UNIVERSITÉ DE NEUCHÂTEL

FNSNF SWISS NATIONAL SCIENCE FOUNDATION

UNIVERSITY OF NEUCHÂTEL

PHD THESIS

to obtain the title of

Doctor of Philosophy in Biological Science

of the University of Neuchâtel

A molecular approach to microeukaryotic diversity, ecology and biogeography associated with *Sphagnum* mosses

Defended by

David Singer

Thesis Advisor: Enrique Lara
defended on February 16, 2017

Jury

Dr. Enrique LARA - University of Neuchâtel
Pr. Edward MITCHELL - University of Neuchâtel
Dr. Micah DUNTHORN - University of Kaiserslautern
Dr. Diego FONTANETO - Institute of Ecosystem Study, CNR
Dr. Thomas POSCH - University of Zürich
Pr. Sergio RASMANN - University of Neuchâtel

UNIVERSITÉ DE NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur David SINGER

Titre:

"A molecular approach to microeukaryotic diversity, ecology and biogeography associated with *Sphagnum* mosses"

sur le rapport des membres du jury composé comme suit:

- Dr Enrique Lara, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Edward Mitchell, co-directeur de thèse, Université de Neuchâtel, Suisse
- Prof. ass. Sergio Rasmann, Université de Neuchâtel, Suisse
- Dr Micah Dunthorn, Université de Kaiserslautern, Allemagne
- Dr Thomas Posch, Université de Zürich, Suisse
- Dr Diego Fontaneto, Italian National Research Council, Italie

Neuchâtel, le 31 mai 2017

Le Doyen, Prof. R. Bshary

"It's a dangerous business, Frodo, going out your door. You step onto the road, and if you don't keep your feet, there's no knowing where you might be swept off to."

J.R.R. Tolkien, The Lord of the Rings

"Biology sometimes reveals its fundamental principles through what may seem at first to be arcane and bizarre."

Elizabeth Blackburn

Contents

Chapter 1

Abstract/Résumé

Abstract

Despite the fact that free-living microeukaryotes compose the major part of Earth's biodiversity and play numerous essential roles in ecosystems, knowledge on their true diversity, ecology and their global patterns of distribution remain limited. In this sense, the objectives of this thesis are 1) to increase the knowledge on the diversity of microeukaryotes 2) characterize the ecological preferences and determine which are the main variables that influence community composition, and finally 3) to understand the rules that shape the communities at both local and global scales. To meet these objectives a specific component of the earth surface was selected: the "Sphagnosphere" i.e. the interstitial water directly influenced by *Sphagnum* mosses. This understudied but unique microenvironment is characterized by low nutrient contents, low pH, and high amounts of organic acids produced by the mosses. It is also very stable over time.

We first explored the diversity of two groups of protists in *Sphagnum* peatlands. The first group was genus *Nebela* (Arcellinida, Hyalospheniidae), a common testate amoeba taxon in acidic soils. We formally described the most abundant one and named it *Nebela gimllii* due to the small and stout shells. The different community profiles revealed that species are not randomly distributed among microhabitats in peatlands. Instead, we observed a strong phylogenetic clustering in nitrogen-poor areas suggesting that little amounts of nitrogen exerted strong environmental filtering. We also surveyed the molecular diversity of Oomycota, a clade of fungi-like stramenopiles which enclose many animal, fungi and plant parasites, as well as saprotrophic species. We revealed a high diversity, which was unexpected for osmotrophic organisms in nutrient-poor habitats unless most are parasitic. Moreover, most phylotypes found were not recorded in previous studies, which suggest the existence of highly specialized organisms.

We also surveyed the diversity of microbial eukaryotes along altitudinal gradients in three different climatic zones, temperate (western Alps), subtropical (Japan) and tropical (Costa Rica). We showed that 25 percent of phylotypes were shared in the three climatic zones. We found also a significant negative correlation between the proportion of phylotypes related to mixotrophic organisms and temperature. This, in line with other lines of evidence in the literature corroborates the idea that mixotrophy is disadvantageous under warm climates. Finally, we studied the spatial distribution of an emblematic morphospecies of testate amoeba found in the northern hemisphere peatlands: *Hyalosphenia papilio*. A total of 13 lineages were found, from which nine showed narrowly restricted distributions, and four were well distributed across the Holarctic realm. We showed, based on phylogenetic analyses and ancestral character reconstructions that *H. papilio* most probably appeared somewhere in the West Coast of North America.

