The DNA was extracted from a 1g subsample of the dried composite soil using the MoBio PowerSoil DNA extraction kit (MoBio Laboratories, USA). The DNA and PCR product quality was checked by agarose gel electrophoresis (2%) and all concentrations of DNA and PCR products were measured using Nanodrop 1000 (Thermo Fisher Scientific, Germany). PCR amplification and next generation sequencing (NGS) preparation followed exactly the protocol described in Chapter 3. This protocol was developed for the analysis of the highly variable V4 region of the 18SrDNA. The outcoming 710bp long fragment can be amplified from most eukaryotic groups, enabling detailed analysis down to species level over a wide range of eukaryotes groups. PCR reactions were carried out as triplets and pooled to one sample before NGS to minimize PCR biases.

The pooled samples were used for NGS by using a 454 GS-FLX Roche sequencer system (GATC Biotech AG, Germany). All Samples were sequenced from the forward primer (5' end).

Data analyses and statistics

The obtained raw dataset was filtered based on the following criteria, excluding all sequences from the analysis which did not match these criteria: (i) perfect match to adapter and forward primer sequence, (ii) minimum sequence length of 200nt, (iii) maximum sequence length of 710nt, (v) query coverage of 100% (vi) at least two assigned sequences in one sample to confirm a taxa. Sequence reads were separated by their specific MID-sequence and assigned to the corresponding sampling sites. The sequences were analyzed by using Geneious Pro 6.1 assembler software (Biomatters Ltd.). Different taxonomic levels were defined for levels of super-group/group (≥ 80% sequence identity), genus (≥ 98% sequence identity) and species (≥ 99.7% sequence identity). Additionally, sequences of species were assigned at 100% sequence identity for a high resolution of species diversity.

Taxonomic identifications were assigned by using Blast (Altschul et al. 1990) against the quality filtered Protist Ribosomal Reference database PR^2 (Guillou et al. 2013), which follows the current eukaryotic classification of Adl et al. (2012). The different

levels for taxonomic identification were defined according to recent studies using OTU (operational taxonomic unit) clustering for NGS data analysis. 80% identity was used as proxy for super-groups and groups (narrowing the 80% identity used for class identification by Massana et al. (2011) and Stoeck et al. (2010)), 98% identity was used as proxy for genera (Bachy et al. 2013) and 99.7% identity was used as proxy for species boundaries (narrowing the 99% identity used for species identification by Massana et al. (2011) and Salani et al. (2012)). The boundaries for certain taxonomic levels were narrowed, as OTU clustering thresholds are usually made very carefully to avoid overestimates of OTU numbers per taxonomic level (Huse et al. 2010; Kunin et al. 2010). Protistan super-groups were named according to Adl et al. (2012) and the following proximate taxonomic level was named as protistan group. Taxa were confirmed if at least two sequences were assigned, excluding singletons. Sequence analyses were performed by using Geneious Pro 6.1 assembler software (Biomatters Ltd.). A rough classification was performed for all eukaryotes and the subsequent detailed analysis was focused mainly on the unicellular protistan taxa (excluding fungi).

Obtained data were transformed before statistical analyses to approximate homogeneity of variance, as protistan species abundances were log(X+1) transformed and species proportions were arcsin-transformed. Protistan species abundances were calculated form the number of sequence reads representing each species. Cluster analyses and ANOSIM analysis (one-way ANOVA) of the different samples were performed by using Primer 6 software package (Clarke and Gorley 2006). Distance matrices were calculated based on the Bray-Curtis similarity using unweighted matrices based on protistan species presence/absence transformed data and weighted matrices based on protistan species abundance.

Spearman rank correlation was used for quantification of species log(X+1) transformed abundance and species presence to soil properties. Two-way ANOVA analyses, Spearman rank correlation, Shapiro-Wilk test of normality and paired t-test were performed by using R statistical software (http://www.r-project.org/).

The diversity was calculated with the Shannon-Weaver index (H) (Shannon and Weaver 1949) using the relative abundance of sequence reads representing the taxa on a certain identity level, by:

$$H=-\sum_{i} p_i \cdot \ln p_i$$

, where p_i is the proportion of the i-th morphotype.

Additionally the evenness of the taxa was calculated by:

$E=H/\ln S$, where S is the total number of identified taxa on a certain level.

Results

A total of 715.001 sequences was recovered from the eight different soil samples. The number of 540.513 eukaryotic sequence reads remained after the stringent quality filtering, ranging from 7.018 (LIG1) to 206.584 (HIF1). The remained analyzed sequences had a mean length of 426 nt. The proportion of sequences which could not be assigned to at least 80% sequence identity varied between 15.9% (LIG2) and 26.5% (HIG2). At a level of 100% sequence identity, up to 6.410 (HIF1) eukaryotic sequences and up to 517 (HIG2) protozoan sequences could be assigned, which corresponded to 3.1% respectively 0.3% of total eukaryotic sequences. More information about the number of assigned sequences and metadata of the sampling sites are shown in Table 1.

The soil community analyses were performed on eight different sites with different land use intensities (low and intensive) and different habitats types (grassland and forest). The detailed analyzed sites were LIG (low intensively used grassland), HIG (high intensively used grassland), LIF (low intensively use forest) and HIF (high intensively used forest). Each site was replicated, referred to as 1 and 2 (see Tab. 1), summed up for the general comparisons of habitat types. The sum up was constituted by clustering of the replicates due to species presence (see Appx. Fig. 4.1). The community structures of the different soil samples were analyzed based on three, respectively four different levels of sequence identity: 80% identity as proxy for super-group/group, 98% identity as proxy for genera, 99.7% identity as proxy for

species boundaries and additionally 100% as proxy for distinct species boundaries. Based on the different identity levels, a total of 22 eukaryotic and 17 protistan groups (80% identity), 361 eukaryotic and 111 protistan genera (98% identity), 134 eukaryotic and 40 protistan species (99.7% identity) and 61 and 31 protistan distinct eukaryotic species (100% identity) were identified.

Community comparisons based on 80% sequence identity

At this level, the eukaryotic communities were dominated by fungi and representatives of the SAR group (Stramenopiles, Aleveolata, Rhizaria), see Figure 1. However, the five protistan super-groups included 13 different groups: Opisthokonta (Choanomonada, Mesomycetozoa), SAR (Stramenopiles, Alveolates, Rhizaria), Amoebozoa (Lobosa, Conosa, Breviatea, undetermined. Amoebozoa), Archaeplastida (Chlorophyta), *Insertae sedis* Eukaryota (Apusomonadida) and Excavata (Discoba, Malawimonadidae) (Fig. 2).

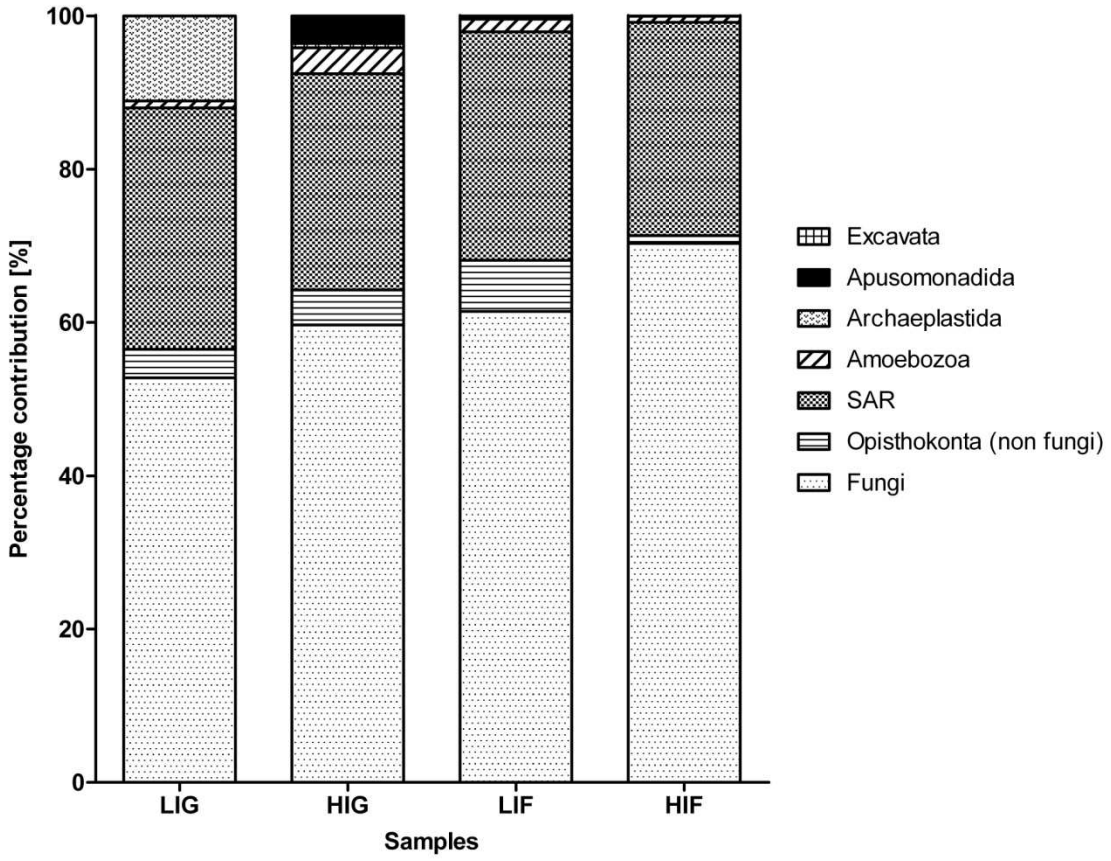


Figure 1. Relative abundances of eukaryotic super-groups grouped by 80% sequence identity.

The protistan community was dominated by the SAR group, with different contributions of Stramenopiles (41.4% HIF), Cercozoa (34.2% HIG), Apicomplexa (22.3% HIF) and Ciliophora (11.2% HIG) (Fig. 2). The highest number of protistan groups was found in HIG, which included 16 groups, followed by HIF with 14 groups. Both low intensively used sites included only 10, respectively 11 groups. Appreciable contributions of Choanomonada (4.7%) and Hilmonadea (9.1%) were identified only in the HIG sample, as both taxa were only of minor importance in the other samples. The higher classification to at least genus level led to a loss of these rarely recorded groups.

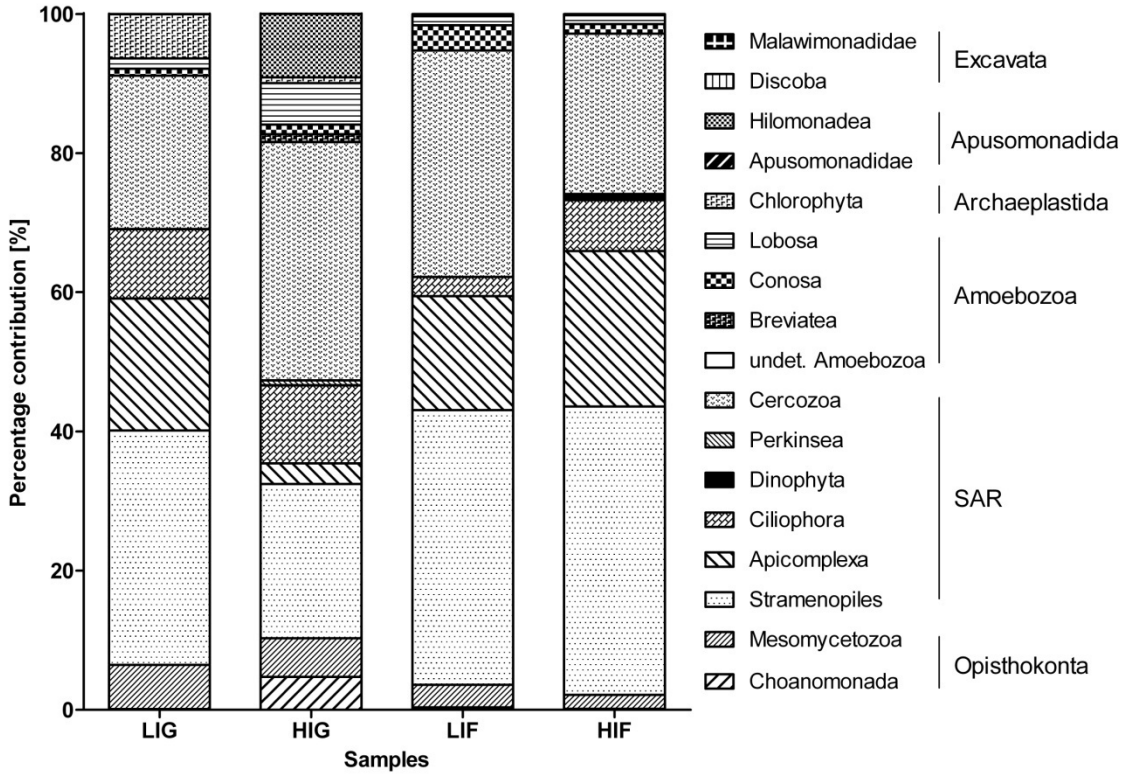


Figure 2. Relative abundances of protistan groups assigned at 80% sequence identity.

Community comparisons on 98% sequence identity

The highest number of eukaryotic genera was identified in HIF (179), combined with the lowest diversity and thus a high proportion of rare taxa. The by far highest number of protistan genera was found in the LIG (70) compared to the other grassland sample HIG (44) and both forest samples LIF (44) and HIF (45). As on group level, the protistan communities were dominated by genera of the SAR group.

Table 1. Number of assigned sequences at different taxonomic levels and additional metadata information of the different sampling sites.

	LIG1	LIG2	HIG1	HIG2	LIF1	LIF2	HIF1	HIF2
N 454-reads total	8.005	10.075	71.262	242.461	9.929	8.261	283.743	81.265
Total eukaryotic high-quality tags	7.018	8.647	58.784	179.530	8.805	7.232	206.584	63.913
Assigned tags 80% id.	5.816	7.269	44.427	131.939	7.159	5.511	171.364	47.928
Assigned tags 80% id. (protists)	2.099	2.457	14.230	51.678	2.239	1.777	55.729	7.190
Assigned tags 98% id.	3.196	4.106	14.123	48.164	4.317	2.503	111.979	21.209
Assigned tags 98% id. (protists)	954	1.194	5.232	15.188	1.208	1.058	33.703	2.833
Assigned tags 99.7% id.	211	250	1.024	5.680	331	238	18.357	1.209
Assigned tags 99.7% id. (protists)	75	76	416	1.847	142	171	4.712	106
Assigned tags 100% id.	50	43	256	1.597	78	70	6.402	895
Assigned tags 100% id. (protists)	23	27	133	612	46	58	1.566	104
Unassigned tags	1.202 (17.1%)	1.378 (15.9%)	14.356 (24.4%)	47.591 (26.5%)	1.646 (18.7%)	1.720 (23.8%)	35.220 (17.0%)	15.985 (25.0%)
Location	RW 4388100 HW 5649700	RW 4390100 HW 5652600	RW 4389180 HW 5683280	RW 4389700 HW 5683010	RW 4396800 HW 5692270	RW 4388100 HW 5664300	RW 4382900 HW 5673600	RW 4386200 HW 5676270
Soil Type	Brown soil	Pelosol	Pseudogley	Pseudogley	Parabrown soil	Parabrown soil	Pseudogley	Parabrown soil
Land use	Meadow fertilized (mown 3 times)	Meadow fertilized (mown 3 times)	Cattle pasture unfertilized	Cattle pasture unfertilized	Beech Plenter forest, very late succession	Unfarmed beech forest, late succession	Spruce forest early succession	Spruce forest early succession
Soil water content [%]	33.98	28.32	27.19	23.82	41.58	30.20	41.62	29.70
pH	6.1	7.11	6.39	6.26	6.42	4.2	5.88	5.38
Bacteria $*10^9$ Ind./g	12.09	11.62	14.95	20.08	17.41	21.83	9.99	13.58
C/N	10.11	14.24	10.85	10.39	13.38	12.79	13.327	14.38

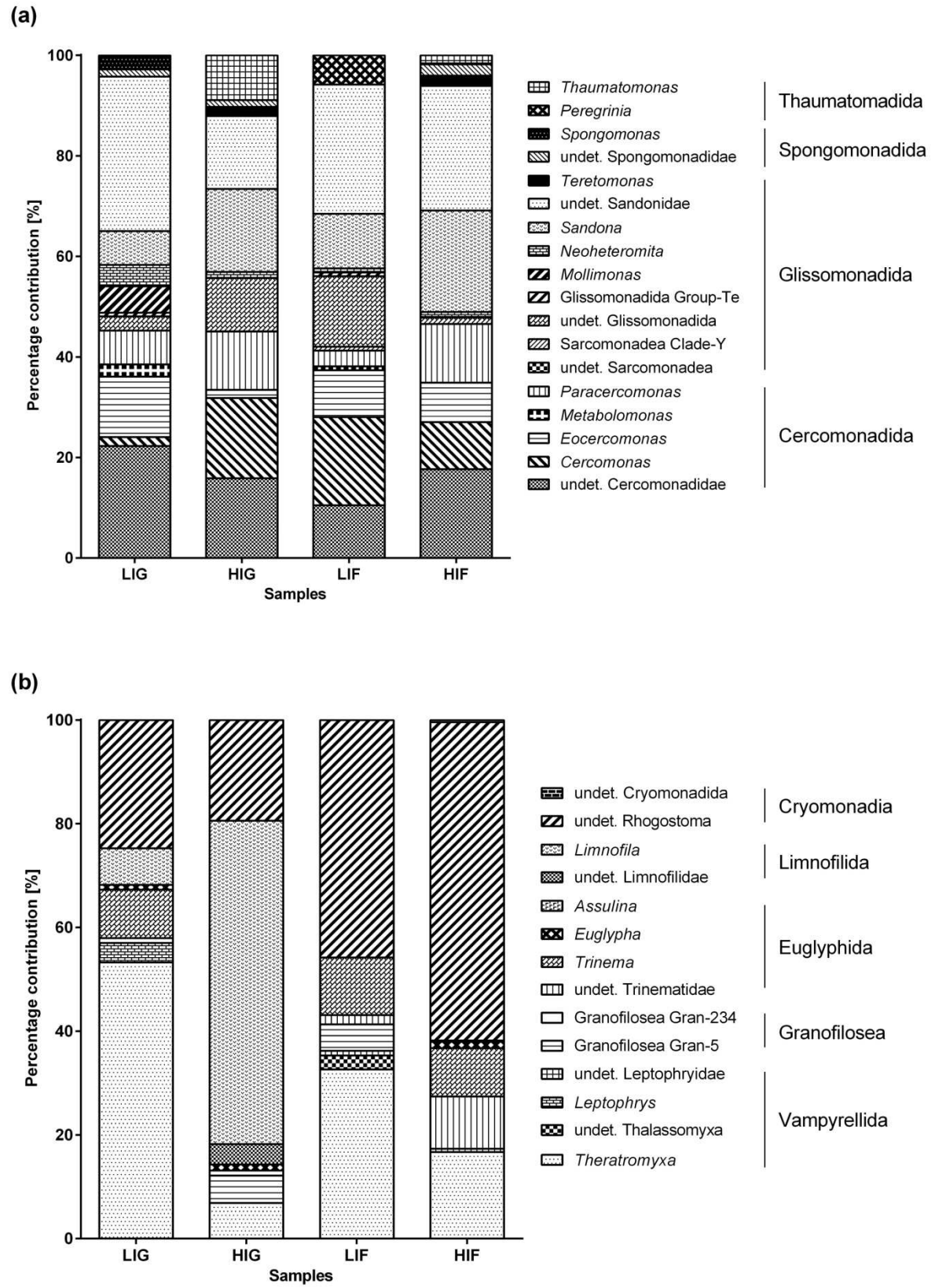


Figure 3. Relative abundances of (a) flagellated and (b) amoeboid cercozoan genera grouped by 98% sequence identity.

Stramenopiles had highest proportion, where cercozoan and ciliphoran genera showed highest richness and were analyzed in more detail. A total of 32 cercozoan genera could be identified, from which 18 belonged to flagellated cercozoans (Fig. 3a) and 14 to amoeboid cercozoans (Fig. 3b). The presence (Fig. 4) and abundance (number of sequences) (Appx. Fig. 4.2) of different flagellated cercozoans was influenced by different intensity of land use.

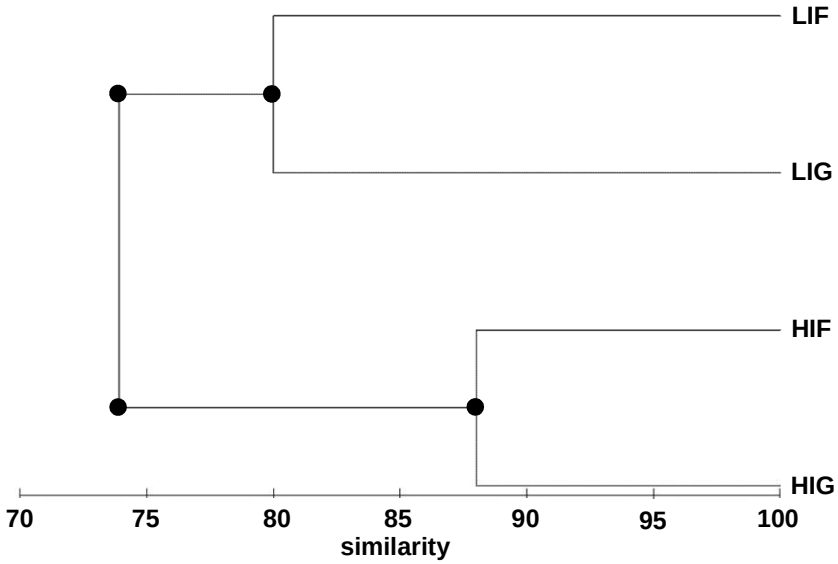


Figure 4. Similarity of flagellated cercozoan soil communities in grassland and forest. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weighted matrix based on genera presence/absence transformed data. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.

As the presence of amoeboid cercozoans differed between grassland and forest sites (Fig. 5), the abundance (number of sequences) was influenced by the intensity of land use (Appx. Fig. 4.3). Genera of glissomonads (44.7-52.9%) and cercomonads (41.2-46.5%) dominated the flagellated cercozoan soil community very constantly across the different sites. The amoeboid cercozoan community was dominated by *Theratromyxa* and two granofilosean genera, *Limnofila* and undetermined. Rhogostoma. Other amoeboid genera were of minor importance.

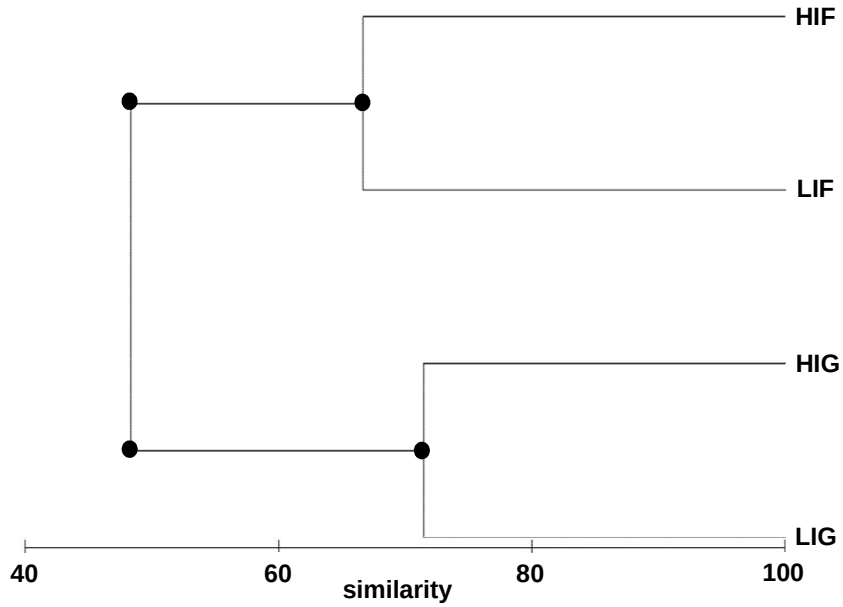


Figure 5. Similarity of amoeboid cercozoan soil communities in grassland and forest. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weighted matrix based on genera presence/absence transformed data. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.

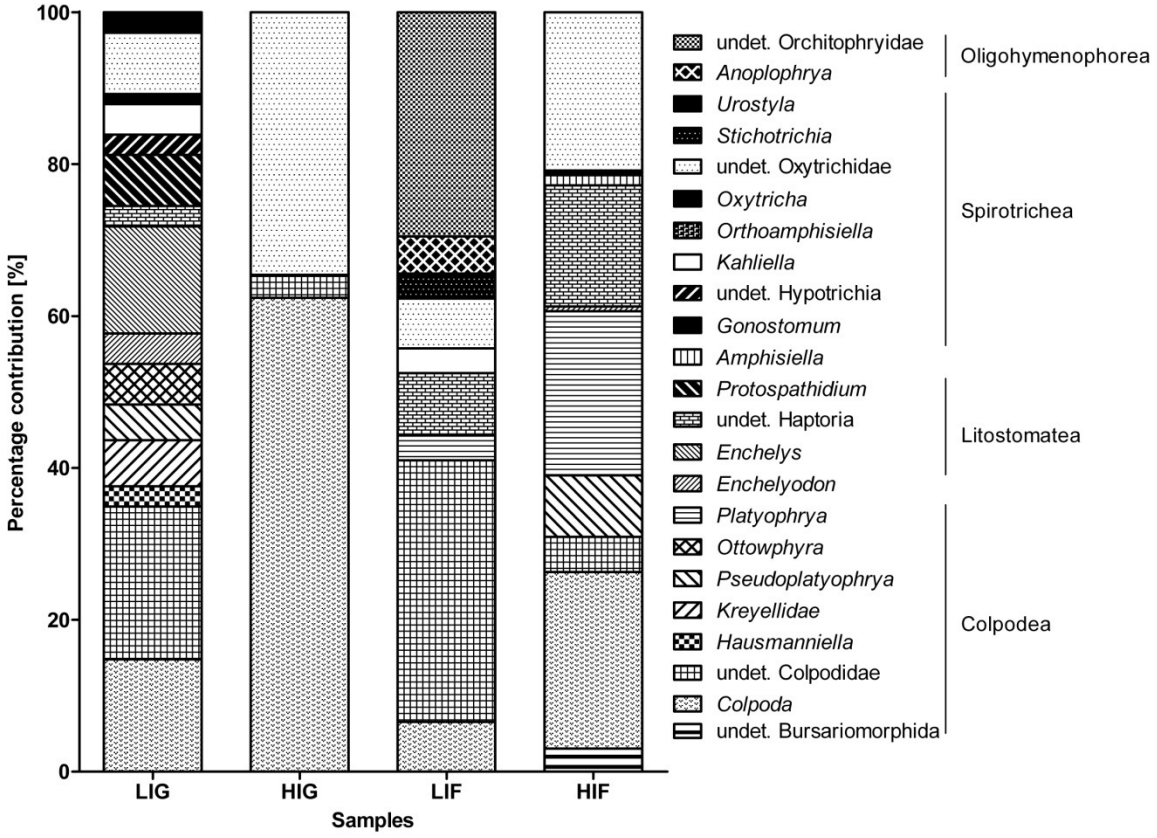


Figure 6. Relative abundance of ciliophoran genera grouped by 98% sequence identity.

The ciliate community indicated high variations between the different sites (Fig. 6). The most diverse community was found in LIG (15 genera), compared to five genera in HIG (Fig. 7). Nevertheless, all communities were dominated by colpodean genera, with contributions of up to 65% (HIG), followed by Spirotrichea (only represented by the order Hypotricha) with up to 35%.

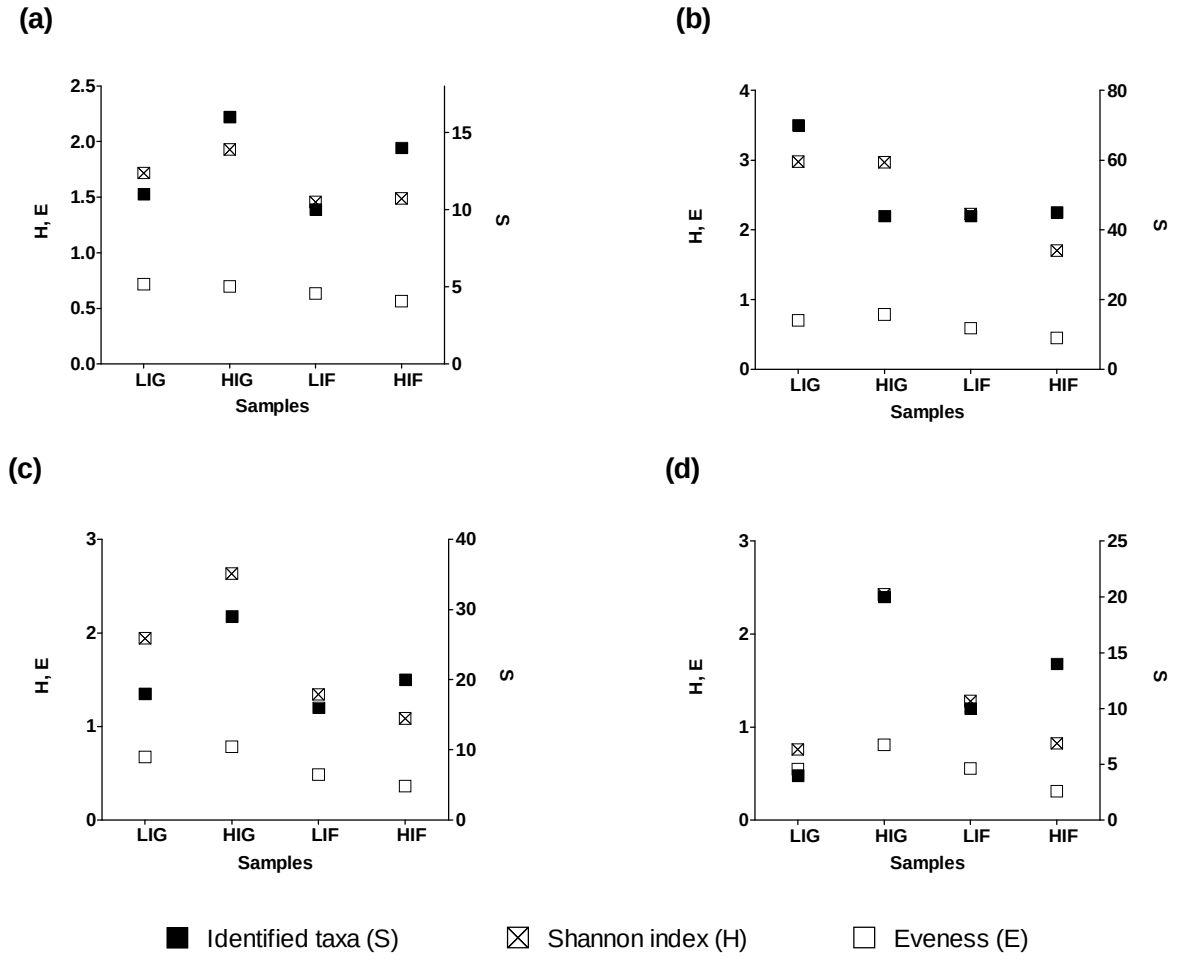


Figure 7. Calculated number of identified protistan taxa (S), Shannon diversity (H) and Evenness (E) on different levels of sequence identity, as 80% identity was proxy for (a) group, 98% identity was proxy for (b) genus, 99.7% identity was proxy for (c) species and 100% was proxy for (d) distinct species. Taxa which were assigned only once per sampling site were excluded from the analysis.

Community comparisons on 99.7% and 100% sequence identity

The highest number of identified eukaryotic species was found in both high intensively used sites, as HIG included 70 and HIF 71 species, from which nearly half of the species were protists. The protistan species abundance differed significantly between the low and high intensively used sites (R=0.43; P=0.029), mainly caused

by variability's of cercozoan species (Appx. Tab. 4.1). High abundances of undetermined Rhogostoma ($F_{[1,4]}$=10.47; P=0.032 for land use intensity), undetermined Sandonidae ($F_{[1,4]}$=24.21; P=0.008 for land use intensity) and undetermined Trinematidae ($F_{[1,4]}$=21.62; P=0.010 for land use intensity, site type and land use intensity x site type) in high intensively used sites were mainly responsible for the difference. The presence of all three taxa was highly significant influenced by the land use intensity, site type and the interaction of both ($F_{[1,4]}$<0.0001; p<0.0001). At the species level, the presence and abundance of the different protists correlated with different soil probabilities (Appx. Tab. 4.2). No specific soil property influenced more than one phylum of protists, resulting in no general correlation of soil properties on species richness and abundance.

In General, 31 of the 55 protistan species were heterotrophic protists from which 22 were cercozoans (Tab. 2). The highest number of heterotrophic protists was found in both high intensively used sites, HIG (21) and HIF (16), compared to LIG (four) and LIF (eight). Thus, the highest number of heterotrophic protists and additionally the highest contribution of heterotrophic protists to total protistan community (86.2%) were found in HIG. The other samples had lower contributions, such as LIG (20.6%), LIF (24.8%) and HIF (12.2%).

Table 2. Enumeration of protistan community based on the relative contribution of determined protistan species. Species which were assigned only once per sampling site were excluded from the analysis. Shades of grey indicate the relative contribution of each determined species

Group	Class/order	Species	LIG	HIG	LIF	HIF	G-Soil
Stramenopiles	Synurophyceae	Clade-C sp.					
		Clade-H sp.					
		Pedospumella encystans					
		Spumella sp.					
	Hyphochytrydiales	*Hyphochytrium catenoides*					
	Oomyceta	*Achlya apiculata*					
		undet. Oomyceta					
		Pythiopsis cymosa					
	Xanthophyceae	*Chlorellidium tetrabotrys*					
Apicomplexa	Gregarines	undet. Gregarines					
Ciliophora	Colpodida	*Colpoda sp.*					
		undet. Colpodidae					
	Haptoria	undet. Haptoria					
	Hypotrichia	*Amphisiella magnigranulosa*					
		undet. Oxytrichidae					

Table 2 continued.

Group	Class/order	Species	LIG	HIG	LIF	HIF	G-Soil
Cercozoa	Vampyrellida	*Theratromyxa weberi*					
	Granofilosea	undet. Granofilosea Gran-5					
	Limnofilida	*Limnofila anglica*					
		Limnofila longa					
		undet. Limnofilidae					
	Euglyphida	undet. Trinematidae					
	Thaumatomonadida	*Thaumatomonas sp.*					
	Cercomonadida	undet. Cercomonadidae					
		Cercomonas pellucida					
		Cercomonas media					
		Cercomonas sp.					
		Eocercomonas expoloratorii					
		Eocercomonas ramosa					
		Metabolomonas insania					
		Paracercomonas compacta					
		Paracercomonas sp.					
		Paracercomonas vonderheydeni					
	Glissomonadida	undet. Allapsidae					
		Flectomonas ekelundi					
		undet. Glissomonadida					
		Sandona sp.					
		undet. Sandonidae					
	Cryomonadida	undet. Rhogostoma					
Conosa	Dictyosteliida	*Dictyostelium sphaerocephalum*					
Chlorophyta	Chlorophyceae	*Bracteacoccus cohaerens*					
		Bracteacoccus sp.					
		Pseudotetracystis sp.					
		Tetracystis vinatzeri					
	Chlorellales	undet. Chlorellales					
		Muriella sp.					
		Myrmecia bisecta					
		Nannochloris sp.					
	Microthamniales	undet. Microthamniales					
		Myrmecia astigmatica					
		Myrmecia sp.					
	Prasiolales	*Diplosphaera sp.*					
		undet. Prasiolales					
	Trebouxiophyceae	*Elliptochloris subsphaerica*					
		undet. Trebouxiophyceae					
Mesomycetozoa	Ichthyosphonida	*Anurofeca sp.*					

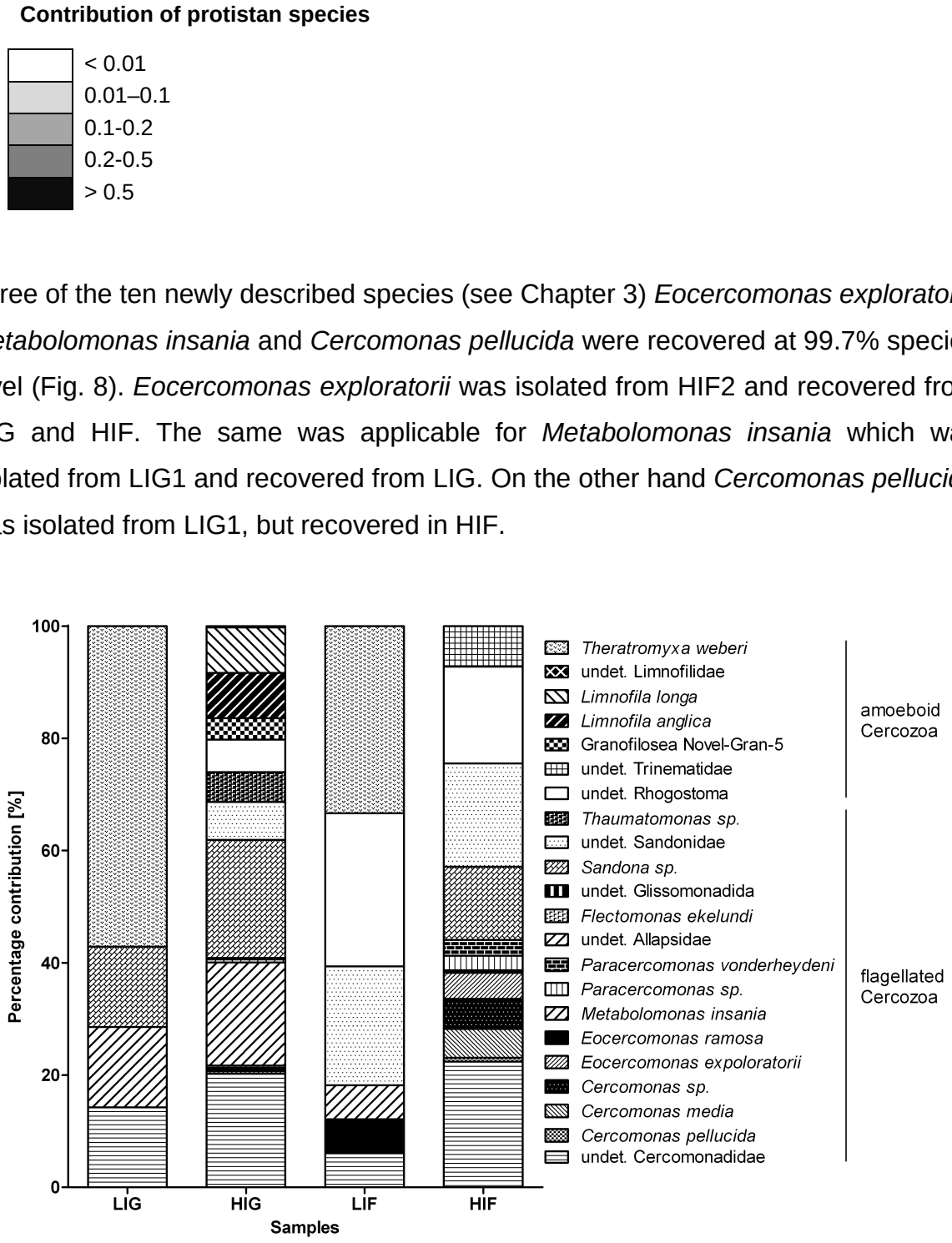


Figure 8. Relative abundance of cercozoan species grouped by 99.7% sequence identity.

Three of the ten newly described species (see Chapter 3) *Eocercomonas exploratorii*, *Metabolomonas insania* and *Cercomonas pellucida* were recovered at 99.7% species level (Fig. 8). *Eocercomonas exploratorii* was isolated from HIF2 and recovered from HIG and HIF. The same was applicable for *Metabolomonas insania* which was isolated from LIG1 and recovered from LIG. On the other hand *Cercomonas pellucida* was isolated from LIG1, but recovered in HIF.

As the number of cercozoan genera was constant between the different sites (19-20), the number of cercozoan species was low in the low intensively used sites, LIG (four) and LIF (six) and increased in the intensively used sites HIG (15) and HIF (12) (Fig. 8). However, the contribution of flagellated and amoeboid cercozoans also differed

between low and intensively used sites, as the high intensively used sites were dominated by 75% of flagellated cercozoan species and low intensively used sites by only 42%. This pattern was also found at the distinct species level of 100% sequence identity, where number of flagellated cercozoans and a lower diversity of cercozoan species were found in low intensively used sites (Fig. 9).

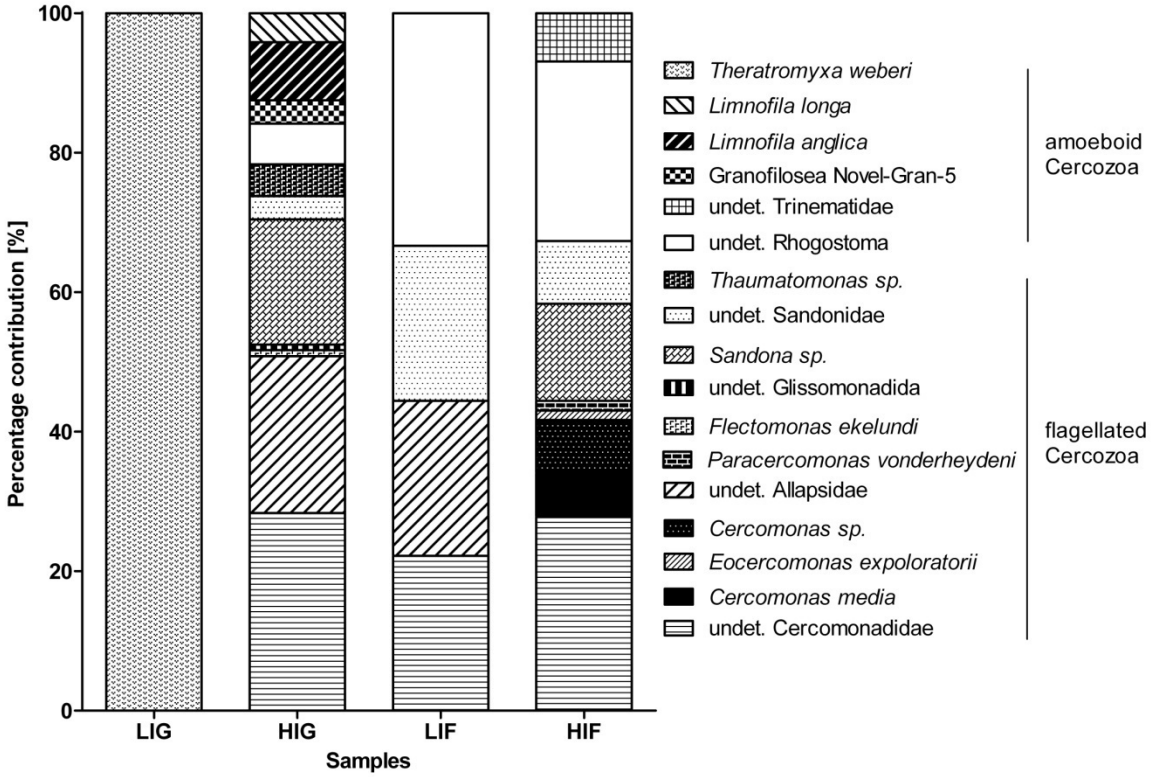


Figure 9. Relative abundance of cercozoan species grouped by 100% sequence identity.

Likewise, on the distinct level of 100% sequence identity, highest protistan species were found in HIG (20) and HIF (14) and lower numbers in the low intensively used sites LIG (four) and LIF (10). The determined protistan species were dominated by SAR (mean 97%), mainly due to high contributions of two Stramenopiles taxa, undetermined Oomyceta and Clade-C (Fig. 10). The HIG included the highest species richness and diversity (Fig. 7), resulting in similar contributions of stramenopiles, alveolates and cercozoans therein.

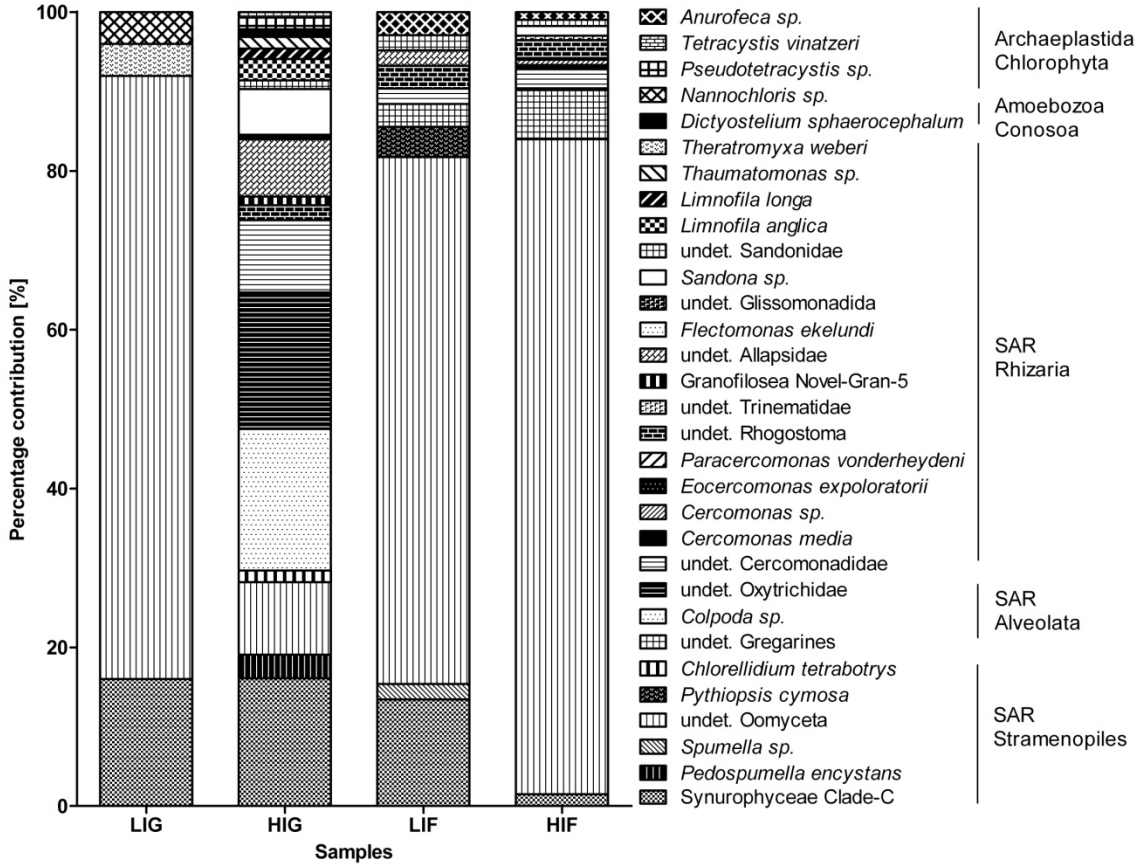


Figure 10. Relative abundance of protistan species grouped by 100% sequence identity.