In summary, my PhD revealed that the Sphagnosphere environment hosts high and unique diversity. This diversity is driven by physicochemical factors at the local scale and by climate and geographical distance at the global scale. We identified and quantified the main local abiotic variables, amongst which micro-topography and nitrogen content appeared to be the most significant in shaping micro-eukaryotic diversity within the same climate zone. These variables exerted strong environmental filtering, which appeared to be fundamental process of community assembly. On the other hand, at a global scale, we demonstrated that temperature was the factor that best explain community composition, and notably the abundance of mixotrophs (and hence a different functioning). At both scales, community composition, and therefore biotic interactions (and most probably ecosystem functioning) change drastically.

Keywords

Microeukaryotes, Protists, Diversity, Ecology, Biogeography, Peatland, *Sphagnum* sp., Bryosphere, *Nebela collaris* s.I., Oomycetes, *Hyalosphenia papilio*, mitochondrial cytochrome oxidase c 1 (COI) gene, V9 SSU rRNA gene, Next generation sequencing, Metabarcoding, Cloning, High Throughput Sequencing (HTS), Sphagnosphere

Résumé

Malgré le fait que les micro-eucaryotes composent la majeure partie de la biodiversité terrestre et jouent de nombreux rôles essentiels dans le maintien des écosystèmes, la connaissance de leur diversité, de leur écologie ainsi que de leurs aires de répartition reste très lacunaire. Dans ce sens, les objectifs de cette thèse sont 1) d'accroître la connaissance de la diversité des micro-eucaryotes 2) de caractériser les préférences écologiques et de déterminer quelles sont les principales variables qui influencent la composition des communautés et enfin 3) de comprendre les règles qui dirigent les communautés à l'échelle locale et globale. Pour atteindre ces objectifs, un milieu spécifique a été sélectionné : la "sphagnosphère", celui-ci désigne l'eau interstitielle sous l'influence des mousses de sphaignes (*Sphagnum*). Cet environnement est un excellent modèle en biologie car il se caractérise par une faible teneur en éléments nutritifs, un faible pH, des quantités élevées d'acides organiques et une grande stabilité dans le temps.

Nous avons d'abord exploré la diversité de deux groupes de protistes vivant dans les sphaignes. Le premier groupe est le genre *Nebela* (Arcellinida, Hyalospheniidae), un groupe d'amibes à thèque composé d'espèces étroitement apparentées. Nous avons décrit formellement la plus abondante et l'avons nommée *Nebela gimllii* en raison de la taille de sa thèque. Les différents profils de communautés ont révélé que les espèces ne sont pas distribuées de manière aléatoire dans les tourbières. Au contraire, nous avons observé un fort groupement phylogénétique dans les zones oligotrophes, ce qui suggère que les teneurs faibles en azote exercent une forte pression environnementale. Nous avons également étudié la diversité moléculaire du clade d'Oomycota. Ce sont des stramenopiles qui se composent de nombreux parasites d'animaux, de champignons et de végétaux, ainsi que d'espèces saprotrophes. Nous avons révélé une grande diversité dans ce clade ce qui était inattendu pour des organismes osmotrophes vivant dans des habitats oligotrophes. De plus, la plupart des phylotypes trouvés ne sont pour le moment pas décrits morphologiquement ni génétiquement, ce qui suggère l'existence d'organismes hautement spécialisés.

Nous avons également étudié la diversité des micro-eucaryotes vivant dans des Sphaignes situées à différentes altitudes dans trois zones climatiques différentes : tempérée (Suisse-France-Italie), subtropicale (Japon) et tropicale (Costa Rica). Nos résultats suggèrent que 25% des phylotypes étaient communs dans ces trois zones. Nous avons également trouvé une corrélation significativement négative entre la quantité de phylotypes liés aux organismes mixotrophes et des températures élevées. Cela suggère que la mixotrophie est désavantageuse dans un climat chaud. Enfin, nous avons étudié la répartition spatiale d'une espèce emblématique d'amibe à thèque trouvé dans les tourbières de l'hémisphère nord: *Hyalosphenia papilio*. Un total de 13 lignées ont été trouvées, dont neuf présentent des distributions restreintes et quatre sont bien réparties dans tout le domaine holarctique. Nous avons montré, sur la base de reconstructions phylogénétiques et d'une reconstitution des caractères ancestraux, que l'origine de *H. papilio* se situe probablement sur la côte ouest de l'Amérique du Nord.

En résumé, ma thèse démontre que l'environnement « sphagnosphère » accueille une diversité élevée et unique de micro-eucaryotes. Cette diversité est influencée par des variables environnementales physicochimiques à l'échelle locale mais également par le climat et la distance géographique à l'échelle mondiale. Nous avons identifié et quantifié les principales variables abiotiques locales (à savoir la microtopographie et la teneur en azote) qui influencent fortement les communautés au sein d'une même zone climatique. Ces variables ont exercé un fort effet de filtre environnemental, qui semble être un processus fondamental dans la mise en place des communautés. De plus, à l'échelle mondiale, nous avons démontré que la température était le principal paramètre influençant la composition de la communauté, et notamment l'abondance mixotrophique. Aux deux échelles, la composition des communautés, et donc les interactions biotiques (et probablement le fonctionnement des écosystèmes), changent radicalement.

Mots clés

Microeucaryotes, Protistes, Diversité, Ecologie, Biogéographie, Tourbière, *Sphagnum* sp., Bryosphère, *Nebela collaris* s.I., Oomycètes, *Hyalosphenia papilio*, Mitochondrial Cytochrome oxydase c 1 (COI), gène V9 SSU rRNA, métabarcoding, clonage, séquençage à haut débit (HTS), Illumina HiSeq, Sphagnosphère

Chapter 2

Introduction

Diversity

Concept of species

In order to understand the interactions between organisms and their environments, it is important to list and classify all these living forms in a coherent and global system. In this purpose, Karl von Linné introduced a classification system in the 18th century that set the base to the modern taxonomy. This system is based on the hierarchical classification of organisms, with the species as the central element (Linnaeus, 1788). However, it was not until the 19th century and the famous natural selection theory issued by Charles Darwin that the study of diversity was framed in an evolutionary perspective (Darwin, 1872). The zoologist Ernst Mayr wrote that "a species is not just a group of morphologically similar individuals but a group that can breed only among themselves excluding the others" introducing the interfertility as a key factor to define a species (Mayr, 1942) and introduced the biological species concept. In line with this, he postulated that when populations within a species become isolated by geography, feeding strategy, mate choice, or other means, they may start to differ from other populations through a combination of genetic drift and natural selection, and may evolve over time into new species (Mayr, 1963). This statement includes the two main speciation modes, allopatric (Lande, 1980) and sympatric speciation (Smith, 1966). However, a major limitation of this species concept is that it is not applicable to the many organisms which reproduce by inbreeding or asexually or for which the life cycle is unknown; this limitation applies to many protists (Sonneborn, 1957). When working with organisms that do not breed under laboratory conditions (like many protist species documented in this thesis), separations between biological species is impossible to prove. Therefore, this concept, although appealing due to its intuitive nature, is unpractical for most protist species.

The morphological species concept grew from the view that morphological traits have been considered traditionally as decisive criterion of species. This concept was useful for taxonomists to catalog the species in keys and in collections (Mayr, 1996). However the definition of a species based only on morphological characters is subjective and researchers may disagree on which features to use. Moreover cryptic diversity is particularly common among microeukaryotes where "real" species do not differ at all morphologically or only very slightly (Sonneborn, 1975).

Another species concept has proven useful for micro-organisms, the ecological species concept. Following this concept, a species is a lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range (Van Valen, 1976). In other terms, a species is a set of organisms adapted to a particular niche. It is a concept that needs to be used in combination with others (Van Valen, 1976).

To overcome these issues, a genetic species concept has been proposed. The principle lies in the use of genotypic clusters to define species (Mallet, 1995; Masters and Spencer, 1989). The advantage of this approach is that it relies on objective data (DNA sequences) and that it can be generalized to all types of organisms. A practical development of this concept is the notion of DNA Barcoding. This strategy is based on sequencing a marker gene easy to amplify, and link it to a visual documentation. It was first introduced for metazoa using the first subunit of the mitochondrial cytochrome oxidase (COI) marker (Hebert et al., 2003). Later, this method was adapted to other taxonomic groups, for which in some case better suited markers had to be chosen (Pawlowski et al., 2012). However, this very practical approach has been criticized, notably because a single marker gives only a one-sided, partial view of diversity (Will et al., 2005). Eventually a new, integrative concept of the species was developed, including elements that, taken all together, give an accurate vision of diversity (Hubert and Hanner, 2015).