The grassland and forest sites of the Hainich were compared among each other and with an arctic soil sample (referred to as G-Soil, for details see Chapter 5). Most of the determined species were rare, as they contributed for less than 10% to total community (see Tab. 2). Three protistan species were omnipresent in the Hainich samples and two of them were also indentified in the arctic soil sample. Those were at 99.7% sequence identity the stramenopiles taxa, Clade-C sp. and undetermined. Oomyceta and the cercozoan taxa, undetermined Cercomonadidae. Additionally, the omnipresent undetermined Oomyceta dominated the protistan community in LIG, LIF and HIF, but were of minor importance in HIG. Like Clade-C sp. which was one of the dominating taxa in LIG, HIG and LIF. Species which were restricted to a certain sampling site in the Hainich, contributed with less than 5% to the total protistan community. Most of the identified species in Greenland were also found in the samples of Hainich. The presence of protistan species led to similarities of the high intensively used sites, as the other samples were more different (Fig. 11). On the other hand the species abundances separated high and low intensively used sites

and the Greenland sample (Appx. Fig. 4.4). The differences between the Hainich and Greenland soil samples were mainly due different identified Chlorophyta species. Interestingly, chlorophytes were only found in grassland samples of the Hainich and in Greenland, with highest species number in LIG. The Mesomycetozoa *Anurofeca sp.* was only found in the forest samples of the Hainich.

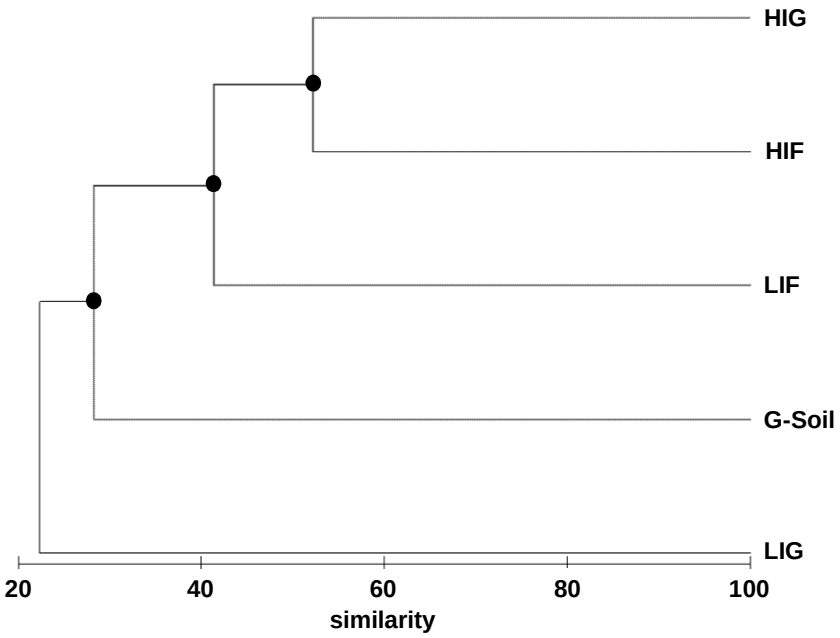


Figure 11. Similarity of protistan species soil communities in grassland and forest of Thuringia and Greenland. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weighted matrix based on species presence/absence transformed data. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.

Only five different ciliate species were identified from the different soil samples (Tab. 2). Nevertheless, ciliates contributed with up to 30% (HIG) to total protistan community. The different classes of ciliates were restricted to the different land use intensities, as colpodids and hypotrichs were only found in HIG and HIF, haptorids were only found in LIF and no ciliates were identified from LIG.

Discussion

Protistan diversity is often underestimated in morphological studies, as some organisms occur in low abundances and inactive resting stages and are not recognized based on the selectivity of determination and the enrichment procedure (Weisse 2008). The present study aims at applying molecular methods the first detailed analysis of protistan soil communities. The applied molecular study revealed a very diverse protistan community at a high resolution. Depending on the sequence identity which was proxy for the different taxonomic levels, a total of 17 protistan groups, 111 protistan genera and 40 protistan species was identified. Moreover, 31 of the species could also be identified at 100 sequence identity. The protistan community structures were thereby dominated by the SAR group (Stramenopiles, Alveolata and Rhizaria), mainly due to high stramenopile abundances. The highest species richness was found for cercozoan species, dependent on the land use intensity. High abundances and diversity of cercozoans was already indicated by the results of Chapters 1 and 2. The typical soil cercozoans are very adaptive to changing environmental conditions, due to their morphology and ability to form cysts. This was also reflected by the higher contribution and species richness of cercozoans in high intensively used sites. Heterotrophic flagellates had highest species richness, followed by autotrophic protists, amoebae, slime moulds and ciliates. Thus, the present study revealed the importance of flagellated organisms in protistan soil communities, in addition to the already known importance of ciliates and testate amoebae (Foissner 1987a, 1999b).

Influence of land use intensity on protistan soil communities

The comparison of the different sampling sites indicated highest species richness and diversity in the high intensively used grassland site. The intensity of land use had a generally significant effect on the species richness. Independent of site type, a high intensity of land use indicated higher species richness and significantly higher number of protistan species sequence reads (abundances). This assumption supported the findings of Chapter 1, where an effect of land use intensity on species abundance could be revealed for the different grassland sites. In the present study, the effect of land use was reflected by significant chances of abundance and presence of one glissomonad and two testate amoebae taxa. As in the morphological

and molecular studies land use effects were reflected by changes of amoebae, the potential of this protistan group as bioindicator is implicated. Foissner (1999b) described in his comprehensive work several advantages of protists as bioindicators and recommended the use of the highly abundant testate amoebae. He assumed that testate amoebae are easily counted and identified and that by the detailed autecological data which are available from most of the common species, testate amoebae can indicate a wide range of biotic and abiotic variables. Therefore, testate amoebae were already used as bioindicators for monitoring of moorlands (Turner and Swindles 2012), of metals such as copper (Asada and Warner 2009) and of soil age, substrate and stocking (Wanner and Dunger 2001).

Also ciliates are known as useful indicators of soil changes (Lara et al. 2007; Petz and Foissner 1989; Sauvant et al. 1999). The identified ciliate species in the present study revealed certain distribution boundaries, as colpodids were restricted to the intensively used sites and haptorids to the low intensively used sites. Lara and Acosta-Mercado (2012) identified some ciliate classes which can provide information on soil conditions. Thus, colpodids can indicate disturbed soils and haptorids can indicate stable soils. This was also supported by the study of Lara et al. (2007) where polluted sites inhabited a comparative low diverse ciliate community, dominated by colpodids, whereas haptorids were restricted to the unpolluted sites.

The results revealed an effect of land use intensity on the protistan soil community and especially on testate amoebae, ciliates and heterotrophic flagellates. Significant correlations to certain soil properties were tested, as a correlation of microbial activity to several soil properties was assumed (Adl and Gupta 2006; Weisse 2008). However, a significant correlation of certain soil properties on the total protistan community could not be revealed within this study. Even the testate amoebae and ciliates were not influenced by a certain soil property. Thus, the certain factors which caused the land use effects on the protistan soil community could not be revealed. This might be due to the analyzed mineral soil, as an independency of protistan α-diversity on environmental factors was already indicated for protistan communities in mineral soils (Bates et al. 2013) and also for bacteria and soil yeasts in various different mineral soils (Birkhofer 2012). The analysis of the litter layer in especially forest might have indicated more differences in between the different sites, as a higher microbial activity is indicated for leaf litter compared to needle litter (Priha and Smolander 1997).

Comparison of soils in the Hainich and a site in the Arctic Greenland

The comparison of the Hainich sites in Central Europe with the Arctic Greenland site indicated a far higher diversity in the Hainich soil as in Greenland soil. The comparison might indicate ubiquitous species and similar diversity patterns in soil independent on other conditions, as both sampling regions are different in habitat type (grassland/forest vs. tundra), annual mean temperature (6.5-7.5℃ vs. -0.5℃ cited in van As et al. 2012), annual mean precipitation (750-800 mm vs. 246mm cited in Mernild et al. 2010) and region (Central Europe vs. Arctic). These differences were reflected in significant differences of species abundance and richness. Thereby, the low intensively used sites were more similar to the Greenland sample than the intensively used sites. A high contribution of restricted rare taxa in the different sites was indicated. On the other hand, only few taxa dominated the communities and were present in all sites. That communities are dominated by only few taxa, but generally constituted by many restricted rare taxa was already indicated from soil community structures using molecular techniques (Bates et al. 2013; Caron and Countway 2009; Elshahed et al. 2008).

Comparison of morphological and molecular investigations

The use of NGS helped to increase the resolution of the protistan community structures and revealed further effects on the protistan communities in soil, compared to the morphological analysis (Chapter 1). By the high amount of data and the resulting analysis depth, 111 different protistan genera and 40 protistan species were identified. Thus, NGS increased the explanatory power, as only 22 protistan taxa were morphologically determined. Moreover, both methods identified taxa, which were not identified by the other study. Naked amoebae and discobian species contributed with 50%, respectively 10% to the protistan community (Chapter 1), but failed to be detected in the present study on species level. However, ciliates were only rarely recorded in the morphological investigations (Chapter 1), but could be identified by NGS. This might due to a low number of excysted organisms, as most of the active organisms are restricted to the litter layer and heavily disturbed soils (Foissner 1987b; Weisse 2008) and might be overlooked due to their low abundances in morphological studies. Nevertheless, several isolated and newly described species could be recovered by NGS in various samples, indicating that

some common protozoan species in soil might have been overlooked by morphological studies.

Nevertheless, molecular as well as morphological techniques have certain biases, leading to an under-, respectively overestimation of species numbers. The used NGS primers, like all primers, are selective to certain taxa and even the use of a general eukaryotic primer or a mix of specific primers will lead in incomplete coverage. Species which are only determined by sequences always bare the potential of false positive identification, due to sequencing errors, the proxy for taxa boundaries and the applied quality filtering (Huse et al. 2010; Kunin et al. 2010; Stoeck et al. 2010). However, the analyzed V4 region of the 18S rDNA is indicated as useful region to identify environmental diversity (Dunthorn et al. 2012). Furthermore, it provides the largest database of sequences to which the reads can be assigned to. On the other hand, morphological identifications are always limited by culture dependency and the expert knowledge. Hence, a high level of sequence identity which was proxy for genus (98%) and species (99.7%) level and a strict quality filtering was applied to minimize the overestimation of richness. Thus, a high number of sequences could also be assigned to the distinct species level of 100% sequence identity. Nevertheless, assumptions of species abundances based on sequence numbers should be made very carefully (see Chapter 3, Weber and Pawlowski 2013).

Summary

Unicellular protists are common members of the microbial soil community. Protists can be found globally and their distribution via air is mainly restricted by the given conditions of the habitat where they migrate to. Their high abundances, short generation times and dependency on given conditions, make them very suitable to indicate changes of environmental conditions. A detailed study on protistan community structures, based on next generation sequencing (NGS) of the SSU rDNA, in eight different soil sites was applied. Up to now only very few data on detailed analysis of protistan soil communities are available. Due to the establishment of the high-quality PR^2 database the present detailed analysis was facilitated, giving the first detailed analysis of different protistan soil communities. The soil sites differed in type (grassland/forest) and land use intensity (low/high), which is assumed to affect the protistan community structure. This detailed study was applied

to reveal in which way these factors change and influence the protistan community structure. Furthermore, this study continues a preliminary morphological study of the same sites. Therefore morphological and molecular data could be compared. The high amount of NGS data increased the resolution and explanatory power of protistan community structures in soil and influences on the same. A high species richness of different protistan taxa could be revealed, indicating the importance of soil as protistan habitat and reservoir for species distribution. The site type was of minor importance for the protistan community, as no significant changes were indicated between forest and grassland. On the other hand significant influences of land use intensity on protistan abundance and presence could be revealed. This effect of land use was mainly reflected by changes of testate amoebae species abundance and presence. Furthermore, the different ciliate classes were restricted to certain land use intensities. As a result, protistan communities are sensitive to changing environmental conditions, where certain taxa can especially be used to indicate those differences.

Chapter 5 – Protistan soil communities revealed by NGS: A comparison of globally distant sites and indications of distribution pathways

Introduction

A previous study of the entire protistan community in different grassland and forest sites local habitat in Central Europe, revealed an overall very diverse protistan community in soil (see Chapter 4). This local study was performed by applying next generation sequencing (NGS) leading to a very detailed analysis of species composition and by this a general increased understanding of the total protistan community structure in soil. Up to now, other studies based on NGS concentrated on certain protistan taxa or identified the protistan community only to a very rough taxonomic level (Bates et al. 2013; Kamono et al. 2013; Triado-Margarit and Casamayor 2012; Urich et al. 2011). As a high diversity of soil protists was indicated in the local study, a high diversity in other soils was assumed. The present study aims at applying molecular methods to unscramble protist diversity at globally distant sites. Therefore, soil samples were taken from globally distributed soil samples from the Arctic, Antarctic and North America (California) and compared to the samples of Central Europe of the previous study.

The understanding of protistan biodiversity and their biogeography is a current issue of ecological studies (e.g. Nolte et al. 2010; Scheckenbach et al. 2010; Stoeck et al. 2010). The paradigm of Laurens Baas Becking that ‘Everything is everywhere, but the environment selects’ (cited in de Wit and Bouvier 2006) is supported by different studies indicating a global distribution of microorganisms (Fenchel et al. 1997; Finlay 2002; Finlay and Esteban 1998; Finlay and Fenchel 1999). In contrary, other studies assume a non-global distribution and the existence of endemic flagship species among soil protists (Foissner 1999a; Smith and Wilkinson 2007). The protistan dispersal is still debated und unrevealed. Nevertheless, globally similar soil protistan communities imply unrestricted long distance dispersal. It is known that protists are able to form small cysts (Rivera et al. 1992) by which especially soil protists are distributed by air through blow up of soil particles (Altenburger et al. 2010; Wanner and Xylander 2005; Wilkinson 2001). To reveal the mechanisms of air dispersal, samples from air, ice, snow and soil were taken from a low anthropogenic influenced site in Greenland. Hereby it will be hypothesized that protistan communities are similar in the connected habitats of air, ice, snow and soil and a gradient of species richness is expected from soil > snow > air > ice due to the specific characteristics of each habitat. It is assumed that airborne protistan species have relatively low

abundances (thus rare species might be underrepresented) and species richness due to selective transport into the air by species specific differences in cyst formation and survival. By rain, respectively snow (Yamagata et al. 2009) the organisms are then transported to the ground where they accumulate in the snow layer, reflected by higher numbers of organisms per area. The organisms are then transported by the melting snow to the soil, where they are metabolically active and reproduce. Thus, the highest number of organisms (and most diverse community is assumed in soil. On the other hand snow can accumulate, forming glacial ice. Here, only very small interconnected water-filled veins (approximately 1.9µm^3, Mader et al. 2006) are present, where metabolic activity is mainly restricted to very small and psychrophilic microorganisms, such as bacteria and archaea (Price 2000). Thus, lowest numbers of inactive encysted protist species are expected in the ice samples.

The present study aims at applying molecular methods to give first insights into protist dispersal between air, ice, snow and soil using a low anthropogenic influenced model site in Greenland and to compare protist species richness of globally distant sites.

Material and Methods

Study site and sampling

Figure 1. Global distribution of the investigated sampling sites. (http://user.uni-frankfurt.de/~heuer/POKITO/images/Weltkarte.gif)

Environmental samples were collected from soil, snow, ice and air from the Arctic in Greenland and additional soil samples from sites in Iceland, USA and Antarctica (Fig. 1, for further details about the location and the sampling sites see Table 1). Concerning the different environments, sampling and DNA-extraction was performed as followed.

Kangerlussuaq, Greenland

The glacial ice (G-Ice) samples were taken by drilling 15 ice cores using an ice core drill (10cm diameter and 50cm length). The top layer of the cores was scraped by a knife that has been sterilized using a burner. The ice cores were transferred into 5l sterile infusion bags (Nündel Kunststoff-Technik GmbH, Germany) by using sterile laboratory gloves. The ice cores were melted at room temperature. The infusion bags with the melted water inside were connected to a MicroFunnel™ Filter Unit (Pall Corporation, USA) with Supor® Membrane filters (0.2 8m, white, gridded). The melted ice water was then filtered down by using an electric vacuum pump at -300 mbar. The filters for DNA analysis were stored in 15 ml tubes with ethanol absolute (molecular laboratory grade) at -20℃. The G-Ice an d G-BlIce cores were taken at the "Point 660" in the inland glacial ice of Greenland, about 40 km east of Kangerlussuaq. The black ice (G-BlIce) was an ancient compressed ice layer, which contained very low amounts of atmospheric gasses. Due to the uneven surface of the ground and the movement of the glacier, the black ice layers are folded up and break through the younger ice layers. For sampling of black ice, the upper 10cm of the ice core were removed.

Snow (G-Snow) was cut out with a sterilized saw, picked up with sterile disposable gloves and collected in sliced 3l or 5l infusion bags (Nündel Kunststoff-Technik GmbH, Germany). Samples were melted at room temperature and concurrently filtered with a vacuum pump at approximately -300 mbar by using a MicroFunnel™ Filter Unit (Pall Corporation) with Supor® Membrane (0.2µm, white, gridded). Filters were stored in ethanol absolute (molecular biology grade) at -20℃ for further DNA analysis.

The aeroplankton (G-Air) was collected with two MicroBio MB2 air samplers (Advanced Instruments, Inc., USA) at a constant flow rate of 100l/min. The air impinged onto a sterile 90mm petri dish filled with ethanol. After exposure to the air

stream for a fixed period of time, the petri dish was removed and incubated (MicroBio Manual). In total, 100m^3 air were collected per dish. The ethanol of the petri dishes was transferred into falcon tubes for transport and stored at -20℃ until DNA extraction. Samples were provided by Frank Nitsche and Hartmut Arndt.

Vatuschard, Iceland; Jubany, King George Island, Death Valley, USA and Hainich, Germany

Soil samples were collected from three different geographic regions, Arctic (G-Soil, I-Soil and I-Lava), Antatctic (J-Soil) and North America, USA (C-Desert). Soil samples were taken by removing the first 2cm of the soil face and collecting the underlying soil in 50ml Falcon tubes. Soil samples were dried and stored by room temperature until further analysis. Samples were provided by Frank Nitsche.
For further comparison of protistan diversity at globally distant sites with a site in Central Europe, two low intensively used grassland samples of the Hainich (Thuringia, Germany), LIG1 and LIG2 were chosen (see Chapter 4). LIG1 was referred to as Ger-Soil1 and LIG2 was referred to as Ger-Soil2. These sites were selected according to a similar number of sequences obtained in comparable studies. Sampling, DNA extraction, sequencing and data analysis were performed as described in Chapter 4.

DNA extraction, PCR amplification and NGS of 18S rDNA

The obtained filters from air, snow and ice were washed with ethanol absolute (molecular laboratory grade) and the solution was centrifuged at 4.000rpm for 10min. The supernatant was discharged and the pellet was resuspended and 1ml of the suspension was used for DNA extraction and handled as described for the soil samples. The DNA of 1g subsample of the dried composite soil was extracted by using the MoBio PowerSoil DNA extraction kit (MoBio Laboratories, USA). The DNA and PCR product quality was checked by agarose gel electrophoresis (2%) and all concentrations of DNA and PCR products were measured using Nanodrop 1000 (Thermo Fisher Scientific, Germany). PCR amplification and NGS preparation followed exactly the protocol described in Chapter 3, which was developed for the analysis of the highly variable V4 region of the 18SrDNA. The 710bp long fragment

can be amplified from most eukaryotic groups, enabling detailed analysis down to species level over a wide range of eukaryotic groups. PCR reactions were carried out as triplets and pooled to one sample before NGS to minimize PCR biases.

The pooled samples were used for NGS by using a 454 GS-FLX Roche sequencer system (GATC Biotech AG, Germany). All samples were sequenced from the forward primer (5' end).

Data processing

The raw data were analyzed and filtered on the basis of adapter and primer mismatch, sequence length, sequence similarity and sequence identity (for details see Chapter 4). Taxonomic identifications were assigned by using Blast (Altschul et al. 1990) against the quality filtered Protist Ribosomal Reference database PR^2 (Guillou et al. 2013), which follows the current eukaryotic classification of Adl et al. (2012). The assigned sequence reads were identified at different taxonomic levels based on sequence identity of 80%, 98% and 99.7% (see Tab. 1). The different levels for taxonomic identification were defined according to recent studies using OTU (operational taxonomic unit) clustering for NGS data analysis for protists. 80% identity was used as proxy for super-groups and groups (narrowing the 80% identity used for class identification by Massana et al. (2011) and Stoeck et al. (2010)), 98% identity was used as proxy for genera (Bachy et al. 2013) and 99.7% identity was used as proxy for species boundaries (narrowing the 99% identity used for species identification by Massana et al. (2011) and Salani et al. (2012)). The boundaries for certain taxonomic levels were narrowed, as OTU clustering thresholds are usually made very carefully to avoid overestimates of OTU numbers per taxonomic level (Huse et al. 2010; Kunin et al. 2010). Protistan super-groups were named according to Adl et al. (2012) and the following proximate taxonomic level was named as protistan group. Taxa were confirmed if at least two sequences were assigned, excluding singletons. Sequence analyses were performed by using Geneious Pro 6.1 assembler software (Biomatters Ltd.). A rough classification was performed for all eukaryotes and the subsequent detailed analysis was focused mainly on the unicellular protistan taxa (excluding fungi).

Cluster analyses of the different samples were performed by using Primer 6 software package (Clarke and Gorley 2006). Distance matrices were calculated based on the Bray-Curtis similarity using unweighted matrices based on protistan species presence/absence transformed data and weighted matrices based on protistan species log(X+1) transformed abundance. Protistan species abundances were calculated form the number of sequence reads representing each species.

T-test of presence, diversity and contribution to total community at different taxonomic levels in the Greenland ice, snow and soil samples, were performed by using R statistical software (http://www.r-project.org/).

The diversity was calculated with the Shannon-Weaver index (H) (Shannon and Weaver, 1949) using the relative abundance of sequence reads representing the taxa on a certain identity level, by:

$$H=-\sum_{i} p_i \cdot \ln p_i$$

, where p_i is the proportion of the i-th morphotype.

Additionally the evenness of the taxa was calculated by:

E= H/ln S, where S is the total number of identified taxa on a certain level.

Results

Prior to analyzing the NGS data, all obtained sequence reads were assigned to their specific origin and underwent the quality filtering. Finally, between 842 (J-Soil) and 23.329 (G-Snow) sequence reads remained of the received sequence reads (mean 31.2%) after the stringent quality filtering. The remained and analyzed sequences had a mean length of 418nt.

Table 1. Number of assigned sequences at different taxonomic levels and additional metadata information of the different sampling sites.

	G-Air	G-Ice (3 samples)	G-Snow (3 samples)	G-Soil (3 samples)	G-BlIce	I-Soil	I-Lava	C-Desert	J-Soil
N 454-reads total	69.726	66.767	26.459	14.126	4.642	3.533	4.172	12.413	2.914
N 454-reads high quality	2.263	13.269	23.329	11.041	3.505	2.749	3.233	9.946	842
Total eukayotic tags	2.259	13.268	22.698	11.004	3.508	2.742	3.232	5.965	721
Assigned tags 80% id.	1.867	11.464	18.226	9.199	3.207	2.112	2.586	30	455
Assigned tags 80% id. (protists)	751	582	6.014	2.463	2.481	1.642	1.138	30	454
Assigned tags 98% id.	1.134	7.955	4.219	4.610	2.579	1.648	927	13	4
Assigned tags 98% id. (protists)	487	503	2.229	1.068	1.963	1.252	91	13	4
Assigned tags 99.7% id.	41	693	308	296	231	289	78	0	0
Assigned tags 99.7% id. (protists)	16	74	138	98	176	198	0	0	0
Unassigned tags	396 (17.5%)	1.805 (13.6%)	5.103 (21.9%)	1.842 (16.7%)	298 (8.5%)	637 (23.2%)	647 (20.0%)	9.916 (99.7%)	387 (46.0%)
Location	67°8.3' N 50°1.4' W	67°8.3' N 50°1.4' W	67°8.3' N 50°1.4' W	67°8.1' N 50°1.5' W	67°8.1' N 50°1.4' W	65°27.85'N 18°52.1'W	65°27.85'N 18°52.1'W	36°13.2'N 116°46'W	62°14.3'S 58°39.2'W
Sampling date	02.2010	02.2010	02.2010	02.2010	02.2010	06.2008	06.2008	02.2009	12.2007
Habitat type	Air	Ice	Snow	Soil	Black ice	Soil	Lava	Desert	Soil
Country	Greenland	Greenland	Greenland	Greenland	Greenland	Iceland	Iceland	USA, California	Antarctic peninsula
Provenance	Kangerlussuaq	Kangerlussuaq	Kangerlussuaq	Kangerlussuaq	Kangerlussuaq	Vatuschard	Vatuschard	Death Valley	Jubany

The number of reads which could not be assigned to at least the level of protistan group (>80% sequence identity) ranged between 298 (G-BIIce) and 9.916 (C-Desert) sequences. For the C-Desert, this corresponded to 99.7% of all quality filtered eukaryotic sequences. For all other samples, the vast majority of sequences could be assigned at least to group level. A summary of the number of assigned sequence reads and additional metadata is shown in Table 1.

Protistan community comparisons at the level of super-groups

The overall comparison of the samples on the level of 80% sequence identity indicated a high contribution of fungi in most of the samples, see Figure 2. The Greenland samples were dominated by fungi, accounting for up to 83% (G-Ice) of an individual sample, followed by SAR group (including Stramenopiles, Alveolata and Rhizaria) with up to 77% (G-BIIce), mainly due to a high number of Cercozoa in the G-BIIce sample.

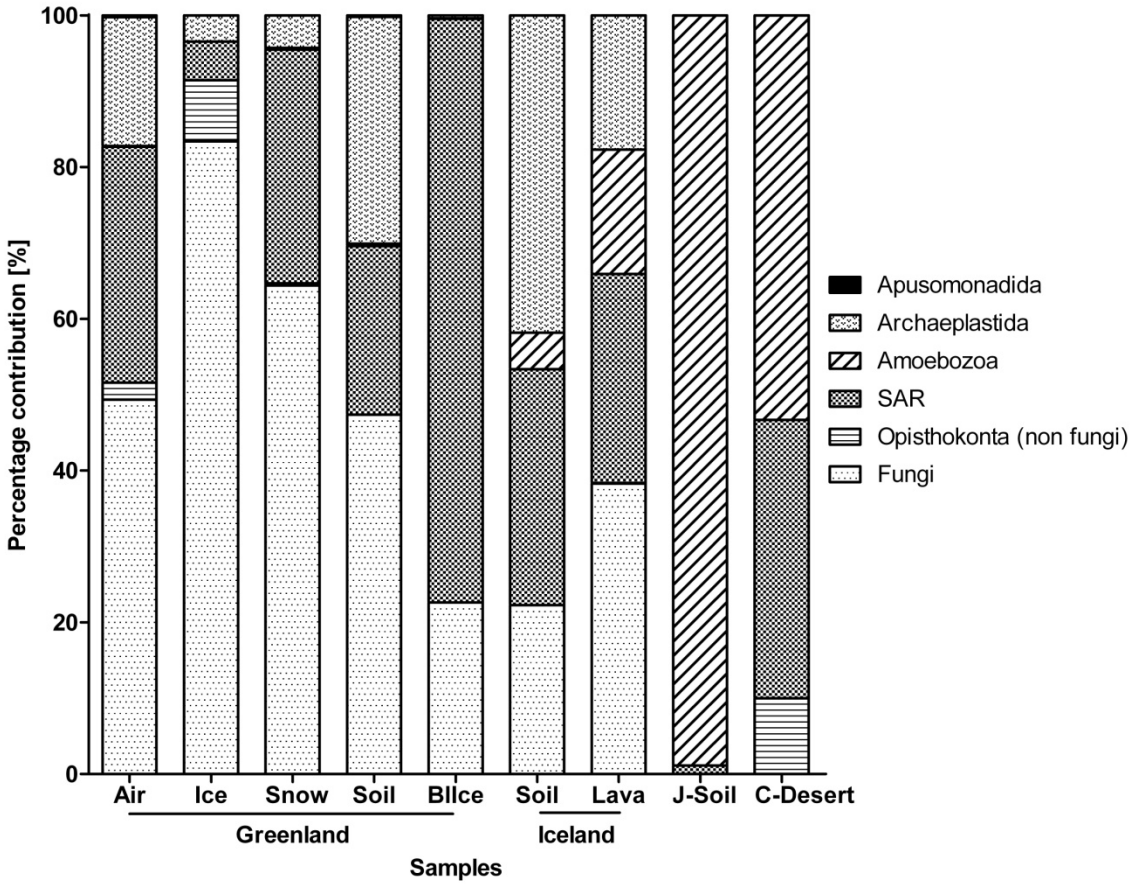


Figure 2. Relative abundances of eukaryotic super-groups grouped by 80% sequence identity.

Within the SAR group, the major groups were Rhizaria and Alveolata (Figs. 3 and 4). Generally the non fungi Opisthokonta and Apusomonadida were of minor importance. A high contribution of Amoeboza in J-Soil (99%) and C-Desert (53%) lead to a very different community structure in those samples compared to the other samples. Additionally, only two respectively three super-groups were detected in those samples, see Figure 2. The SAR group was the only super-group which was discovered in all samples. All other super-groups could be found in most of the samples, except for the eukaryote *incertae sedis* group Apusozomonadida which was only identified in G-Air and G-Soil. These samples were also those where all other five super-groups could be identified.

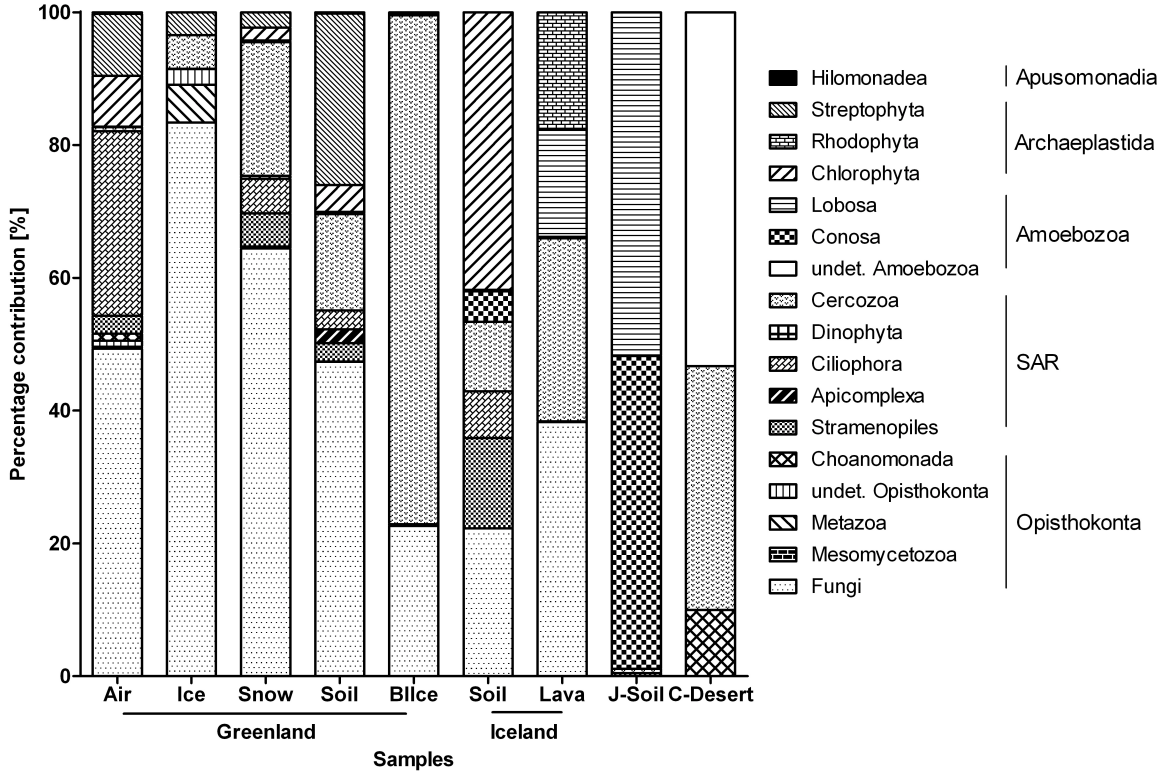


Figure 3. Relative abundances of eukaryotic groups assigned by 80% sequence identity.

The highest number of groups (taxonomic level below super-group) was identified for the Greenland air, snow and soil samples, with 12 respectively 11 groups. Therein, fungi dominated, with variable proportion of the other taxa. The other samples were generally more different to the three Greenland samples. The number of groups varied in the other samples from three (C-Desert) to seven (I-Soil).

A more detailed analysis of the protistan groups showed members of the SAR group, mainly Cercozoa and ciliated Alveolata, to be the most important groups in most of the samples (Fig. 4). Both Greenland ice samples, G-Ice and G-BIIce, were very similar and dominated by cercozoans. Matching to the super-group level, the Greenland air (G-Air) and soil (G-Soil) samples showed highest number of detected protistan groups (eight, respectively nine). Within the identified protistan groups, Choanomonada were rare and identified in four of the samples (G-Air, G-Soil, I-Lava, C-Desert). The highest proportion was found in the C-Desert with 10% of total protistan community. Choanomonads of the C-Desert and I-Lava samples were identified on genus level (98% identity) as undetermined. Stephanoecidae, which accounted for 23% (C-Desert) and 2% (I-Lava) to total protistan community. Chlorophytes were identified with high contribution in the I-Soil sample (54%), see Figure 3. The other soil samples had only low contributions of chlorophytes, with 15% in G-Soil and no contribution in J-Soil. Higher numbers of sequence reads were found in G-Air, with 19% of total protistan groups, but none in the extremes of I-Lava and C-Desert.

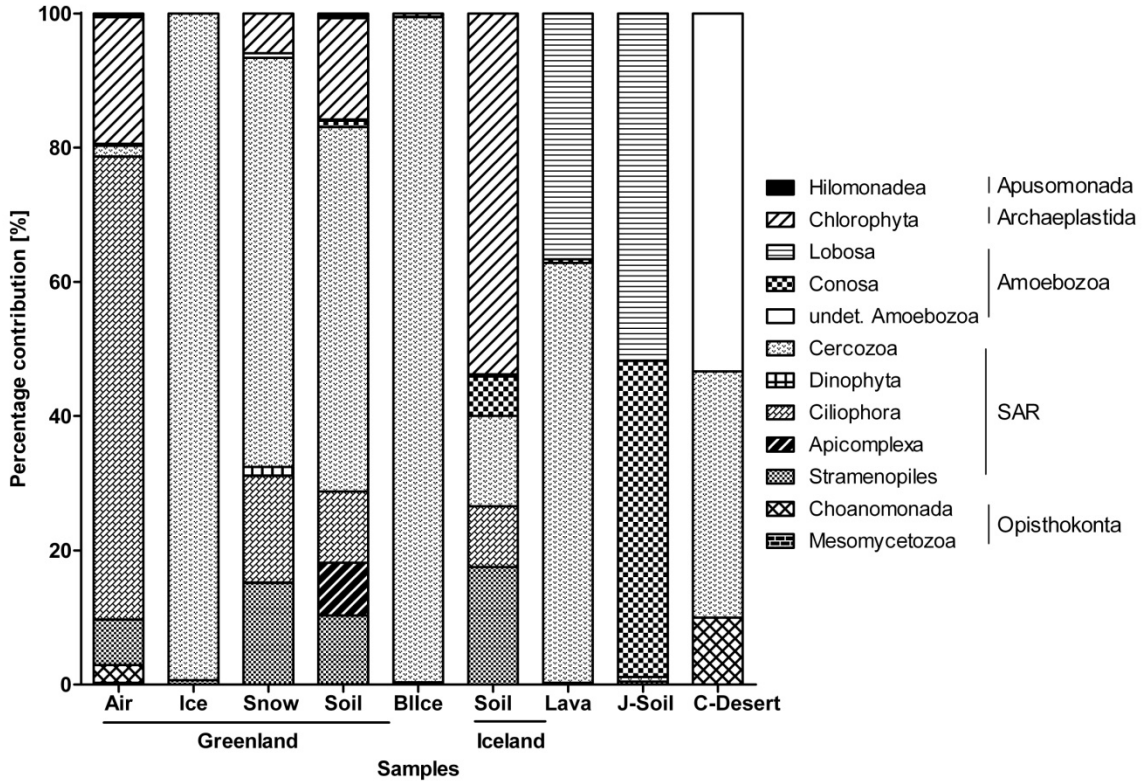


Figure 4. Relative abundances of protistan groups assigned by 80% sequence identity.

Protistan community comparisons at the level of genera

The highest number of eukaryotic genera (98% sequence identity) could be identified in G-Soil (103) and G-Snow (95), see Figure 9. The other samples showed extremely low numbers of identified genera, with mean numbers of identified genera in G-Air (43) and G-Ice (41) and very low numbers in J-Soil (2) and I-Lava (4). Nevertheless, a similar diversity was found in G-Air (2.8), G-Snow (3.2) and G-Soil (3.2). Both, G-Soil (39) and G-Snow (28) included also by far the highest number of identified protistan genera and highest diversity (Fig. 10). Thus, the number of identified protistan genera was low within the other samples ranging between 2 (G-Ice, I-Lava, J-Soil) and 9 (G-Air).

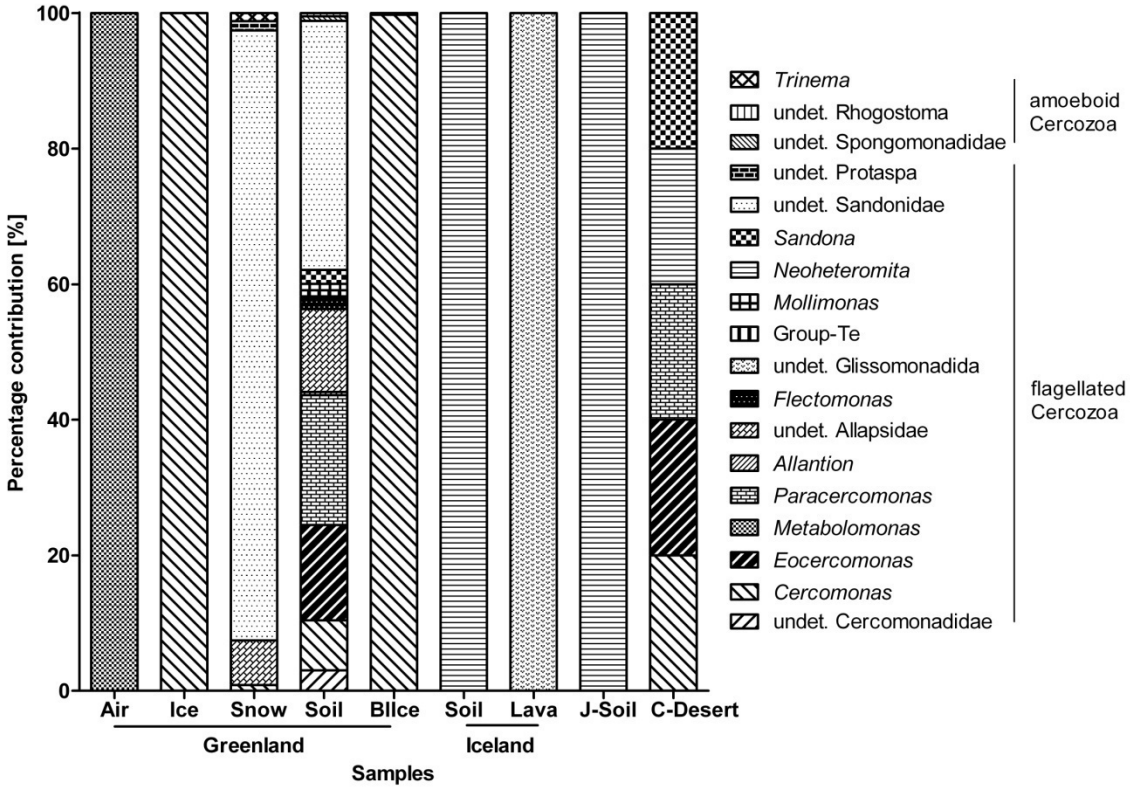


Figure 5. Relative abundance of cercozoan genera grouped by 98% sequence identity

The identified protistan genera included representatives of seven protistan groups: Choanomonada (one genus), Stramenopiles (eight genera), Apicomplexa (two genera), Ciliophora (11 genera), Dinophyta (three genera), Cercozoa (18 genera) and Chlorophyta (24 genera), see Table 2. Cercozoan genera were the most abundant and diverse heterotrophic protistan genera in the analysis, see Figures 4

and 5. The higher resolution of the different cercozoan genera indicated obvious differences between the different samples (Fig. 5). Samples were constituted by different numbers of cercozoan genera, ranging between 1 (G-Ice, I-Soil, I-Lava and J-Soil) and 13 (G-Soil), also dominated by different cercozoan genera. The cercozoan genera were dominated by flagellated cercozoans and amoeboid genera were only found in G-Snow (*Trinema*) and G-Soil (undetermined. Rhogostoma). Ciliophoran genera were only found in Greenland air, ice, snow and soil and Iceland soil (Tab. 2).

Table 2. Percentage contribution of protistan genera to the total number of identified protistan genera. Genera which were assigned only once per sampling site were excluded from the analysis. Shades of grey indicate the relative contribution of each determined genus to total protistan community.

Group	Class/order	Genus	G-Air	G-Ice	G-Snow	G-Soil	G-BIIce	I-Soil	I-Lava	C-Desert	J-Soil	Ger-Soil1	Ger-Soil2
Choanoflagellida	Acanthoecida	undet. Stephanoecidae											
Mesomycetozoa	Ichthyosphonida	*Anurofeca*											
	Rhynosporida	undet. Rhynosporidae											
Stramenopiles	Bacillariophyta	*Entomoneis*											
		Eolimna											
		Hantzschia											
		Pinnularia											
		undet. Raphid-pennate											
		Surirella											
	Synurophyceae	undet. Synurophyceae											
		Chrysophyta											
		Clade-C											
		Clade-D											
		Pedospumella											
		Spumella											
	Hyphochytrydiales	*Hyphochytrium*											
	Oomyceta	*Aphanomyces*											
		undet. Oomyceta											
	Pirsonia	*Thelebolus*											
	Xanthophyceae	*Botryochloris*											
		Chlorellidium											
		Xanthophyceae											
Apicomplexa	Colpodellidae	undet. Colpodellidae											
	Gregarines	undet. Gregarines											
		Monocystis											
Ciliophora	Colpodida	*Colpoda*											
		undet. Colpodidae											
		Hausmanniella											
		Kreyellidae											
		Pseudoplatyophrya											
	Haptoria	*Enchelyodon*											
		Enchelys											
		undet. Haptoria											
		Protospathidium											
		Spathidium											
		Trachelophyllum											

Table 2 continued.

Group	Class/order	Genus	G-Air	G-Ice	G-Snow	G-Soil	G-BIIce	I-Soil	I-Lava	C-Desert	J-Soil	Ger-Soil1	Ger-Soil2
	Hypotrichia	*Amphisiella*											
		Gonostomum											
		undet. Hypotrichia											
		Kahliella											
		Oxytricha											
		undet. Oxytrichidae											
		Stichotrichia											
		Urostyla											
	Peniculia	*Peniculia*											
	Platyophryida	*Ottowphyra*											
		Platyophrya											
Dinophyta	Dinophyceae	undet. Dinophyceae											
		Gymnodinium											
		Woloszynskia											
Cercozoa	Cercomonadida	undet. Cercomonadidae											
		Cercomonas											
		Eocercomonas											
		Metabolomonas											
		Paracercomonas											
	Cryomonadida	undet. Protaspa											
		undet. Rhogostoma											
	Euglyphida	*Euglypha*											
		Trinema											
	Granofilosea	Granofilosea Gran-5											
	Glissomonadida	*Allantion*											
		undet. Allapsidae											
		Flectomonas											
		undet. Glissomonadida											
		Group-Te											
		Mollimonas											
		Neoheteromita											
		Sandona											
		undet. Sandonidae											
	Limnofilida	*Limnofila*											
	Plasmodiophorida	*Polymyxa*											
	Spongomonadida	undet. Spongomonadidae											
		Spongomonas											
	Vampyrellida	*Leptophrys*											
		Theratromyxa											
Conosa	Variosea	AND16-lineage											
Lobosa	Centramoebida	*Acanthamoeba*											
	Echinamoebida	*Vermamoeba*											
Chlorophyta	Chlorellales	*Actinastrum*											
		Auxenochlorella											
		Chlorella											
		undet. Chlorellales											
		Muriella											
		Myrmecia											
		Nannochloris											
	Chlorophyceae	*Bracteacoccus*											
		Chlamydomonadaceae											
		Chlamydomonas											
		Chloromonas											
		Chlorosarcinopsis											
		undet. CW-Chlamydomonadales											

Table 2 continued.

Group	Class/order	Genus	G-Air	G-Ice	G-Snow	G-Soil	G-BIIce	I-Soil	I-Lava	C-Desert	J-Soil	Ger-Soil1	Ger-Soil2
Chlorophyta	Chlorophyceae	*Dictyococcus*											
		Dunaliella											
		Pseudotetracystis											
		Tetracystis											
		Vitreochlamys											
	Microthamniales	*Dictyochloropsis*											
		undet. Microthamniales											
		Myrmecia											
		Trebouxia											
	Prasiolales	*Desmococcus*											
		Diplosphaera											
		undet. Prasiolales											
		Raphidonema											
	Trebouxiophyceae	*Coccomyxa*											
		Elliptochloris											
		undet. Trebouxiophyceae											
	Trentepohliales	*Trentepohlia*											

Contribution of protistan genera

< 0.1
0.1–0.2
0.2-0.4
0.4-0.6
> 0.6

Protistan community comparisons at the level of species

On species level (99.7% sequence identity), 66 eukaryotic species were identified from 1936 sequence reads. Most of the species were protists (700 reads) and fungi (915 reads), each represented by 28 species. As no sequence reads could be assigned at the high sequence identity of 99.7% in C-Desert and J-Soil, no eukaryotic species and accordingly no protist species could be identified here. Two fungi species were detected in the I-Lava sample but no other protist species. The highest number of eukaryotic species was identified in the Greenland samples and varied from four (G-BIIce), eight (G-Air), 14 (G-Ice) and 22 (G-Soil) to 30 (G-Snow) species (Fig. 9). The number of identified species was much lower in the other samples ranging from 0 to 6 species. The same bias was found for protistan species. Here, G-Soil (13) and G-Snow (12) included most species, followed by G-Air (three) and G-Ice (two) as in the other samples only rarely or no protistan taxa were found (Fig. 10). Most of the identified protistan species were chlorophytes (14 species), followed by

cercozoans (six species) and stramenopiles (four species) (Tab. 3). The protistan community of the different Greenland samples clustered with the two samples from Central Europe (Ger-Soil1 and 2) and the I-Soil sample.