In this manuscript we chose to follow a three-sided concept of species: morphology, genetics and ecology.

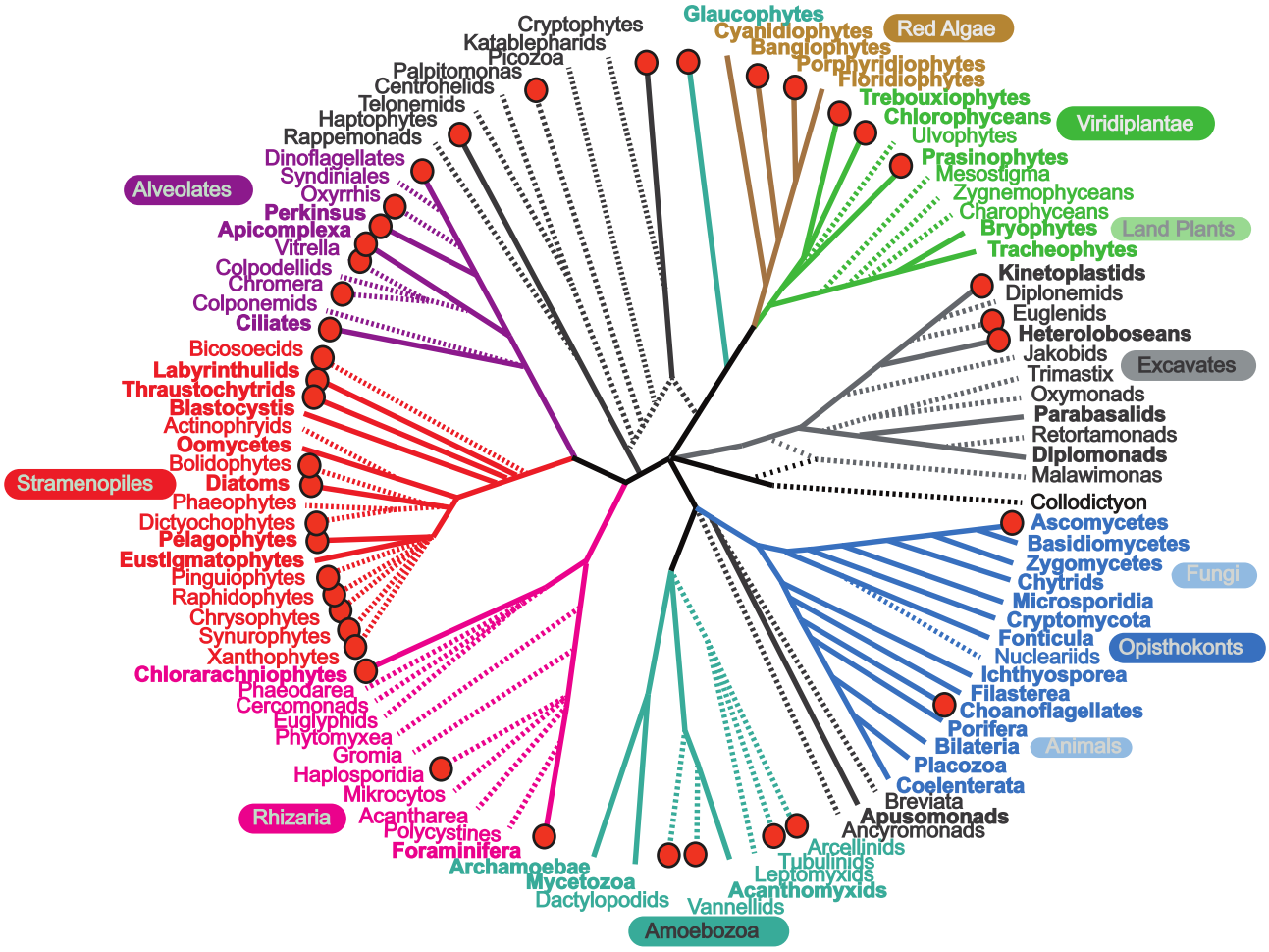

Figure 2.1 – Figure from (Keeling et al., 2014) showing the major lineages in the eukaryotic tree of life, the relationships between lineages for which genomic resources are currently available. Red dots show the lineages that have been targeted by the The Marine Microbial Eukaryote Transcriptome Sequencing Project.