Based on the species abundances, the community of G-Air and I-Soil, G-Soil and G-Snow and G-Ice and G-BIIce were significantly similar (Fig. 6). The presence of certain protistan species indicated similarities of the G-soil and the Ger-Soil1 and of the G-Air and the Ger-Soil2 (Fig. 7). A high similarity of both ice samples, G-Ice and G-BIIce were also indicated by the presence of protistan species. All other samples were separated.

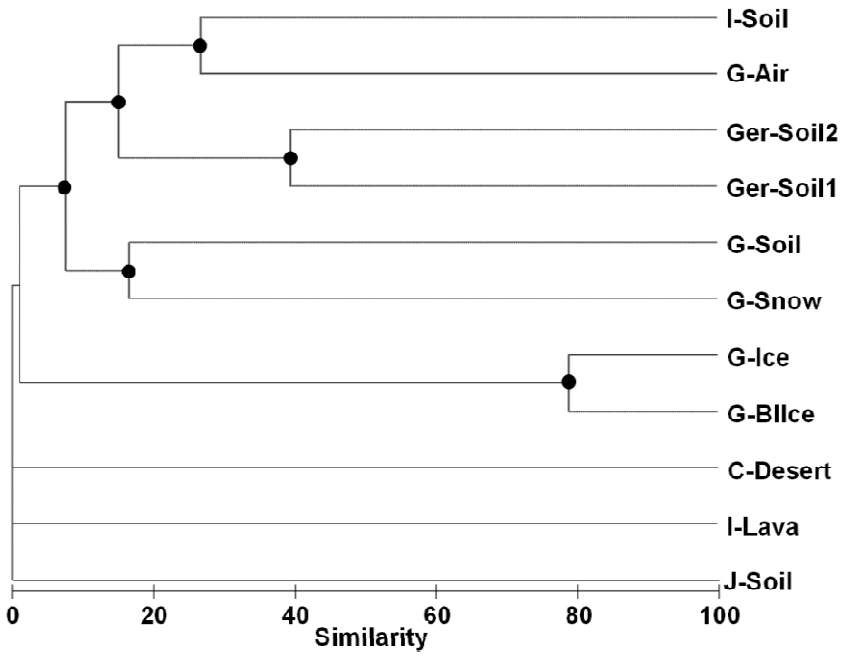


Figure 6. Similarity of protistan species communities in globally distributed samples. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weight matrix based on species log(X+1) transformed abundances. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.

The calculated diversity indicated a high diversity within the different samples among the different taxonomic levels (Figs. 9, 10). Highest protistan diversity was found in soil for species (2.23) and genera (2.95), which was close to the diversity calculated in the Ger-Soil samples with 1.97 for protistan species and 2.97 for protistan genera. Interestingly, a high diversity was also determined for samples with low richness, for example eukaryotic genera in G-Air. This was mainly due to an uneven distribution of taxa. The highest diversity was found in the different Greenland samples, G-Soil, G-Snow and G-Air.

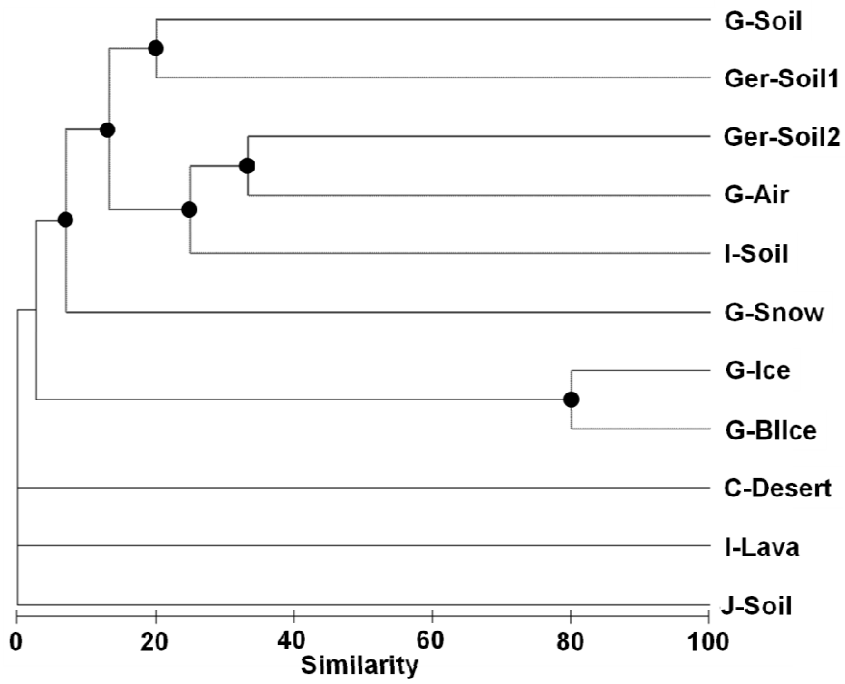


Figure 7. Similarity of protistan species communities in globally distributed samples. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using unweighted matrix based on species presence/absence transformed data. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.

Comparison of different Greenland habitats

Within the Greenland samples, ciliophoran genera were not the most abundant genera but after the cercozoans the most diverse heterotrophic group. The different ciliophoran genera were mainly restricted to a certain habitat, as colpodean genera were identified only from G-Soil, litostomatean taxa were identified from G-Soil and G-Snow, oligohymenophorean taxa were only identified from G-Snow and hypotrichian taxa were identified from G-Soil and G-Air. The most diverse group was the autotrophic protistan kingdom, Chlorophyta from which 24 genera were determined. Most of the genera were found in G-Snow (14) and G-Soil (12), with no significant difference in genus number, diversity and evenness. Though significantly differences were indicated for the ice as the diversity and evenness of protistan genera differed significantly between G-Soil and G-Ice ($F_{[2,2]}$=425.4; P=0.007 for diversity and $F_{[2,2]}$=4.9; P<0.001 for evenness) and the evenness between G-Snow and G-Ice ($F_{[2,2]}$=57.5; P=0.027 for evenness), see Appendix Table 5.1.

The cluster analysis of the Greenland samples indicated a similarity between G-Soil and G-Snow, both clustering with G-Air. Both ice samples branched on another cluster with high similarity between them (Fig. 8).

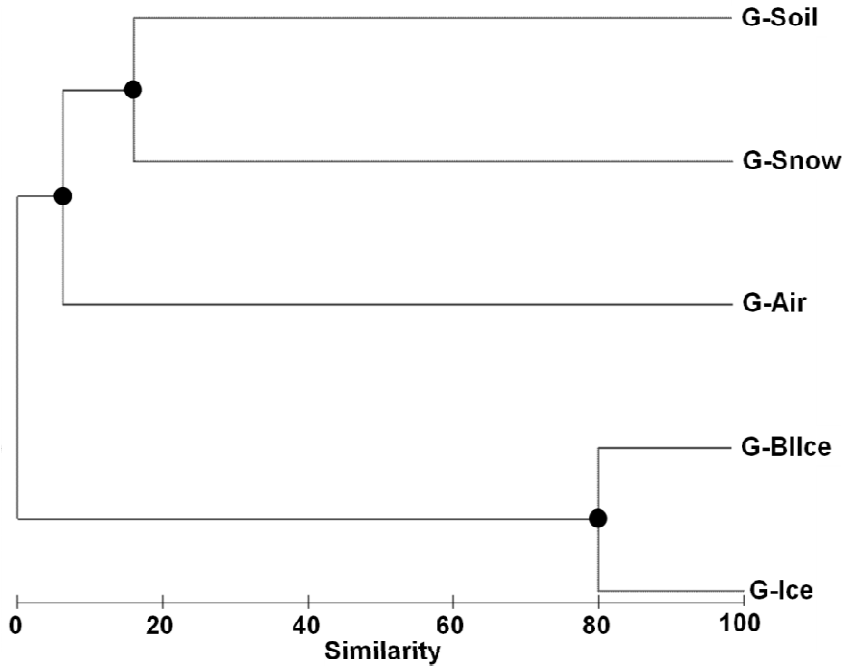


Figure 8. Similarity of protistan species communities in Greenland samples. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using unweighted matrix based on species presence/absence transformed data and weighted matrices based on species log(X+1) transformed abundances. Significant clusters were indicated by dots, significance level was set at $p<0.05$. The distance matrix was calculated by 1000 repeats. As dendrograms of unweight and weight data were identical only one cluster analysis is presented here.

Table 3. Percentage contribution of protistan species to the total number of identified protistan species. Species which were assigned only once per sampling site were excluded from the analysis. Shades of grey indicate the relative contribution of each determined species to total protistan community.

Group	Class/order	Species	G-Air	G-Ice	G-Snow	G-Soil	G-BIIce	I-Soil	I-Lava	C-Desert	J-Soil	Ger-Soil1	Ger-Soil2
Stramenopiles	Synurophyceae	Clade-C sp.	0.1–0.2									> 0.6	0.2-0.4
		Clade-D sp.			0.1–0.2								
		Pedospumella encystans											
	Pirsonia	*Thelebolus microsporus*											
Ciliophora	Haptoria	undet. Haptoria			0.4-0.6								
	Hypotrichia	*Amphisiella magnigranulosa*											
		undet. Oxytrichidae				0.1–0.2		0.1–0.2					
Dinophyta	Dinophyceae	*Gymnodinium sp.*											
Cercozoa	Cercomonadida	undet. Cercomonadidae											
		Cercomonas paraglobosa											
		Cercomonas sp.		> 0.6			> 0.6						
		Paracercomonas compacta				0.1–0.2							
	Glissomonadida	undet. Allapsidae											
		undet. Sandonidae											
Chlorophyta	Chlorophyceae	*Bracteacoccus sp.*											
		Chlamydomonadaceae *sp.*											
		Chloromonas paraserbinowii											
		Chloromonas sp.			0.4-0.6								
		undet. CW-Chlamydomonadales						> 0.6					
		Pseudotetracystis sp.	0.2-0.4										
	Microthamniales	undet. Microthamniales				0.4-0.6							
		Myrmecia astigmatica	0.4-0.6					0.1–0.2					0.2-0.4
		Myrmecia sp.											0.1–0.2
		Trebouxia jamesii											
	Prasiolales	*Diplosphaera sp.*											
		undet. Prasiolales											
	Trebouxiophyceae	*Elliptochloris subsphaerica*											0.1–0.2
		undet. Trebouxiophyceae											

Contribution of protistan species

- < 0.1
- 0.1–0.2
- 0.2-0.4
- 0.4-0.6
- > 0.6

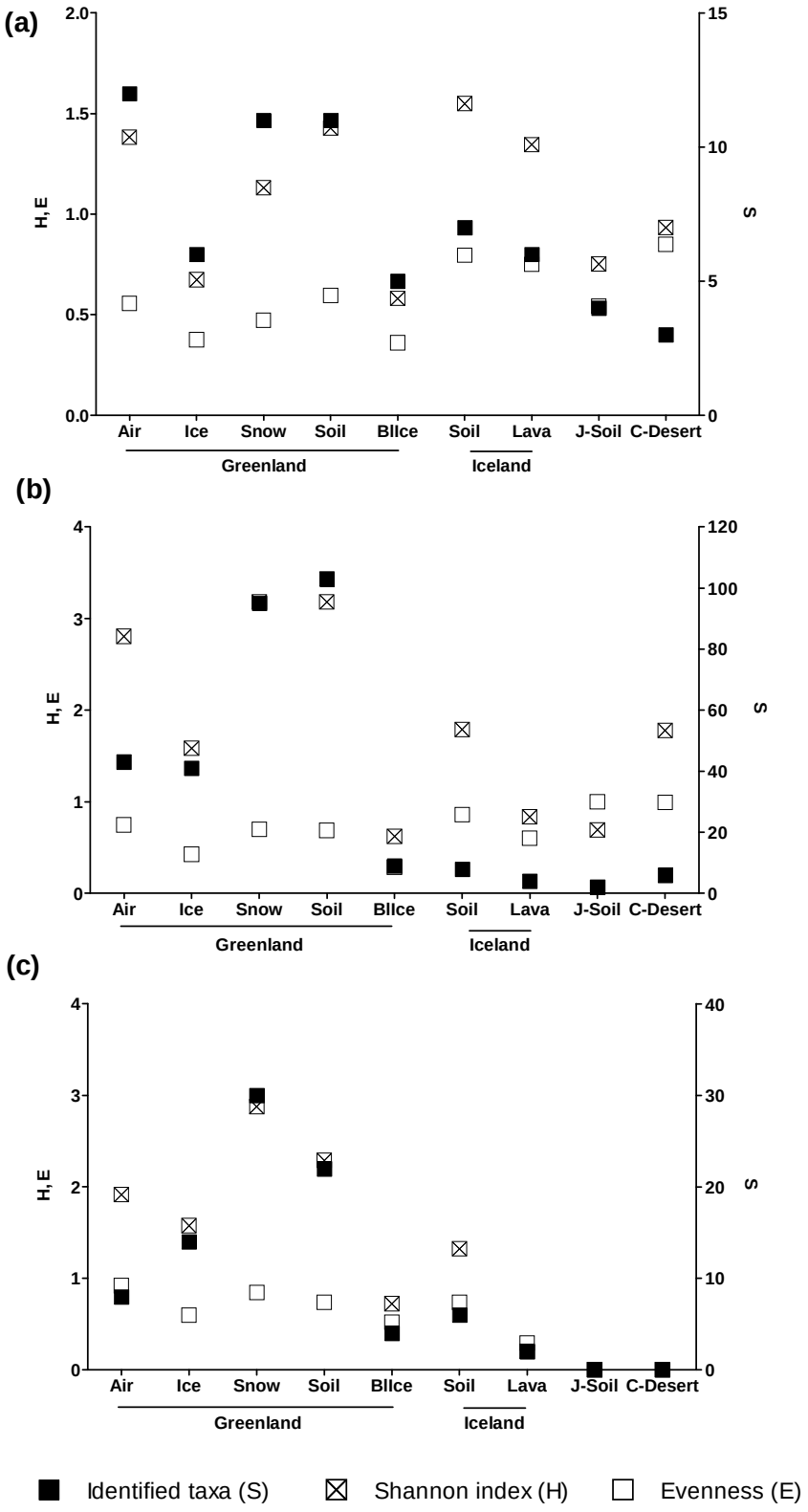


Figure 9. Calculated number of all identified eukaryotic taxa (S), Shannon diversity (H) and Evenness (E) on different levels of sequence identity, as 80% identity was proxy for (a) group, 98% identity was proxy for (b) genera and 99.7% identity was proxy for (c) species boundaries. Taxa which were assigned only once per sampling site were excluded from the analysis.

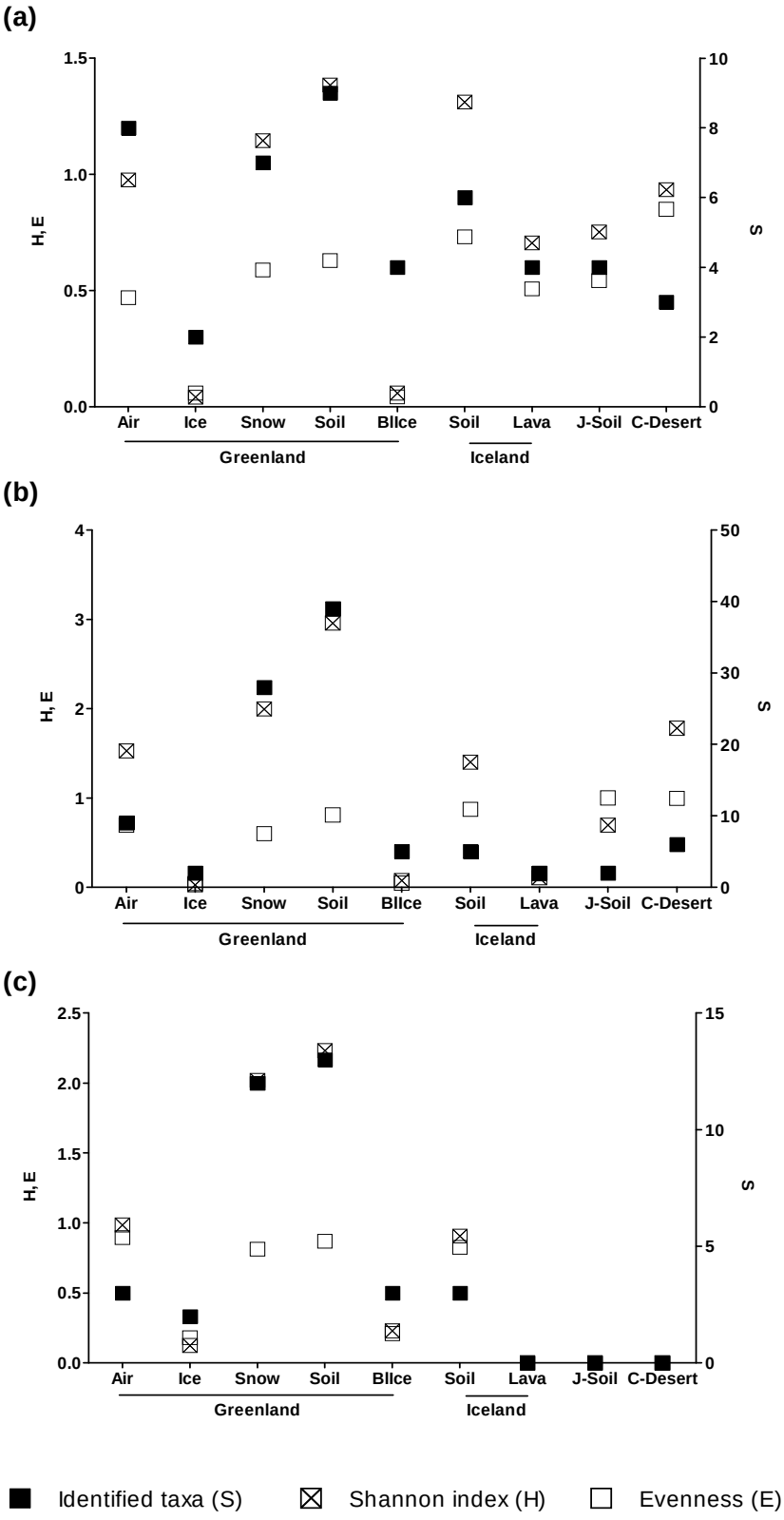


Figure 10. Calculated number of identified protistan taxa (S), Shannon diversity (H) and Evenness (E) on different levels of sequence identity, as 80% identity was proxy for (a) group, 98% identity was proxy for (b) genera and 99.7% identity was proxy for (c) species boundaries. Taxa which were assigned only once per sampling site were excluded from the analysis.

Discussion

Similarities between different protistan communities

Based on the applied molecular study, the results confirmed protistan communities dependent on the habitat, but independent on the geographic region. As the presence and abundance (read number) of protistan species was similar in soils of Germany and Greenland and both were similar to the Iceland soil. It is assumed that the distribution by air is globally unrestricted for organisms smaller than 20µm (Wilkinson et al. 2012). Consequently globally similar protist communities have to be assumed. The similar protistan communities of the Greenland, Iceland and German soils support this idea. The much lower species richness in the soils of Death Valley, the Antarctic and Iceland lava might be explained by the extreme environmental conditions.

The comparison of the different Greenland habitats revealed the assumed pathways of air dispersal. Based on the presence and abundance of protistan species, soil and snow were similar and connected to the air. Both ice samples were very similar and distant to the other Greenland habitats. The high relation of both clustered Greenland ice samples and their low species richness indicated similar selection processing on deposited protist communities.

Protistan community comparison of the different habitats of Greenland

The expected gradient of protistan species richness of soil > snow > air > ice was confirmed by the number of species and the number of genera found in each environment. No ubiquitous protistan species occurring in all four habitats could be identified. This was mainly due to low numbers of sequence reads and identified taxa in air and ice. The low numbers of species in ice and air might be due to the specific characteristics of both environments. In ice, protists withstand the conditions generally as encysted resting stages. The water-filled veins in the ice are too small for metabolically active organisms and the organisms are additionally stressed by melting and freezing in the upper ice layer during the day and night in the summer season. In air, it is assumed that airborne species are mainly originate from the soil (Rogerson and Detwiler 1999). Hence, typical soil protists such as cercozoans, amoebae and ciliates (Bates et al. 2013; Comeau et al. 2013;

Domonell et al. 2013; Esteban et al. 2006; Harding et al. 2011; Urich et al. 2008) were expected to be found. Ciliates had the highest proportion of protistan genera in the air samples dominated by hypotrichs, which are known to be the most abundant and diverse ciliates in soil (Foissner 1998). On the other hand, only one cercozoan genus and no amoebae genus or species was identified in the air sample. This was in contrary to a preliminary morphological study were four cercozoan morphotypes could be identified from the same sample (Selse 2010). However, cercozoans (Rhizaria) and ciliates (Alveolata) had high contributions in the soil and snow samples from Greenland.

Community structures of Arctic, Antarctic, USA and Central Europe

The richness and diversity of the different soil samples was relatively low especially in the extremes of Death Valley, Arctic soil and Iceland lava. The low richness and often low diversity might correspond to the notably high number of unassigned sequences in Death Valley (99.7%) and Antarctic soil (46.0%). These sequences might indicate on one hand a high hidden diversity (meaning sequences which have not been incorporated in gene banks), or on the other hand, it might reflect the specificity of the used general eukaryotic primers. However, the number of assigned sequence reads of Iceland lava was in the same range as that of other samples. Thus, the low richness in the lava sample might indicate the selective conditions in this habitat. Cercozoans were the dominating protists at genus level (80% sequence identity) in the Iceland lava and the C-Desert sample. This revealed the selectivity of these very dry and differently structured habitats, as only a low number of taxa may survive the extreme conditions. Cercozoans are globally distributed organisms (Bass and Cavalier-Smith, 2004). They are most important soil flagellates (Bass and Cavalier-Smith 2004; Brabender et al. 2012), also inhabiting other environments. Their metabolic cell shape, the possession of flagella and the ability to form cysts within a short period of time, make them very suitable to survive extreme conditions. Cercozoans were the only protistan phylum, which was found in all samples reflecting their generality and unrestricted occurrence. Bamforth et al. (2005) found flagellates and amoebae as the most abundant taxa in an Antarctic dry valley. Most of the flagellates he identified were cercozoans and kinetoplastids, amoebae were

dominated by *Acanthamoeba.* In the present study, the used primers were not suitable for amplification of kinetoplastids. Hence kinetoplastids were not detected. Nevertheless, cercozoans and lobose amoebae (and undetermined Amoebozoa) dominated the community in the dry samples of this study.

While most of the protistan species contributed only little to the total protistan community, two species, *Cercomonas sp.* and undetermined CW-Chlamydomonadales, were extremely abundant and contributed more than 60% to the total protistan community. Thereby *Cercomonas sp.* was restricted to both ice types. The most common genus was the stramenopile Clade-C (relative to *Spumella*), as it was found on a wide range of samples. The less diverse samples were dominated by certain genera which were restricted mainly to that specific site. As the air was dominated by *Stichotrichia* (205 sequence reads), the ice (501 sequence reads) and black ice (1942 sequence reads) by *Cercomonas*, the snow by undetermined Haptoria (601 sequence reads) and undetermined Sandonidae (837 sequences), no dominant genus was identified for the most rich and diverse Greenland soil sample. The samples with the highest species richness were characterized by numerous rare genera, while samples with lower richness contained only a few but one dominant genus. Most autotrophic species were found in the snow sample. A high contribution of autotrophic organisms in the snow was also indicated by other studies (Felip et al. 1995; Harding et al. 2011).
This applied molecular study revealed a highly diverse protistan community, especially in the Greenland samples. The high throughput method enabled a highly resolved view into the community structure. As the protist diversity is often underestimated, based on undersampling, low abundances, resting stages and methodological biases based on selective determination and enrichment (Weisse 2008), culture independent molecular techniques will increase the understanding of protistan diversity. The present highly resolved analysis of the protistan communities indicated on the one hand that transport via air should be selective and on the other hand that Greenland and German soil protist communities are rather similar. The latter supports the idea that most protist species can be found everywhere depending on the environment, which is in agreement with several other studies on ciliates (Fenchel et al. 1997; Finlay 2002; Finlay and Clarke 1999; Finlay and Esteban 1998). Different aspects support microbial global and long

distance dispersal, i. the ability to form cysts, allowing outlive of various harsh conditions (Rivera et al. 1992), ii. the low weight and small body/cyst size, facilitating distribution via wind or attached to dust particles and other organisms (Rivera et al. 1992; Rogerson and Detwiler 1999; Wilkinson 2001) and iii. the large population sizes increasing the chance for individuals dispersing (Wilkinson et al. 2012).

Summary

Protists are globally distributed organisms which can be found in all different kinds of habitats, where they contribute significantly to nutrient cycling. The mechanisms of protistan dispersal are poorly understood. Their small body size, high abundances and their ability to form cysts lead to their potential of unrestricted distribution. Once they are airborne, they could disperse, migrate and colonize new habitats.

Based on the unrestricted dispersal, similar protistan communities are expected in similar habitats. To test this hypothesis, soil samples of the Arctic, Antarctic and USA, as well as soil samples from Central Europe (Chapter 4) were analyzed by applying next generation sequencing of the highly variable V4 region of the SSU rDNA. Similarities between the Central European soil samples and Arctic soil samples support the idea of a global distribution of protists. The protistan communities of these soil samples were dominated by cercozoan, stramenopile and chlorophyte genera. Samples from extremely dry habitats were dominated by cercozoans.

Based on the hypothesized ways of dispersal a gradient of richness was assumed for soil > snow > air > ice. This could be confirmed by the studies carried out at low anthropogenic influenced site in Greenland, on the richness and presence of protistan species and genera indicating a selective transport via air. A strong correlation between soil and snow and between air and soil was revealed.

Concluding Remarks and Perspective

Despite the ecological significance of protists in soil environments, little is known about their taxonomic composition and the factors influencing abundance and community structure. This thesis could reveal the first detailed analysis of protistan community structures in soils down to the species level using modern sequencing techniques.

Protistan community structures in local grassland and forest sites

The effects of site type and land use intensity on the protistan soil community structure were investigated by applying morphological and molecular techniques. Hence, local grassland and forest sites with different land use of a region in Thuringia were chosen for the comparisons. The morphological investigations revealed an influence of land use intensity on the abundance of protozoans in grassland soils, mainly due to differences in amoebaean abundances. Flagellates (dominated by cercomonads and glissomonads) and naked amoebae (dominated by fan-shaped amoebae) shared a similar part of the morphological determined communities. As cercomonads, amongst glissomonads, dominated the flagellated protozoan community and had highest diversity, this important and morphological variable soil flagellates were analyzed in more detail. This was carried out, by a combination of morphological investigations, including detailed ultrastructural studies and phylogenetic analyses in collaboration with other scientists. The detailed analysis of 23 isolated monoclonal cultures indicated a high hidden diversity of these important soil flagellates, as ten new species, including two new genera could be described, where I contributed to the analysis with the description of two new species. These morphological analyses indicated the general importance of further research on protists in soil and additionally the requirement of very detailed studies to increase the taxonomic resolution.

Therefore next generation sequencing on the same sampling sites was applied to analyze the total protistan community. The very detailed analysis indicated in addition to flagellates and naked amoebae, ciliates, testate amoebae, unicellular algae and slime moulds to contribute evidently to the protistan soil community. The dominant groups were representatives of the SAR group (Stramenopiles, Alveolata and Rhizaria), where stramenopiles had highest abundances (number of

sequences) and cercozoans had highest species richness in the soil communities. The site type was of minor importance for the protistan community, as no significant changes were indicated between forest and grassland. On the other hand a significant influence of land use intensity on protistan presence and abundance (number of sequences) could be revealed. This land use effect was indicated by changes of testate amoebae species abundance and presence and restricted appearance of certain ciliate classes. The communities were analyzed on different sequence identities as proxy for different taxonomic levels. A sequence identity of at least 98% is required for significant conclusions on the protistan communities. As a result, protistan communities could be shown to be sensitive to changing environmental conditions, where certain taxa can especially be used to indicate those differences. The applied molecular approach could refine and enlarge the assumptions made by morphological studies.

Protistan community structures in globally distant soil sites and indications of protistan dispersal

Based on the applied molecular study, the results confirmed protistan community structures dependent on the habitat, but independent on the geographic region. As the presence and abundance (number of sequences) of protistan species was similar in soils of Germany and Greenland and both were similar to the Iceland soil, the idea of unrestricted global distribution by air is supported for small organisms. The much lower species richness in the soils of Death Valley, the Antarctic and Iceland lava might be explained by the extreme environmental conditions.

The comparison of the different Greenland habitats revealed the assumed pathways of air dispersal. Based on the presence and abundance (number of sequences) of protistan species and genera, samples from soil and snow were similar and related to air samples. Young and ancient ice were very similar and different to the other Greenland habitats. The high similarity of both Greenland ice samples and their low species richness indicated similar selection processing on deposited protist communities.

Application of next generation sequencing for protistan community identification

The evaluation of next generation sequencing (NGS) to identify in-vitro simulated flagellate community structures indicated a strong bias of PCR amplification in large protistan communities and revealed the pitfalls by using culture independent molecular techniques, emphasizing a strict quality filtering during the data processing and analysis. Thus, NGS can lead to assumptions regarding species richness, but assumptions of species abundances should be made carefully. The quality of the assumptions made by NGS studies depends on the database where the sequence reads are assigned to. Since the development of the high-quality database PR^2 a highly resolved analyses of species richness are possible. This database needs continuous update to include as much species as possible to the analysis.

Conclusion

This thesis shows for the first time detailed analyses of protistan soil communities down to species level. Based on preliminary morphological studies a high species richness and diversity was estimated and could be revealed. Soil communities are generally dominated by members of the SAR group, where environmental changes are indicated mainly by changes of testate amoebae and ciliates presence. A strong influence of land use on protistan species was revealed. Similar soil communities in distant globally distributed soil sites as well as analyses of air, snow and ice samples indicated a global distribution of many soil protists. The applied molecular technique, showed its potential for detailed analysis of soil protistan community structure, but the methodological studies also indicated that assumptions regarding abundances should be made with caution.

The pipeline developed in this thesis for protistan community analyses by NGS can be adapted for a wide range of further detailed analysis of protistan communities. A higher amount of samples e.g. on a certain site might reveal indications of small scale patchiness changes in protistan community composition. In addition, further sampling of local samples will increase the understanding of land use and site type effects and further sampling of globally distributed samples might increase the understanding of protistan global distribution.

List of References

Adl, M.S., Gupta, V., 2006. Protists in soil ecology and forest nutrient cycling. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 36, 1805-1817.

Adl, S.M., Coleman, D.C., 2005. Dynamics of soil Protozoa using a direct count method. Biology and Fertility of Soils 42, 168-171.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukes, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., le Gall, L., Lynn, D.H., McManus, H., Mitchell, E.A.D., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C.L., Smirnov, A., Spiegel, F.W., 2012. The revised classification of eukaryotes. Journal of Eukaryotic Microbiology 59, 429-493.

Altenburger, A., Ekelund, F., Jacobsen, C.S., 2010. Protozoa and their bacterial prey colonize sterile soil fast. Soil Biology and Biochemistry 42, 1636-1639.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local aligment search tool. Journal of Molecular Biology 215, 403-410.

Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. Microbiological Reviews 59, 143-169.

Arndt, H., Dietrich, D., Auer, B., Cleven, E.-J., Gräfenhan, T., Weitere, M., Mylnikov, A.P., 2000. Functional diversity of heterotrophic flagellates in aquatic ecosystems, in: Leadbeater, B.s.C., Green, J.C. (Eds.), The Flagellates. Taylor & Francis, London. 240-268.

Asada, T., Warner, B.G., 2009. Plants and testate amoebae as environmental indicators in cupriferous peatlands, New Brunswick, Canada. Ecological Indicators 9, 129-137.

Auer, B., Arndt, H., 2001. Taxonomic composition and biomass of heterotrophic flagellates in relation to lake trophy and season. Freshwater Biology 46, 1–14.

Babaei, Z., Oormazdi, H., Rezaie, S., Rezaeian, M., Razmjou, E., 2011. Giardia intestinalis: DNA extraction approaches to improve PCR results. Experimental Parasitology 128, 159-162.

Bachy, C., Dolan, J.R., Lopez-Garcia, 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 Journal 7, 244-255.

Bamforth, S.S., Wall, D.H., Virginia, R.A., 2005. Distribution and diversity of soil protozoa in the mcmurdo dry valleys of antarctica. Polar Biology 28, 756-762.

Bass, D., Cavalier-Smith, T., 2004. Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). International Journal of Systematic and Evolutionary Microbiology 54, 2393-2404.

Bass, D., Howe, A.T., Mylnikov, A.P., Vickerman, K., Chao, E.E., Edwards Smallbone, J., Snell, J., Cabral Jr, C., Cavalier-Smith, T., 2009. Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas, Eocercomonas, Paracercomonas*, and *Cavernomonas* gen. nov. Protist 160, 483-521.

Bass, D., Richards, T.A., Matthai, L., Marsh, V., Cavalier-Smith, T., 2007. DNA evidence for global dispersal and probable endemicity of Protozoa. BMC Evolutionary Biology 7.

Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R., Fierer, N., 2013. Global biogeography of highly diverse protistan communities in soil. ISME Journal 7, 652-659.

Bernard, C., Simpson, A.G.B., Patterson, D.J., 2000. Some free-living flagellates (Protista) from anoxic habitats. Ophelia 52, 113–142.

Birkhofer, K.I.S., Fabian Alt, Nadine Herold, Bernhard Klarner, Mark Maraun, Sven Marhan, Yvonne Oelmann, Tesfaye Wubet, Andrey Yurkov, Dominik Begerow, Doreen Berner, François Buscot, Rolf Daniel, Tim Diekötter, Roswitha B. Ehnes, Georgia Erdmann, Christiane Fischer, Bärbel Foesel, Janine Groh, Jessica Gutknecht, Ellen Kandeler, Christa Lang, Gertrud Lohaus, Annabel Meyer, Heiko Nacke, Astrid Näther, Jörg Overmann, Andrea Polle, Melanie M. Pollierer, Stefan Scheu, Michael Schloter, Ernst-Detlef Schulze, Waltraud Schulze, Jan Weinert, Wolfgang W. Weisser, Volkmar Wolters, Marion Schrumpf, 2012. General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. PLoS One 7.

Boenigk, J., Arndt, H., 2002. Bacterivory by heterotrophic flagellates: Community structure and feeding strategies. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 81, 465-480.

Bonkowski, M., 2004. Protozoa and plant growth: The microbial loop in soil revisited. New Phytologist 162, 617-631.

Bonkowski, M., Brandt, F., 2002. Do soil Protozoa enhance plant growth by hormonal effects? Soil Biology and Biochemistry 34, 1709-1715.

Bonkowski, M., Cheng, W.X., Griffiths, B.S., Alphei, G., Scheu, S., 2000. Microbial-faunal interactions in the rhizosphere and effects on plant growth. European Journal of Soil Biology 36, 135-147.

Bonkowski, M., Clarholm, M., 2012. Stimulation of plant growth through interactions of Bacteria and Protozoa: Testing the auxiliary microbial loop hypothesis. Acta Protozoologica 51, 237-247.

Brabender, M., Kiss, Á.K., Domonell, A., Nitsche, F., Arndt, H., 2012. Phylogenetic and morphological diversity of novel soil cercomonad species with a description of two new genera (*Nucleocercomonas* and *Metabolomonas*). Protist 163, 495-528.

Buee, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S., Martin, F., 2009. 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytologist 184, 449-456.

Butler, H., Rogerson, A., 1995. Temporal and spatial abundance of naked amebas (Gymnamoebae) in marine benthic sediments of the Clyde-Sea area, Scotland. Journal of Eukaryotic Microbiology 42, 724-730.

Caron, D., Davis, P., Sieburth, J., 1989. Factors responsible for the differences in cultural estimates and direct microscopical counts of populations of bacterivorous nanoflagellates. Microbial Ecology 18, 89-104.

Caron, D.A., Countway, P.D., 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquatic Microbial Ecology 57, 227-238.

Cavalier-Smith, T., Chao, E.E.Y., 2003. Phylogeny and classification of phylum Cercozoa (Protozoa). Protist 154, 341-358.

Cavalier-Smith, T., Karpov, S.A., 2012. *Paracercomonas* kinetid ultrastructure, origins of the body plan of Cercomonadida, and cytoskeleton evolution in Cercozoa. Protist 63, 47-75.

Cavalier-Smith, T., Oates, B., 2012. Ultrastructure of *Allapsa vibrans* and the body plan of Glissomonadida (Cercozoa). Protist 163, 165-187.

Clarholm, M., 1985. Interactions of Bacteria, Protozoa and plants leading to mineralization of soil nitrogen. Soil Biology and Biochemistry 17, 181-187.

Clarholm, M., 1989. Effects of plant-bacterial-amoebal interactions on plant uptake of nitrogen under field conditions. Biology and Fertility of Soils 8, 373-378.

Clarholm, M., Popovic, B., Rosswall, T., Soderstrom, B., Sohlenius, B., Staaf, H., Wiren, A., 1981. Biological aspects of nitrogen mineralization in humus from a pine forest podsol incubated under different moisture and temperature conditions. Oikos 37, 137-145.

Clarke, K.R., Gorley, R.N., 2006. Program Primer, Version 6, Primer -E, Plymouth, UK.

Comeau, A.M., Philippe, B., Thaler, M., Gosselin, M., Poulin, M., Lovejoy, C., 2013. Protists in Arctic drift and land-fast sea ice. Journal of Phycology 49, 229-240.

Darbyshire, J.F., Wheatley, R.E., Greaves, M.P., Inkson, R.H.E., 1974. Rapid micromethod for estimating bacterial and protozoan populations in soil. Revue d'ecologie et de Biologie du sol 11, 465-475.

de Wit, R., Bouvier, T., 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environmental Microbiology 8, 755-758.

Domonell, A., Brabender, M., Nitsche, F., Bonkowski, M., Arndt, H., 2013. Community structure of cultivable protists in different grassland and forest soils of Thuringia. Pedobiologia 56, 1-7.

Dujardin, F., 1841. Histoire Naturelle des Zoophytes Infusoires. Roret, Paris.

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. The Journal of Eukaryotic Microbiology 59, 185-187.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., Holder, M., Taylor, G.T., Suarez, P., Varela, R., Epstein, S., 2011. Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. ISME Journal 5, 1344-1356.

Ekelund, F., Daugbjerg, N., Fredslund, L., 2004. Phylogeny of *Heteromita, Cercomonas* and *Thaumatomonas* based on SSU rDNA sequences, including the description of *Neocercomonas jutlandica* sp. nov., gen. nov. European Journal of Protistology 40, 119-135.

Ekelund, F., Patterson, D.J., 1997. Some heterotrophic flagellates from a cultivated garden soil in Australia. Archiv für Protistenkunde 148, 461-478.

Ekelund, F., Ronn, R., 1994. Notes on Protozoa in agricultural soil with emphasis on heterotrophic flagellates and naked amebas and their ecology. FEMS Microbiology Reviews 15, 321-353.

Ekelund, F., Rønn, R., Griffiths, B.S., 2001. Quantitative estimation of flagellate community structure and diversity in soil samples. Protist 152, 301-314.

Elshahed, M.S., Youssef, N.H., Spain, A.M., Sheik, C., Najar, F.Z., Sukharnikov, L.O., Roe, B.A., Davis, J.P., Schloss, P.D., Bailey, V.L., Krumholz, L.R., 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. Applied and Environmental Microbiology 74, 5422-5428.

Esteban, G.F., Clarke, K.J., Olmo, J.L., Finlay, B.J., 2006. Soil protozoa - An intensive study of population dynamics and community structure in an upland grassland. Applied Soil Ecology 33, 137-151.

Felip, M., Sattler, B., Psenner, R., Catalan, J., 1995. Highly-active microbial communities in the ice and snow cover of high-mountain lakes. Applied and Environmental Microbiology 61, 2394-2401.

Fenchel, T., 1974. Intrinsic rate of natural increase: Relationship with body size. Oecologia 14, 317-326.

Fenchel, T., Esteban, G.F., Bland, J.F., 1997. Local versus global diversity of microorganisms: Cryptic diversity of ciliated Protozoa. Oikos 80, 220-225.

Finlay, B.J., 2002. Global dispersal of free-living microbial eukaryote species. Science 296, 1061-1063.

Finlay, B.J., Black, H.I.J., Brown, S., Clarke, K.J., Esteban, G.F., Hindle, R.M., Olmo, J.L., Rollett, A., Vickerman, K., 2000. Estimating the growth potential of the soil protozoan community. Protist 151, 69-80.

Finlay, B.J., Clarke, K.J., 1999. Ubiquitous dispersal of microbial species. Nature 400, 828-828.

Finlay, B.J., Esteban, G.F., 1998. Freshwater Protozoa: Biodiversity and ecological function. Biodiversity and Conservation 7, 1163-1186.

Finlay, B.J., Fenchel, T., 1999. Divergent perspectives on protist species richness. Protist 150, 229-233.

Finlay, B.J., Fenchel, T., 2001. Protozoan community structure in a fractal soil environment. Protist 152, 203-218.

Finlay, B.J., Fenchel, T., 2004. Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 155, 237-244.

Fischer, M., Kalko, E.K.V., Linsenmair, K.E., Pfeiffer, S., Prati, D., Schulze, E.-D., *et al.*, 2010. Exploratories for large-scale and long-term functional biodiversity research, in: Müller, F., Baessler, C., Schubert, H., Klotz, S. (Eds.), Long-term ecological research - between theory and application. Sringer, Heidelberg.

Foissner, W., 1987a. The micro-edaphon in ecofarmed and conventionally farmed dryland cornfields near Vienna (Austria). Biology and Fertility of Soils 3, 45-49.

Foissner, W., 1987b. Soil Protozoa: Fundamental problems, ecological significance, adaptations in ciliates and testaceans, bioindicators, and guide to the literature. Progress in Protistology 2, 69-212.

Foissner, W., 1991. Diversity and ecology of soil flagellates, in: Patterson, D.J., Larsen, J. (Eds.), Biology of free-living heterotrophic flagellates. Oxford University Press, Oxford, pp. 93-112.

Foissner, W., 1992. Comparative studies on the soil life in ecofarmed and conventionally farmed fields and grasslands of Austria. Agriculture, Ecosystems & Environment 40, 207-218.

Foissner, W., 1998. An updated compilation of world soil ciliates (Protozoa, Ciliophora), with ecological notes, new records, and descriptions of new species. European Journal of Protistology 34, 195-235.

Foissner, W., 1999a. Protist diversity: Estimates of the near-imponderable. Protist 150, 363-368.

Foissner, W., 1999b. Soil Protozoa as bioindicators: Pros and cons, methods, diversity, representative examples. Agriculture, Ecosystems and Environment 74, 95-112.

Foissner, W., Buchgraber, K., Berger, H., 1990. Bodenfauna, Vegetation und Ertrag bei ökologisch und konventionell bewirtschaftetem Grünland: Eine Feldstudie mit randomisierten Blöcken. Mitteilungen der österreichischen bodenkundlichen Gesellschaft 41, 5-33.

Foissner, W., Strüder-Kypke, M., van der Staay, G.W.M., Moon-van der Staay, S.-Y., Hackstein, J.H.P., 2003. Endemic ciliates (Protozoa, Ciliophora) from tank bromeliads (Bromeliaceae): A combined morphological, molecular, and ecological study. European Journal of Protistology 39, 365-372.

Glücksman, E., Bell, T., Griffiths, R.I., Bass, D., 2010. Closely related protist strains have different grazing impacts on natural bacterial communities. Environmental Microbiology 12, 3105-3113.

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.

Gottel, N.R., Castro, H.F., Kerley, M., Yang, Z.M., Pelletier, D.A., Podar, M., Karpinets, T., Uberbacher, E., Tuskan, G.A., Vilgalys, R., Doktycz, M.J., Schadt, C.W., 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Applied and Environmental Microbiology 77, 5934-5944.

Griffiths, B.S., Ritz, K., Wheatley, R., Kuan, H.L., Boag, B., Christensen, S., Ekelund, F., Sørensen, S.J., Muller, S., Bloem, J., 2001. An examination of the biodiversity-ecosystem function relationship in arable soil microbial communities. Soil Biology and Biochemistry 33, 1713-1722.

Guillard, R.R., Lorenzen, C.J., 1972. Yellow-green algae with chlorophyllide c. Journal of Phycology 8, 10-14.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W., Lara, E., Le Bescot, N., Logares, R., Mahe, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R.,

2013. The protist ribosomal reference database (PR^2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41, D597-D604.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52, 696–704.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

Hänel, K., 1979. Systematik und Ökologie der farblosen Flagellaten des Abwassers. Archiv für Protistenkunde 121, 73–137.

Harding, T., Jungblut, A.D., Lovejoy, C., Vincent, W.F., 2011. Microbes in high arctic snow and implications for the cold biosphere. Applied and Environmental Microbiology 77, 3234-3243.

Hausmann, K., Hülsmann, N., Radek, R., 2003. Protistology, third ed. Science Publishers, Stuttgart, Germany.

Herdler, S., Kreuzer, K., Scheu, S., Bonkowskia, M., 2008. Interactions between arbuscular mycorrhizal fungi (*Glomus intraradices*, Glomeromycota) and amoebae (*Acanthamoeba castellanii*, Protozoa) in the rhizosphere of rice (*Oryza sativa*). Soil Biology and Biochemistry 40, 660-668.

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

Howe, A.T., Bass, D., Vickerman, K., Chao, E.E., Cavalier-Smith, T., 2009. Phylogeny, taxonomy, and astounding genetic diversity of Glissomonadida ord. nov., the dominant gliding zooflagellates in soil (Protozoa: Cercozoa). Protist 160, 159-189.

Huse, S.M., Welch, D.M., Morrison, H.G., Sogin, M.L., 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology 12, 1889-1898.

Jones, D.L., Nguyen, C., Finlay, R.D., 2009. Carbon flow in the rhizosphere: Carbon trading at the soil–root interface. Plant and Soil 321, 5-33.

Jumpponen, A., Jones, K.L., 2009. Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. New Phytologist 184, 438-448.

Kamono, A., Meyer, M., Cavalier-Smith, T., Fukui, M., Fiore-Donno, A.M., 2013. Exploring slime mould diversity in high-altitude forests and grasslands by environmental RNA analysis. FEMS Microbiology Ecology 84, 98-109.

Karpov, S.A., Bass, D., Mylnikov, A.P., Cavalier-Smith, T., 2006. Molecular phylogeny of Cercomonadidae and kinetid patterns of *Cercomonas* and *Eocercomonas* gen. nov. (Cercomonadida, Cercozoa). Protist 157, 125-158.