According to (Schlegel and Meisterfeld, 2003) the combination between morphology and genetics is essential as small difference in both components could be interpreted to different species (Kosakyan et al., 2013). But in order to incorporate this species in a global context and understand its relations with the rest of the community, it is important to define its ecological preferences (Finlay, 2004; Škaloud and Rindi, 2013).

Microeukaryotic diversity

The Tree of life (ToL) becomes more and more accurate over time following the development of new methods of analyses. The first attempts to build a phylogenetic tree based on (rRNA) sequence data date back to 25 years ago and the work of (Woese et al., 1990), who revealed the existence of the three domains of life: Bacteria, Archaea and Eukarya. The phylogeny of the latter changed gradually as taxon –and gene sampling increased from a tree with long branches at the base ("Archaezoans", (Sogin and Silberman, 1998)) towards a tree comprising five major clades or "supergroups". These are Opisthokonta, Viridiplantae, Excavata, Amoebozoa and SAR (Stramenopiles, Alveolata and Rhizaria) (Baldauf, 2003; Burki, 2014) (fig. 2.1). Besides these large supergroups, several "microkingdoms" still of uncertain affinities can be found (Pawlowski, 2013). The root of the eukaryotic tree is still one of the "Holy Grails" sought by eukaryotic microbiologists, and several options are still under consideration (Burki, 2014). Within each branch, the majority of eukaryotes are unicellular forms usually referred to as "protists" or if fungal clades are included, as "microeukaryotes". Plants and animals each appear as single branches in the tree, which is resolutely microbial and mainly constituted by protists. Although traditional viewpoints consider animals as the most diversified group on Earth (e.g. (Mora et al., 2011)) recent results tend to prove this assumption wrong. Environmental surveys are currently revealing an immense amounts of diversity in oceans (De Vargas et al., 2015), tropical soils (Appendix E (Mahe et al., 2017), in press) where animal and plants appear overwhelmed under protistan richness. The reasons for the underestimation of eukaryotic diversity are multiple: (1) many symbiotic/parasitic forms exist, and cannot be

readily identify through conventional microscopy. Entire hyperdiverse clades are composed by these organisms (del Campo et al., 2015; Guillou et al., 2008; Lara et al., 2009a) (2) Many free living forms are extremely small and inconspicuous, and escape morphological surveys (Blandenier et al., 2016; Tarnawski and Lara, 2015), or cannot be readily distinguished from surrounding sediments (Berney et al., 2015); and (3) many morphotypes resemble strongly each other but differ genetically and have been erroneously pooled together, sometimes hiding a huge diversity. This is almost a rule in small heterotrophic forms (e.g. (Bass et al., 2009b; Karpov et al., 2006; Koch and Ekelund, 2005)) but also in larger forms such as testate amoebae (Chatelain et al., 2013; Kosakyan et al., 2012). To the contrary, environmental diversity surveys are based on single gene sequences (Blaxter et al., 2005), and are supposed to mirror the richness present. This approach has been applied to the study of environmental prokaryotic diversity since the early nineties (Giovannoni et al., 1990) and is particularly relevant in organisms where gene fluxes do not function in a "conventional" eukaryotic way. It is only at the turn of the century that environmental microbiologists started to focus on eukaryotic diversity, revealing a wealth of novel groups in many environments, starting on aquatic systems (for methodological reasons partly) (Diez et al., 2001; López-García et al., 2001; van Hannen et al., 1999). However, this approach did not allow retrieving a complete picture of environmental diversity as the number of sequences obtained was still too low to cover all organisms present in the environment.