Kiss, Á.K., Ács, É., Kiss, K.T., Török, J.K., 2009. Structure and seasonal dynamics of the protozoan community (heterotrophic flagellates, ciliates, amoeboid Protozoa) in the plankton of a large river (River Danube, Hungary). European Journal of Protistology 45, 121–138.

Koller, R., Rodriguez, A., Robin, C., Scheu, S., Bonkowski, M., 2013. Protozoa enhance foraging efficiency of arbuscular mycorrhizal fungi for mineral nitrogen from organic matter in soil to the benefit of host plants. New Phytologist 199, 203-211.

Kreuzer, K., Bonkowski, M., Langel, R., Scheu, S., 2004. Decomposer animals (Lumbricidae, Collembola) and organic matter distribution affect the performance of *Lolium perenne* (Poaceae) and *Trifolium repens* (Fabaceae). Soil Biology and Biochemistry 36, 2005-2011.

Kunin, V., Engelbrektson, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology 12, 118-123.

Lara, E., Acosta-Mercado, D., 2012. A molecular perspective on ciliates as soil bioindicators. European Journal of Soil Biology 49, 107-111.

Lara, E., Berney, C., Ekelund, F., Harms, H., Chatzinotas, A., 2007. Molecular comparison of cultivable Protozoa from a pristine and a polycyclic aromatic hydrocarbon polluted site. Soil Biology and Biochemistry 39, 139-148.

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., Osteras, M., Farinelli, L., Pawlowski, J., 2011. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proceedings of the National Academy of Sciences of the United States of America 108, 13177-13182.

Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R., Cary, S.C., 2012. Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS One 7, e44224.

Lee, J., Leedale, G., Bradbury, P., 2000. An illustrated guide to the Protozoa: Organisms traditionally referred to as Protozoa, or newly discovered groups, second ed. Society of Protozoologists, Lawrence, Kansas.

Leuko, S., Goh, F., Ibanezz-Peral, R., Burns, B.P., Walter, M.R., Neilan, B.A., 2008. Lysis efficiency of standard DNA extraction methods for *Halococcus spp.* In an organic rich environment. Extremophiles 12, 301-308.

Mader, H.M., Pettitt, M.E., Wadham, J.L., Wolff, E.W., Parkes, R.J., 2006. Subsurface ice as a microbial habitat. Geology 34, 169-172.

Massana, R., Pernice, M., Bunge, J.A., del Campo, J., 2011. Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. ISME Journal 5, 184-195.

Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwalder, B., Schlotterer, C., Boenigk, J., 2010. Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Molecular Ecology 19, 32-40.

Mernild, S.H., Liston, G.E., Steffen, K., van den Broeke, M., Hasholt, B., 2010. Runoff and mass-balance simulations from the Greenland ice sheet at Kangerlussuaq (Søndre Strømfjord) in a 30-year perspective, 1979-2008. Cryosphere 4, 231-242.

Mokhele, B., Zhan, X., Yang, G., Zhang, X., 2012. Review: Nitrogen assimilation in crop plants and its affecting factors. Canadian Journal of Plant Science 92, 399-405.

Mylnikov, A.P., Karpov, S.A., 2004. Review of the diversity and taxonomy of cercomonads. Protistology 3, 201–217.

Mylnikov, A.P., 1985. A new species *Cercobodo amoebinus* (class Zoomastigophorea Calkins). Biologia Vnutrenih vod Informationi Biuleten bul 65, 22–25 (in Russian).

Nacke, H., Thurmer, A., Wollherr, A., Will, C., Hodac, L., Herold, N., Schoning, I., Schrumpf, M., Daniel, R., 2011. Pyrosequencing-based assessment of bacterial community structure along different management types in german forest and grassland soils. PLoS One 6, e17000.

Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwalder, B., Boenigk, J., Schlotterer, C., 2010. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. Molecular Ecology 19, 2908-2915.

Nylander, J.A.A., 2003. MrAIC.pl. Program distributed by the author.

Paterson, E., 2003. Importance of rhizodeposition in the coupling of plant and microbial productivity. European Journal of Soil Science 54, 741-750.

Pawlowski, J., Christen, R., Lecroq, B., Bachar, D., Shahbazkia, H.R., Amaral-Zettler, L., Guillou, L., 2011. Eukaryotic richness in the Abyss: Insights from pyrotag sequencing. PLoS One 6, e18169.

Petz, W., Foissner, W., 1989. Spatial seperation of terrestrial ciliates and testaceans (Protozoa): A contribution to soil ciliatostasis. Acta Protozoologica 27, 249-258.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics B 19, 1572–1574.

Price, P.B., 2000. A habitat for psychrophiles in deep Antarctic ice. Proceedings of the National Academy of Sciences of the United States of America 97, 1247-1251.

Priha, O., Smolander, A., 1997. Microbial biomass and activity in soil and litter under *Pinus sylvestris, Picea abies* and *Betula pendula* at originally similar field afforestation sites. Biology and Fertility of Soils 24, 45-51.

Rivera, F., Lugo, A., Ramirez, E., Bonilla, P., Calderon, A., Rodriguez, S., Ortiz, R., Gallegos, E., Labastida, A., Chavez, M.P., 1992. Seasonal distribution of air-borne Protozoa in Mexico-City and its suburbs. Water Air and Soil Pollution 61, 17-36.

Rogerson, A., Detwiler, A., 1999. Abundance of airborne heterotrophic protists in ground level air of South Dakota. Atmospheric Research 51, 35-44.

Ronn, R., Vestergard, M., Ekelund, F., 2012. Interactions between Bacteria, Protozoa and Nematodes in soil. Acta Protozoologica 51, 223-235.

Rosenberg, K., Bertaux, J., Krome, K., Hartmann, A., Scheu, S., Bonkowski, M., 2009. Soil amoebae rapidly change bacterial community composition in the rhizosphere of *Arabidopsis thaliana*. ISME Journal 3, 675-684.

Salani, F.S., Arndt, H., Hausmann, K., Nitsche, F., Scheckenbach, F., 2012. Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. ISME Journal 6, 713-723.

Sauvant, N.P., Pepin, D., Piccinni, E., 1999. *Tetrahymena pyriformis*: A tool for toxicological studies. A review. Chemosphere 38, 1631-1669.

Scheckenbach, F., Hausmann, K., Wylezich, C., Weitere, M., Arndt, H., 2010. Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. Proceedings of the National Academy of Sciences of the United States of America 107, 115-120.

Scheckenbach, F., Wylezich, C., Mylnikov, A.P., Weitere, M., Arndt, H., 2006. Molecular comparisons of freshwater and marine isolates of the same morphospecies of heterotrophic flagellates. Applied and Environmental Microbiology 72, 6638-6643.

Selse, S., 2010. Investigation of arctic aeroplankton regarding the global dispersal of protists via the air. Bachelor thesis at the University of Cologne, Zoological Institute.

Shannon, C.E., Weaver, W., 1949. The mathematical theory of communication, University of Illinois Press, Urbana, IL.

Sherr, E.B., Sherr, B.F., 2002. Significance of predation by protists in aquatic microbial food webs. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 81, 293-308.

Simpson, E.H., 1949. Measurement of diversity. Nature 163, 688.

Skuja, H., 1939. Beitrag zur Algenflora Lettlands II. Acta Horti Botanici Universitatis Latviensis 11/12: 41–169.

Skuja, H., 1948. Taxonomie des Phytoplanktons einiger Seen in Uppland, Schweden. Symbolae Botanicae Upsalienses 9, 1–399.

Skuja, H., 1956. Taxonomische und biologische Studien über das Phytoplankton Schwedischer Binnengewässer. Nova acta Reggiae Societatis societarum upsalinensis 16, 1–404.

Smirnov, A., Goodkov, A., 1999. An illustrated list of the basic morphotypes of Gymnamoebia (Rhizopoda, Lobosea). Protistology 1, 20-29.

Smit, E., Leeflang, P., Gommans, S., van den Broek, J., van Mil, S., Wernars, K., 2001. Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. Applied and Environmental Microbiology 67, 2284-2291.

Smith, H.G., Wilkinson, D.M., 2007. Not all free-living microorganisms have cosmopolitan distributions - the case of *Nebela* (*Apodera*) *vas* Certes (Protozoa: Amoebozoa: Arcellinida). Journal of Biogeography 34, 1822-1831.

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 underexplores "rare biosphere". Proceedings of the National Academy Sciences USA 103, 12115-12120.

Staley, J.T., Konopka, A., 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats Annual Review of Microbiology 39, 321-346.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., 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.

Strien, J., Sanft, J., Mall, G., 2013. Enhancement of PCR amplification of moderate GC-containing and highly GC-rich DNA sequences. Molecular Biotechnology 54, 1048-1054.

Tikhonenkov, D.V., 2007. New species of amoeboid flagellates of the genus *Cercomonas* (Cercomonadida, Cercozoa) from freshwater bodies at the Brazilian coast. Zoologicheskii Zhurnal, Moscow 86, 136–146 (in Russian with English summary).

Thompson, J.D., Gibson, T.J., Plewniak, F., Higgins, D.G., Jeanmougin, F. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 24, 4876-4882.

Torsvik, V., Sorheim, R., Goksoyr, J., 1996. Total bacterial diversity in soil and sediment communities - a review. Journal of Industrial Microbiology 17, 170-178.

Treonis, A.M., Austin, E.E., Buyer, J.S., Maul, J.E., Spicer, L., Zasada, I.A., 2010. Effects of organic amendment and tillage on soil microorganisms and microfauna. Applied Soil Ecology 46, 103-110.

Triado-Margarit, X., Casamayor, E.O., 2012. Genetic diversity of planktonic eukaryotes in high mountain lakes (Central Pyrenees, Spain). Environmental Microbiology 14, 2445-2456.

Triadó-Margarit, X., Casamayor, E.O., 2013. High genetic diversity and novelty in planktonic protists inhabiting inland and coastal high salinity water bodies. FEMS Microbiology Ecology 85, 27-35.

Turner, T.E., Swindles, G.T., 2012. Ecology of testate amoebae in moorland with a complex fire history: Implications for ecosystem monitoring and sustainable land management. Protist 163, 844-855.

Urich, T., Lanzen, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C., 2011. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS One 3, e2527.

van As, D., Hubbard, A.L., Hasholt, B., Mikkelsen, A.B., van den Broeke, M.R., Fausto, R.S., 2012. Large surface meltwater discharge from the Kangerlussuaq sector of the Greenland ice sheet during the record-warm year 2010 explained by detailed energy balance observations. Cryosphere 6, 199-209.

Wamberg, C., Christensen, S., Jakobsen, I., Müller, A.K., Sørensen, S.J., 2003. The mycorrhizal fungus (*Glomus intraradices*) affects microbial activity in the rhizosphere of pea plants (*Pisum sativum*). Soil Biology and Biochemistry 35, 1349-1357.

Wanner, M., Dunger, W., 2001. Biological activity of soils from reclaimed open-cast coal mining areas in upper Lusatia using testate amoebae (protists) as indicators. Ecological Engineering 17, 323-330.

Wanner, M., Xylander, W.R., 2005. Biodiversity development of terrestrial testate amoebae: Is there any succession at all? Biology and Fertility of Soils 41, 428-438.

Weber, A.A.T., Pawlowski, J., 2013. Can abundance of protists be inferred from sequence data: A case study of Foraminifera. PLoS One 8, e56739.

Wegener Parfrey, L., Grant, J., Katz, L.A., 2012. Ribosomal DNA is differentially amplified across life-cycle stages in the forminifer *Allogromia laticollaris* strain CSH. Journal of Foraminiferal Research 42, 151-155.

Weisse, T., 2008. Distribution and diversity of aquatic protists: An evolutionary and ecological perspective. Biodiversity and Conservation 17, 243-259.

Wilkinson, D.M., 2001. What is the upper size limit for cosmopolitan distribution in free-living microorganisms? Journal of Biogeography 28, 285-291.

Wilkinson, D.M., Koumoutsaris, S., Mitchell, E.A.D., Bey, I., 2012. Modelling the effect of size on the aerial dispersal of microorganisms. Journal of Biogeography 39, 89-97.

Yamagata, S., Kobayashi, D., Ohta, S., Murao, N., Shiobara, M., Wada, M., Yabuki, M., Konishi, H., Yamanouchi, T., 2009. Properties of aerosols and their wet deposition in the Arctic spring during ASTAR2004 at Ny-Ålesund, Svalbard. Atmospheric Chemistry and Physics 9, 261-270.

Zhu, F., Massana, R., Not, F., Marie, D., Vaulot, D., 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiology Ecology 52, 79-92.

Zhukov, B.F., 1993. Atlas of freshwater heterotrophic flagellates. Russian Academic of Science, Institute of Biology. Inland Waters, Rybinsk pp. 158 (in Russian).

Appendices

Appendix Chapter 1

Appendix Table 1.1. Site type, intensity of land use and site type in dependence of intensity of land use compared to the percentage contribution of protozoan groups and community assemblage. Values of percentage contribution were arcsin-transformed before being analyzed. F- and P-values of an ANOVA on the effect of parameters on different protozoa were presented here.

Protozoan group	Site type [a]	Intensity of land use [b]	Site type[a] x intensity of land use[b]
Amoebae	$F_{[3,7]} = 0.08$; P = 0.787	$F_{[3,7]} = 2.43$; P = 0.194	$F_{[3,7]} = 0.41$; P = 0.559
Cercozoans	$F_{[3,7]} = 0.14$; P = 0.730	$F_{[3,7]} = 17.03$; P = 0.015	$F_{[3,7]} = 1.18$; P = 0.338
Euglenozoans	$F_{[3,7]} = 0.28$; P = 0.626	$F_{[3,7]} = 0.54$; P = 0.502	$F_{[3,7]} = 0.00$; P = 0.997
Straminopiles	$F_{[3,7]} = 1.25$; P = 0.326	$F_{[3,7]} = 0.08$; P = 0.795	$F_{[3,7]} = 0.01$; P = 0.940
Apusomonads	$F_{[3,7]} = 2.16$; P = 0.215	$F_{[3,7]} = 0.00$; P = 0.990	$F_{[3,7]} = 1.02$; P = 0.369
Other flagellates	$F_{[3,7]} = 0.03$; P = 0.879	$F_{[3,7]} = 0.27$; P = 0.630	$F_{[3,7]} = 0.95$; P = 0.385
Ciliates	$F_{[3,7]} = 0.22$; P = 0.667	$F_{[3,7]} = 0.30$; P = 0.616	$F_{[3,7]} = 0.51$; P = 0.513

[a] Grassland or forest site
[b] Extensive or intensive land use

Appendix Table 1.2. Site type, intensity of land use, site type in dependence of intensity of land use and treatments compared to the protozoan abundance and community assemblage. F-values of an ANOVA on the effect of parameters on different protozoa were presented here (*P<0.05, (*) P<0.1).

Morphotype	Site type [a]	Intensity of land use [b]	Site type[a] x intensity of land use[b]	Treatment[c]
Flagellates	$F_{[3,7]} = 1.05$	$F_{[3,7]} = 1.72$	$F_{[3,7]} = 3.21$	$F_{[3,7]} = 1.99$
Bicosoecids	$F_{[3,7]} = 0.97$	$F_{[3,7]} = 0.94$	$F_{[3,7]} = 0.47$	-
Chrysomonads	$F_{[3,7]} = 1.17$	$F_{[3,7]} = 1.43$	$F_{[3,7]} = 1.48$	-
Rynchomonas nasuta -group	$F_{[3,7]} = 1.53$	$F_{[3,7]} = 0.19$	$F_{[3,7]} = 0.17$	-
Bodo saltans –group	$F_{[3,7]} = 0.31$	$F_{[3,7]} = 1.01$	$F_{[3,7]} = 0.26$	-
Neobodo designis –group	$F_{[3,7]} = 0.14$	$F_{[3,7]} = 1.02$	$F_{[3,7]} = 0.63$	-
Apusomonads	$F_{[3,7]} = 0.00$	$F_{[3,7]} = 0.31$	$F_{[3,7]} = 1.62$	-
Glissomonads	$F_{[3,7]} = 0.04$	$F_{[3,7]} = 0.66$	$F_{[3,7]} = 3.24$	-
Other cercomonads	$F_{[3,7]} = 1.64$	$F_{[3,7]} = 0.57$	$F_{[3,7]} = 0.50$	-
Cercomonad morphotype C	$F_{[3,7]} = 0.75$	$F_{[3,7]} = 0.01$	$F_{[3,7]} = 0.09$	-
Cercomonad morphotype B	$F_{[3,7]} = 2.02$	$F_{[3,7]} = 1.68$	$F_{[3,7]} = 2.49$	-
Cercomonad morphotype A	$F_{[3,7]} = 0.00$	$F_{[3,7]} = 0.14$	$F_{[3,7]} = 0.00$	-
Other flagellates	$F_{[3,7]} = 0.07$	$F_{[3,7]} = 1.34$	$F_{[3,7]} = 0.35$	-
Euglenozoans	$F_{[3,7]} = 0.01$	$F_{[3,7]} = 1.40$	$F_{[3,7]} = 0.12$	-
Stramenopiles	$F_{[3,7]} = 1.11$	$F_{[3,7]} = 1.27$	$F_{[3,7]} = 1.09$	-
Kinetoplastids	$F_{[3,7]} = 0.00$	$F_{[3,7]} = 1.24$	$F_{[3,7]} = 0.09$	-
Cercozoans	$F_{[3,7]} = 0.42$	$F_{[3,7]} = 0.00$	$F_{[3,7]} = 2.09$	-
Amoebae	$F_{[3,7]} = 2.30$	$F_{[3,7]} = 12.99$*	$F_{[3,7]} = 12.12$*	$F_{[3,7]} = 9.13$*
Acanthopodian amoebae	$F_{[3,7]} = 1.15$	$F_{[3,7]} = 0.78$	$F_{[3,7]} = 0.15$	-
Monotactic amoebae	$F_{[3,7]} = 2.60$	$F_{[3,7]} = 4.63$(*)	$F_{[3,7]} = 4.68$(*)	-
Testate amoebae	$F_{[3,7]} = 0.43$	$F_{[3,7]} = 0.05$	$F_{[3,7]} = 2.36$	-
Hartmannelid-like amoebae	$F_{[3,7]} = 0.25$	$F_{[3,7]} = 2.36$	$F_{[3,7]} = 1.62$	-
Fan-shaped amoebae	$F_{[3,7]} = 0.65$	$F_{[3,7]} = 8.90$*	$F_{[3,7]} = 10.52$*	-
Other amoebae	$F_{[3,7]} = 0.50$	$F_{[3,7]} = 0.11$	$F_{[3,7]} = 0.00$	-
Ciliates	$F_{[3,7]} = 0.21$	$F_{[3,7]} = 1.46$	$F_{[3,7]} = 0.02$	$F_{[3,7]} = 0.56$
Protozoa	$F_{[3,7]} = 2.26$	$F_{[3,7]} = 8.15$*	$F_{[3,7]} = 9.39$*	$F_{[3,7]} = 6.60$*

[a] Grassland or forest site
[b] Extensive or intensive land use
[c] Intensive/extensive grassland or intensive/extensive forest site

Appendix Table 1.3. Site type, intensity of land use, site type in dependence of intensity of land use and moisture compared to the percentage contribution of flagellate groups to total flagellate numbers and percentage contribution of amoebae groups to total amoebae. Values of percentage contribution were arcsin-transformed before being analyzed. F-values of an ANOVA on the effect of parameters on different protozoans were presented here (*P<0.05, (*) P<0.1).

Morphotype	Site type [a]	Intensity of land use [b]	Site type[a] x intensity of land use[b]	Moisture	Bacteria
Bicosoecids	$F_{[5,7]} = 0.81$	$F_{[5,7]} = 0.51$	$F_{[5,7]} = 2.54$	$F_{[5,7]} = 2.98$	$F_{[5,7]} = 0.52$
Chrysomonads	$F_{[5,7]} = 2.81$	$F_{[5,7]} = 1.67$	$F_{[5,7]} = 0.48$	$F_{[5,7]} = 1.36$	$F_{[5,7]} = 2.10$
Rynchomonas nasuta -group	$F_{[5,7]} = 0.15$	$F_{[5,7]} = 1.68$	$F_{[5,7]} = 0.24$	$F_{[5,7]} = 1.10$	$F_{[5,7]} = 0.18$
Bodo saltans –group	$F_{[5,7]} = 0.28$	$F_{[5,7]} = 0.04$	$F_{[5,7]} = 0.02$	$F_{[5,7]} = 0.10$	$F_{[5,7]} = 0.14$
Neobodo designis –group	$F_{[5,7]} = 8.95$(*)	$F_{[5,7]} = 0.03$	$F_{[5,7]} = 6.62$	$F_{[5,7]} = 7.64$	$F_{[5,7]} = 10.12$(*)
Apusomonads	$F_{[5,7]} = 0.03$	$F_{[5,7]} = 0.54$	$F_{[5,7]} = 2.64$	$F_{[5,7]} = 11.64$(*)	$F_{[5,7]} = 5.93$
Glissomonads	$F_{[5,7]} = 0.42$	$F_{[5,7]} = 8.06$	$F_{[5,7]} = 0.03$	$F_{[5,7]} = 0.12$	$F_{[5,7]} = 0.11$
Other cercomonads	$F_{[5,7]} = 14.41$(*)	$F_{[5,7]} = 0.02$	$F_{[5,7]} = 0.52$	$F_{[5,7]} = 0.63$	$F_{[5,7]} = 7.23$
Cercomonad morphotype C	$F_{[5,7]} = 1.43$	$F_{[5,7]} = 0.26$	$F_{[5,7]} = 0.11$	$F_{[5,7]} = 0.46$	$F_{[5,7]} = 0.64$
Cercomonad morphotype B	$F_{[5,7]} = 0.03$	$F_{[5,7]} = 0.01$	$F_{[5,7]} = 0.09$	$F_{[5,7]} = 0.54$	$F_{[5,7]} = 0.23$
Cercomonad morphotype A	$F_{[5,7]} = 23.63$*	$F_{[5,7]} = 19.13$*	$F_{[5,7]} = 5.03$	$F_{[5,7]} = 0.28$	$F_{[5,7]} = 12.97$(*)
Other flagellates	$F_{[5,7]} = 3.94$	$F_{[5,7]} = 4.24$	$F_{[5,7]} = 1.56$	$F_{[5,7]} = 0.65$	$F_{[5,7]} = 5.28$
Euglenozoans	$F_{[5,7]} = 1.21$	$F_{[5,7]} = 0.17$	$F_{[5,7]} = 0.74$	$F_{[5,7]} = 1.04$	$F_{[5,7]} = 1.31$
Stramenopiles	$F_{[5,7]} = 2.29$	$F_{[5,7]} = 1.40$	$F_{[5,7]} = 0.94$	$F_{[5,7]} = 1.92$	$F_{[5,7]} = 1.69$
Kinetoplastids	$F_{[5,7]} = 1.39$	$F_{[5,7]} = 0.11$	$F_{[5,7]} = 0.79$	$F_{[5,7]} = 1.21$	$F_{[5,7]} = 1.47$
Cercozoans	$F_{[5,7]} = 0.00$	$F_{[5,7]} = 6.28$	$F_{[5,7]} = 0.24$	$F_{[5,7]} = 0.02$	$F_{[5,7]} = 0.00$
Acanthopodian amoebae	$F_{[5,7]} = 1.58$	$F_{[5,7]} = 0.74$	$F_{[5,7]} = 3.44$	$F_{[5,7]} = 1.42$	$F_{[5,7]} = 0.25$
Monotactic amoebae	$F_{[5,7]} = 3.19$	$F_{[5,7]} = 0.58$	$F_{[5,7]} = 1.65$	$F_{[5,7]} = 0.98$	$F_{[5,7]} = 7.72$
Testate amoebae	$F_{[5,7]} = 2.28$	$F_{[5,7]} = 10.28$(*)	$F_{[5,7]} = 2.26$	$F_{[5,7]} = 2.36$	$F_{[5,7]} = 5.28$
Hartmannelid-like amoebae	$F_{[5,7]} = 0.11$	$F_{[5,7]} = 0.90$	$F_{[5,7]} = 0.01$	$F_{[5,7]} = 0.13$	$F_{[5,7]} = 0.54$
Fan-shaped amoebae	$F_{[5,7]} = 0.06$	$F_{[5,7]} = 5.44$	$F_{[5,7]} = 4.36$	$F_{[5,7]} = 0.27$	$F_{[5,7]} = 3.62$
Other amoebae	$F_{[5,7]} = 0.02$	$F_{[5,7]} = 0.17$	$F_{[5,7]} = 0.05$	$F_{[5,7]} = 0.15$	$F_{[5,7]} = 0.00$
Ciliates	$F_{[5,7]} = 12.93$(*)	$F_{[5,7]} = 9.54$(*)	$F_{[5,7]} = 3.44$	$F_{[5,7]} = 2.49$	$F_{[5,7]} = 14.06$(*)

[a] Grassland or forest site
[b] Extensive or intensive land use

Appendix Chapter 2

Appendix Table 2.1. Morphometric data of the novel species

Species		Mean	Min	Max	Median	SD	SE	CV [%]	n
Cercomonas jendrali	Cell length [µm]	20.2	8.9	26.1	21.2	5.6	1.9	27	9
	Cell width [µm]	11	4.7	15	11.3	3.2	1.1	29	9
	Length of anterior flagellum	31.7	16.5	36.9	33.5	6.5	2.2	20	9
	Length of posterior flagellum	26.4	26.4	13.3	32	29.2	6.3	2.2	23
	Gliding speed [µm/s]	0.9	0.6	1.8	0.7	-	-	-	4
	Beating frequency [Hz]	1.3	0.5	1.9	1.4	-	-	-	4
	Diameter of cysts [µm]	13.1	11.3	14.7	13.3	1.3	0.49	9.9	7
Paracercomonas bassi	Cell length [µm]	10.8	7.4	14.5	11.8	2.5	0.83	23	9
	Cell width [µm]	6.8	4.9	8.7	6.7	1.2	0.40	17	9
	Length of anterior flagellum	13	6.4	16.3	11.8	3.2	1.1	24	9
	Length of posterior flagellum[µm]	16.7	16.7	7.2	20.2	18.2	4.0	1.3	24
	Gliding speed [µm/s]	1.3	1.1	1.6	1.3	-	-	-	4
	Beating frequency [Hz]	1.4	0.8	2.4	1.1	-	-	-	5
	Diameter of cysts [µm]	8.2	6.7	11.3	8.1	1.2	0.30	14	15

Appendix Table 2.2. Comparative table with the general morphological, behavioral and life cycle characteristics of the novel cercomonad species. The metabolicity refers to the capacity of shape changing and pseudopodium formation, regardless of the speed of these processes; the speed of shape changing refers to the latter. The types of pseudopodia and their locations are given in the same rows.

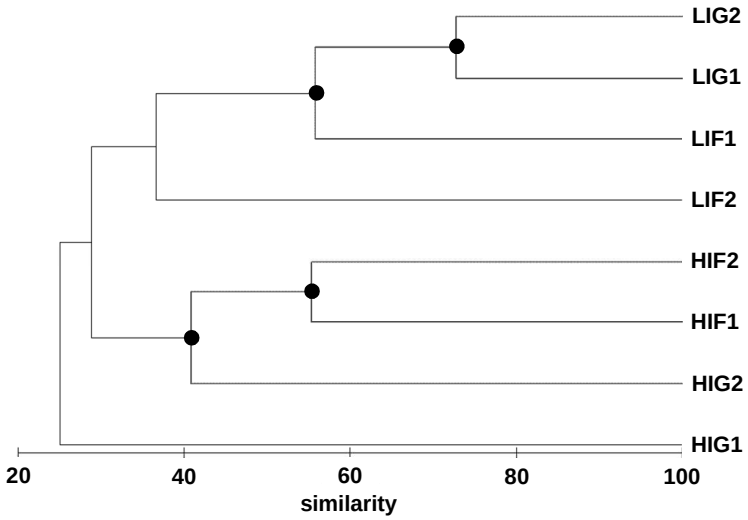
Morphological characters	***Cercomonas jendrali***	***Paracercomonas bassi***
cell shape	elongated oval; oblong; irregularly amoeboid	drop shaped; fusiform; triangular
relative length of anterior flagellum	1.5-1.9	0.9-1.6
relative length of posterior flagellum	1.1-1.6	1-2.2
beating pattern of anterior flagellum	probing, distal part flickers	AF stiff, distal part beating, whole flagellum flaps sporadically
metabolicity	high	high
speed of shape changing	slow	relatively fast
pseudopodium types	lamellar, finger-like, bulbous	finger-like, bulbous filose cytoplasmic tail; lamellar spatulate
pseudopodium location	left lateral	posterior
	posterior	left lateral
gliding direction	many shifts	straight, but changes often
CV position	anterior, lateral, posterior	anterior-lateral
cyst	irregular swellings	spherical
cyst wall	creased, occasion. double layered	smooth
cystoplasm attached	attached	attached
life cycle	gliding	gliding
	unattached, non-gliding, wobbling	unattached, non-gliding, wobbling

Appendix Chapter 3

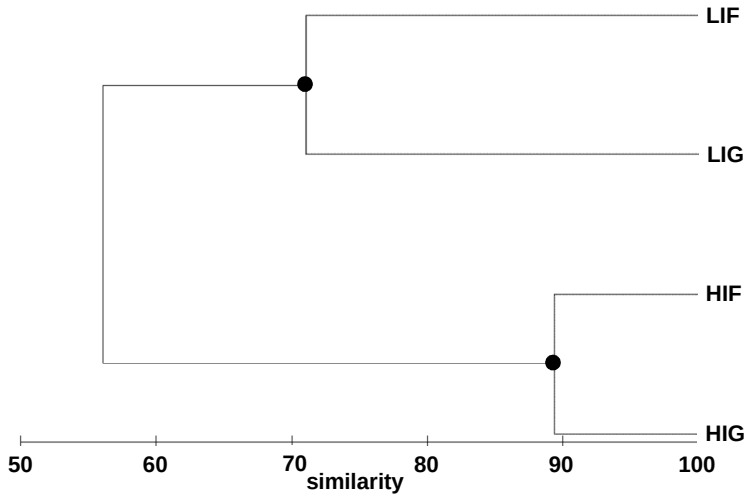
Appendix Table 3.1. Means of observed DNA contribution and observed PCR contributions in comparison to expected contributions of 0.2%, 1.9%, 96.2% and 99.6%. Contributions were arcsin-transformed before analysis of Friedman's Chi^2 test.

	Observed DNA	Observed PCR product
Expected 0.2%	Chi^2=0.67; p=0.414	Chi^2=2.67; p=0.102
Expected 1.9%	Chi^2=0.67; p=0.414	Chi^2=0.67; p=0.414
Expected 96.2%	Chi^2=0.33; p=0.564	Chi^2=0.33; p=0.564
Expected 99.6%	Chi^2=0.33; p=0.564	Chi^2=3.00; p=0.083

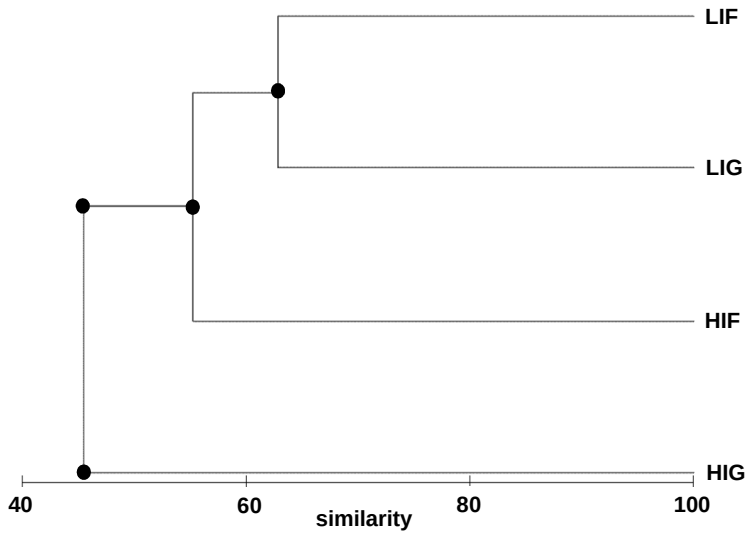
Appendix Chapter 4



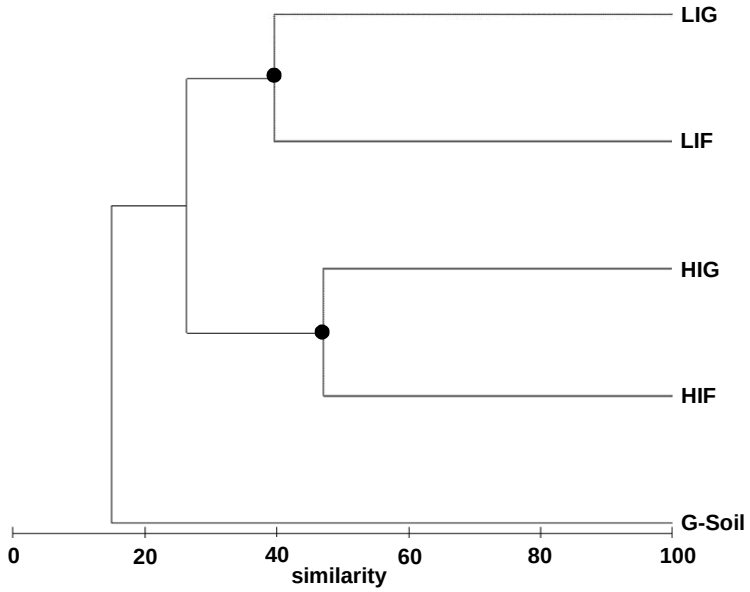
Appendix Figure 4.1. Similarity of distinct species at 100% sequence identity in the replicates of grassland and forest. Dendrogram resulted from distance matrices which were calculated based on Similarity of distinct species. Dendrogram resulted from distance matrices which were calculated by Bray-Curtis similarity using weighted matrix based on species presence/absence transformed data. Significant clusters were indicated by dots, significance level was set at p<0.001 (paired t-test). The distance matrix was calculated by 1000 repeats.



Appendix Figure 4.2. Similarity of flagellated cercozoan soil communities in grassland and forest. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weighted matrices based on genera log(X+1) transformed abundances. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.



Appendix Figure 4.3. Similarity of amoeboid cercozoan soil communities in grassland and forest. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weighted matrices based on genera log(X+1) transformed abundances. Significant clusters were indicated by dots, significance level was set at p<0.05. The distance matrix was calculated by 1000 repeats.



Appendix Figure 4.4. Similarity of protistan species soil communities in grassland and forest of Thuringia and Greenland. Dendrogram resulted from distance matrices which were calculated based on Bray-Curtis similarity using weighted matrices based on species log(X+1) transformed abundances. Significant clusters were indicated by dots, significance level was set at p<0.02 (paired t-test). The distance matrix was calculated by 1000 repeats.

Appendix Table 4.1. Site type (grassland/forest), intensity of land use (low/high), site type in dependence of intensity of land use compared to the percentage contribution of protistan species to total protistan species numbers and protistan abundances. Values of percentage contribution were arcsin-transformed and values of abundances were log(X+1) transformed before being analyzed. F-values of ANOVA on the effect of parameters on different protists presented here. Significant differences were indicated by stars, with p<0.05 (*) and p<0.01 (**).

	Percentage contribution			Abundance		
Species	**Site type**	**Land use**	**Site type x land use**	**Site type**	**Land use**	**Site type x land use**
Achlya apiculata	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Allapsidae	$F_{[1,4]}$=0.32	$F_{[1,4]}$=0.32	$F_{[1,4]}$=1.68	$F_{[1,4]}$=0.58	$F_{[1,4]}$=0.58	$F_{[1,4]}$=1.42
Undet. Cercomonadidae	$F_{[1,4]}$<0.001	$F_{[1,4]}$= 1.10	$F_{[1,4]}$=0.11	$F_{[1,4]}$=0.05	$F_{[1,4]}$=2.38	$F_{[1,4]}$=0.05
Cercomonas media	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Cercomonas pellucida	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Cercomonas sp.	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Chlorellidium tetrabotrys	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Clade-C sp.	$F_{[1,4]}$=8.36*	$F_{[1,4]}$=7.61	$F_{[1,4]}$=2.96	$F_{[1,4]}$=1.28	$F_{[1,4]}$=0.20	$F_{[1,4]}$=1.95
Clade-H .sp.	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Colpoda sp.	$F_{[1,4]}$=0.88	$F_{[1,4]}$=1.12	$F_{[1,4]}$=0.88	$F_{[1,4]}$=0.17	$F_{[1,4]}$=1.83	$F_{[1,4]}$=0.17
Undet. Colpodidae	$F_{[1,4]}$=0.06	$F_{[1,4]}$=1.94	$F_{[1,4]}$=0.06	$F_{[1,4]}$=0.004	$F_{[1,4]}$=2.00	$F_{[1,4]}$= 0.004
Dictyostelium sphaerocephalum	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Eocercomonas expoloratorii	$F_{[1,4]}$=0.11	$F_{[1,4]}$=1.89	$F_{[1,4]}$=0.11	$F_{[1,4]}$=0.25	$F_{[1,4]}$=1.75	$F_{[1,4]}$=0.25
Eocercomonas ramosa	$F_{[1,4]}$=0.40	$F_{[1,4]}$=0.09	$F_{[1,4]}$=1.31	$F_{[1,4]}$=0.03	$F_{[1,4]}$=0.57	$F_{[1,4]}$=0.57
Flectomonas ekelundi	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Glissomonadida	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Gregarines	$F_{[1,4]}$=1.85	$F_{[1,4]}$=0.15	$F_{[1,4]}$=0.15	$F_{[1,4]}$=1.53	$F_{[1,4]}$=0.47	$F_{[1,4]}$=0.47
Undet. Haptoria	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Hyphochytrium catenoides	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Limnofila anglica	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Limnofila longa	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Limnofilidae	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Metabolomonas insania.	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Granofilosea Gran 5	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Oomyceta	$F_{[1,4]}$=21.51**	$F_{[1,4]}$=16.49*	$F_{[1,4]}$=23.04**	$F_{[1,4]}$=1.82	$F_{[1,4]}$=0.01	$F_{[1,4]}$=0.84
Undet. Oxytrichidae	$F_{[1,4]}$=0.46	$F_{[1,4]}$=1.54	$F_{[1,4]}$=0.46	$F_{[1,4]}$=0.04	$F_{[1,4]}$=1.96	$F_{[1,4]}$=0.04
Paracercomonas sp.	$F_{[1,4]}$=0.02	$F_{[1,4]}$=1.98	$F_{[1,4]}$=0.02	$F_{[1,4]}$=0.15	$F_{[1,4]}$=1.85	$F_{[1,4]}$=0.15
Paracercomonas vonderheydeni	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Pedospumella encystans	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Pseudotetracystis sp.	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Pythiopsis cymosa	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Undet. Rhogostoma	$F_{[1,4]}$=5.59	$F_{[1,4]}$=3.89	$F_{[1,4]}$=3.50	$F_{[1,4]}$=1.26	$F_{[1,4]}$=10.47*	$F_{[1,4]}$=1.37
Sandona sp.	$F_{[1,4]}$=0.40	$F_{[1,4]}$=1.23	$F_{[1,4]}$=0.11	$F_{[1,4]}$=0.03	$F_{[1,4]}$=2.33	$F_{[1,4]}$=0.04
Undet. Sandonidae	$F_{[1,4]}$=3.53	$F_{[1,4]}$=3.33	$F_{[1,4]}$=0.12	$F_{[1,4]}$=3.63	$F_{[1,4]}$=24.21**	$F_{[1,4]}$=1.08
Spumella sp.	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Tetracystis vinatzeri	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Thaumatomonas sp.	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1
Theratromyxa weberi	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=1	$F_{[1,4]}$=0.01	$F_{[1,4]}$=1.99	$F_{[1,4]}$=0.01
Undet. Trinematidae	$F_{[1,4]}$= 3.09	$F_{[1,4]}$= 3.09	$F_{[1,4]}$= 3.09	$F_{[1,4]}$=21.62**	$F_{[1,4]}$=21.62**	$F_{[1,4]}$=21.62**

Appendix Table 4.2. Spearman's rank correlation between the protistan species abundance and presence and the different soil properties of the soil sampling sites. The correlation of species was performed by the corresponding protistan phyla. Moreover, total protistan species richness and abundance was correlated. Bold numbers and star indicate a significant correlation of p<0.05.

Group	Phylum	Soil Anorganic Mass		Soil Organic Mass		Total Bacteria		C/N		pH		Soil moisture content	
		Abundance	Presence	Abundance	Presence	Abundance	Presence	Abundance	Presence	Abundance	Presence	Abundance	Presence
Stramenopiles	Synurophyceae	-0.05	-0.25	0.05	0.25	0.31	0.08	**-0.79 ***	-0.58	0.1	0.41	-0.31	0.08
	Hyphochytriomyceta	-0.25	-0.25	0.25	0.25	-0.25	-0.25	-0.58	-0.58	-0.08	-0.08	0.25	0.25
	Oomyceta	0.07	-0.08	-0.07	0.08	-0.48	-0.08	0.6	0.25	0.02	-0.25	0.67	0.41
	Xanthophyceae	-0.41	-0.41	0.41	0.41	0.41	0.41	-0.41	-0.41	0.08	0.08	-0.58	-0.58
Alveolata	Apicomplexa	-0.55	-0.5	0.55	0.5	**-0.33 ***	**-0.25 ***	0.23	0.25	0.06	0.13	**0.76 ***	**0.76 ***
	Colpodea	-0.52	-0.62	0.52	0.62	-0.11	-0.06	-0.35	-0.39	0.08	0.06	-0.25	-0.28
	Litostomatea	-0.08	-0.08	0.08	0.08	0.25	0.25	0.25	0.25	0.41	0.41	0.41	0.41
	Spirotrichea	**-0.73** *	**-0.76 ***	**0.73 ***	**0.76 ***	-0.03	-0.13	-0.3	-0.25	-0.09	-0.13	-0.11	0
Rhizaria	Endomyxa	0.44	0.38	-0.44	-0.38	-0.19	-0.13	0.51	0.5	**0.76 ***	**0.76 ***	0.06	0.13
	Granofilosea	-0.41	-0.13	0.41	0.13	0.41	0.25	-0.41	0.13	0.08	-0.25	-0.58	-0.5
	Imbricatea	-0.35	-0.51	0.35	0.51	-0.08	-0.17	0.25	0.17	-0.41	-0.39	-0.16	-0.06
	Sarcomonadea	-0.55	-0.58	0.55	0.58	0.17	0.41	-0.29	-0.41	-0.57	-0.58	-0.05	0.25
	Thecofilosea	-0.16	-0.25	0.16	0.25	0.35	0.5	0.34	0.13	0.05	-0.38	-0.01	0
Amoebozoa	Mycetozoa-Dictyostelea	-0.41	-0.41	0.41	0.41	0.41	0.41	-0.41	-0.41	0.08	0.08	-0.58	-0.58
Archaeplastida	Chlorophyceae	-0.41	-0.41	0.41	0.41	0.41	0.41	-0.41	-0.41	0.08	0.08	-0.58	-0.58
Protistan species richness		-0.66		0.66		0.37		-0.3		-0.46		0.22	
Protistan species abundance		-0.52		0.52		0.19		-0.17		-0.19		-0.05	

Appendix Chapter 5

Appendix Table 5.1. Comparison of the different Greenland habitats (soil, ice and snow) regarding the calculated number of taxa (S), Shannon index (H) and Evenness (E). The calculations were made on the level of group (80% sequence identity), genus (98% sequence identity) and species (99.7% sequence identity). The calculations and comparisons were made for protists and for all eukaryotes. F-values of a t-test on the variance of means were presented here. Significant differences were indicated by stars, with p<0.05 (*), p<0.01 (**) and p<0.001 (***).

	Protists			All Eukaryotes		
Richness (S)	Group	Genus	Species	Group	Genus	Species
Soil/Ice	$F_{[2,2]}$=2.25*	$F_{[2,2]}$=39.25	$F_{[2,2]}$=1.00	$F_{[2,2]}$=2.25*****	$F_{[2,2]}$=1.32*****	$F_{[2,2]}$=2.08
Soil/Snow	$F_{[2,2]}$=3.00	$F_{[2,2]}$=1.60	$F_{[2,2]}$=1.75	$F_{[2,2]}$=1.33	$F_{[2,2]}$=7.87	$F_{[2,2]}$=19.36
Snow/Ice	$F_{[2,2]}$=1.33******	$F_{[2,2]}$=63.00	$F_{[2,2]}$=1.75	$F_{[2,2]}$=3.00*****	$F_{[2,2]}$=10.40	$F_{[2,2]}$=40.33
Shannon index (H)	Group	Genus	Species	Group	Genus	Species
Soil/Ice	$F_{[2,2]}$=301.5	$F_{[2,2]}$=425.4******	$F_{[2,2]}$=53.43	$F_{[2,2]}$=1.31	$F_{[2,2]}$=5.52*****	$F_{[2,2]}$=2.17
Soil/Snow	$F_{[2,2]}$=1.15	$F_{[2,2]}$=5.93	$F_{[2,2]}$=1.06	$F_{[2,2]}$=1.89	$F_{[2,2]}$=7.91	$F_{[2,2]}$=7.42
Snow/Ice	$F_{[2,2]}$=263.3*****	$F_{[2,2]}$=25.22	$F_{[2,2]}$=50.54	$F_{[2,2]}$=1.44	$F_{[2,2]}$=43.65	$F_{[2,2]}$=16.12
Evenness (E)	Group	Genus	Species	Group	Genus	Species
Soil/Ice	$F_{[2,2]}$=52.11	$F_{[2,2]}$=4.87*******	$F_{[2,2]}$=21.27	$F_{[2,2]}$=1.40	$F_{[2,2]}$=3.26	$F_{[2,2]}$=44.42
Soil/Snow	$F_{[2,2]}$=1.74	$F_{[2,2]}$=11.80	$F_{[2,2]}$=1.23	$F_{[2,2]}$=2.06	$F_{[2,2]}$=1.03	$F_{[2,2]}$=2.30
Snow/Ice	$F_{[2,2]}$=30.02*****	$F_{[2,2]}$=57.48*****	$F_{[2,2]}$=26.23	$F_{[2,2]}$=1.47	$F_{[2,2]}$=3.35	$F_{[2,2]}$=102.0

Teilpublikationen

Domonell, A., Brabender, M., Nitsche, F., Bonkowski, M., Arndt, H., 2013. Community structure of cultivable protists in different grassland and forest soils of Thuringia. Pedobiologia 56, 1-7.

Domonell, A., Arndt, H., Nitsche, F., (subm.). Critical evaluation of next generation sequencing to identify in-vitro simulated flagellate community structures. PLoS One.

Brabender, M., Kiss, Á.K., Domonell, A., Nitsche, F., Arndt, H., 2012. Phylogenetic and morphological diversity of novel soil cercomonad species with a description of two new genera (*Nucleocercomonas* and *Metabolomonas*). Protist 163, 495-528.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen - , die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Hartmut Arndt betreut worden.

Köln, den 02. September 2013