High throughput sequencing (HTS) approaches like 454 pyrosequencing or Illumina appeared in the late 2000's and allowed the generation up to hundreds of millions of sequences per samples. These approaches permitted covering a significant portion of the diversity, including organisms present in very low numbers in environmental samples (i.e. ciliates (Dunthorn et al., 2012), fungi (Buée et al., 2009), plankton (Amaral-Zettler et al., 2009; De Vargas et al., 2015), or in soils (Geisen et al., 2015b)). However, these huge amounts of data needed specific bioinformatics tools to define OTUs that can be interpreted further. To this goal, clustering algorithm have been developed like SWARM (Mahé et al., 2014) or DBC454 (Pagni et al., 2013). However culture independent approaches should still be considered with caution: an inflation of the biodiversity due to sequencing errors has been observed (Kunin et al., 2010). The link between phylotypes diversity and species diversity is still not resolved and remains highly dependent on the studied groups (Nebel et al., 2011). Despite these limitations HTS approaches are promising powerful tools to find new bioindicators for environmental biomonitoring studies (Pawlowski et al., 2016). One of the next challenges in eukaryotic diversity surveys will be to characterize morphologically and functionally the groups for which no information exists (De Vargas et al., 2015; Grossmann et al., 2016). OTUs based solely on SSU have been demonstrated to pool several biological species in many groups (Pawlowski et al., 2012). As this marker is the most commonly used one in eukaryotic environmental DNA surveys, it appears clearly that even HTS-based environmental DNA surveys underestimate diversity. To obtain a correct picture of the (genetic) species level diversity of an environmental sample, other markers need to be used (Pawlowski et al., 2012). Here, sequences presenting infraspecific diversity (haplotypes) must be grouped into genetic species. Different approaches have been developed to evaluate where boundaries between taxa should be set. For example: the percentage of sequence divergence (Kimura, 1980), the Automatic Barcode Gap Discovery (ABGD) (Puillandre et al., 2012), the General Mixed Yule-Coalescent (GMYC) model (Fontaneto et al., 2007; Fujisawa and Barraclough, 2013). These methods are commonly used to evaluate species delineation of various kinds of organisms (i.e. metazoa, (Talavera et al., 2013) plants, (Williams et al., 2015) algae (Leliaert et al., 2014), fungi (Hagen et al., 2015), testate amoeba (Heger et al., 2013), meiofauna (Fontaneto et al., 2015)). Some methods require a phylogenetic reconstruction as input to compute the clustering (i.e. GMYC), which can be an issue with massive dataset containing short length sequences. Barcoding at high phylogenetic resolution is logically very promising in revealing immense levels of environmental diversity. Indeed, what would be considered as a single OTU using SSU rRNA gene as a marker (e.g. sequences within the *Nebela tincta* species complex (Lara et al., 2008), represent probably several different species defined as above (Chapter 3 (Singer et al., 2015)).

Ecology

Fundamental consideration of the "niche" concepts

Since scientists want to understand how the organisms interact with each other and the environment (Sutherland et al., 2013), the evolution of the niche concept has been marked by important theoretical milestones leading to the development of ecology as a science (Chase and Leibold, 2003). This concept evolved and took many meanings, related to the behavior of a species living under specific environmental conditions (Pocheville, 2015). The first attempts to formalize this concept date back to the early 20 th century; the Grinnellian view (Grinnell, 1917), focuses on the habitat where a species lives (environmental conditions). A later development, coined the eltonian niche (Elton, 1927) stressed the importance of biotic interaction with other species (hosts, preys, predators, competitors etc.) and introduced the concept of realized niche, i.e. the practical niche as it can

be observed in nature as a result of interactions with other species. The eltonian niche considers not only the tolerance of a species to a given environment but also the resource availability, the disturbance and the biotic interactions (Ackerly, 2003). A major milestone was attained with the quantitative approach to describing an environmental niche proposed by (Hutchinson, 1957). His concept describes the niche as a hyperspace, merging both previous concepts. His works inspired many others to develop models to explain how similar species can coexist in the same environment leading to the definition of concepts such as "niche breadth", "niche partitioning" and "niche overlap" (Chase and Leibold, 2003). Microeukaryotes interact also with their biotic and abiotic environment, and their immense diversity suggests that these interactions are extremely complex. Whether the same rules that work for macroscopic organisms apply to them is still open to debate. The "paradox of the plankton" is a paradigm that illustrates an alternative functioning. The coexistence of many micro-algal species in a small volume of water can be considered as paradoxical because these organisms apparently share the same homogeneous habitat, and likely the same niche, which should be theoretically impossible (Hutchinson, 1961). Several explanations have been given, most often based on chaotic oscillations of the different populations, which rely on fast population growth and high mortality rates (Scheffer et al., 2003). This paradigm however does not apply to slow-growing organisms, and instances of "classical" niche partitioning have been reported in large protists like foraminiferans (Weiner et al., 2012). Chapter 4 shows a conceptually similar example based on other large and slow-growing protists, the testate amoeba genus *Nebela* in peatlands (Hyalospheniidae; Arcellinida; Amoebozoa)

Functional role of microeukaryotes in the ecosystem

Whatever mechanism applies to the assembly of microeukaryotic communities, their composition influences the function of ecosystems at different scales. Indeed, microeukaryotes are key actors of biogeochemical cycling and are involved in numerous biotic interactions. Autotrophic plankton is responsible for half of global carbon fixation through photosynthesis (Falkowski, 2002). The impact of grazing by small heterotrophic bacterivores and the subsequent release of labile nutrients available to phototrophic organisms (microalgae, plants) has been coined "microbial loop". This concept has been first introduced in oceans (Azam et al., 1983; Pomeroy, 1974) and highlights the importance of micro-organisms in marine food webs. Heterotrophic nanoflagellates prey on bacteria and release nutrients that can be taken up by the phytoplankton. In soils, a similar concept has been introduced, where small bacterial grazers eat bacteria and release labile compounds such as ammonium that stimulate plant growth (Bonkowski and Clarholm, 2012; Clarholm, 1985). In turn, bacteria benefit from the plant's exudates, for which they compete with mycorrhizal Fungi (Bonkowski, 2004). Yet, recent research has highlighted the importance of eukaryvory (the action of eating other eukaryotes) in the input and redistribution of nutrients in soils, for instance by eating Fungi (Dumack et al., 2016b; Geisen et al., 2016). Predation on microalgae present in soils seems to be much more significant than previously thought, and represent therefore certainly another way for nutrient input in soils (Appendix F (Seppey et al., 2017), accepted). And, finally, parasites and parasitoids also play an important roles in the relocation of nutrients (Geisen et al., 2016; Lara et al., 2009b). Microeukaryotes are therefore key actors in soil fertility and plant production at a global scale. Changes in the environment also produce modifications in the communities, and hence in the function of microeukaryotes in the ecosystems. These changes can be induced by human activities, and it can be foreseen that global warning is affecting ecosystems through microbial eukaryotic communities. For instance, an increase in temperature induces planktonic freshwater microalgae to shift from a mixotrophic (i.e. combining photosynthesis and phagocytosis) to a strictly phototrophic behavior (Crane and Grover, 2010; Jones, 2000). Likewise, warming was shown to modify the structure of microbial communities in *Sphagnum*, causing a decrease in mixotrophs and an increase in bacterivores and decomposers (Jassey et al., 2011, 2015), with a probably significant effect on peatlands carbon-storing capacity (Jassey et al., 2015).

Biogeography

The question arises if microeukaryotic communities function in a similar way everywhere on the globe. Indeed, if micro-eukaryotic species are not homogeneously distributed on Earth, one can ask whether the functions are also conserved (Caron, 2009). Up to the early 2000's, the predominant view on eukaryotic biogeography was the old tenet: "everything everywhere, but the environment selects" (Baas-Becking, 1934). In other words, micro-organisms can be found potentially anywhere on Earth as long as their required environmental conditions are met. This implies also that every point on Earth hosts a reservoir of microbial propagules that will potentially develop into large populations when the right conditions are met (Baas-Becking, 1934). Finlay, Fenchel and co-workers defended this position based mostly on freshwater protist (Fenchel, 2005; Finlay, 1998). They stated that the cosmopolitanism of protists was due to their small sizes, huge populations and capacity to form dormant stages that could withstand adverse conditions and be transported over large distances (Fin-

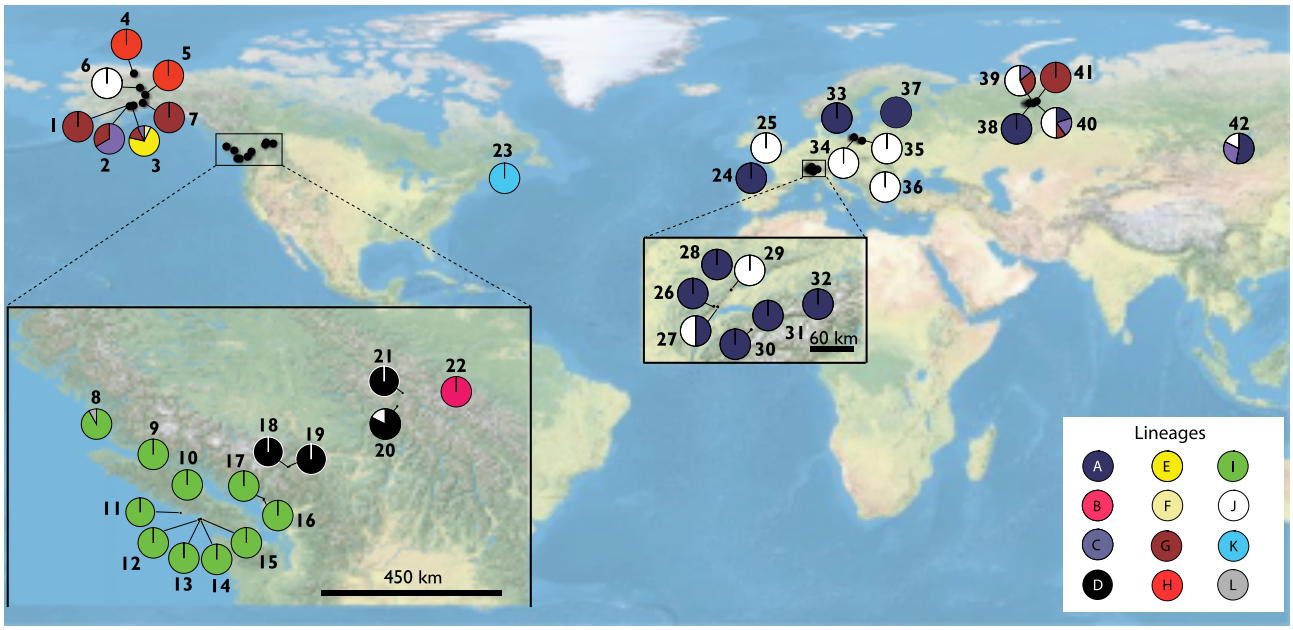

Figure 2.2 – Figure from (Heger et al., 2013) showing the Biogeographical distribution of 301 single-cell isolates of the *Hyalosphenia papilio* morphospecies. Colors and letters indicate the 12 genetic lineages of *H. papilio* shown by phylogenetic analysis of the COI sequences. Pie charts refer to the proportion of sequences representing each genetic lineage at each sampling site.

lay, 2002). As a consequence, because allopatric speciation is considered as an important factor for biological diversification (Mayr, 1942), only sympatric speciation can occur and therefore, global diversity of free-living micro-organisms should be low (Finlay, 1998). In contrast to this dominant view, other authors like Wilhelm Foissner defended the idea that some microeukaryotes at least had a spatially-limited distribution (Foissner, 1999a). They illustrated this viewpoint based on the example given by "flagship species" i.e. species that have an unmistakable morphology (and, thus, could not have been overlooked by previous researchers) and that have a clearly restricted distribution. Such "flagship species" include the characteristic "Gondwanian" testate amoeba *Apodera vas* (Smith and Wilkinson, 2007) and the equally conspicuous "Laurasian" giant ciliate *Bresslauides discoideus* (Foissner, 2006). Based on these findings, a "moderate endemicity" model was proposed, where some (a minority) of protist species have a geographically limited distribution (Foissner, 2006). Still, the causes for such distributions remained to be determined in order to be able to generalize the model. Size seems to influence the ability to disperse, as free-living protists disperse only passively (Wilkinson, 2001; Yang et al., 2010). Models based on atmospheric circulation highlighted the importance of propagule size on dispersal potential, and also demonstrated that distance mattered also for the smallest propagules (Wilkinson et al., 2012).

A weakness of the above-mentioned studies is that inferences on protist distributions rely solely on morphology-based identification. As seen previously, cryptic diversity is extremely common in microeukaryotes, and masks true diversity, but also potentially limited distribution (Mitchell and Meisterfeld, 2005). Molecular-based studies should be theoretically more objective even if the different techniques each present their own biases (already described in environmental diversity paragraph). Different results can be observed with this approach. For example the study of four variable markers on different strains of the marine planktonic micro alga *Micromonas pusilla* failed in showing restricted distribution patterns at a global scale (Šlapeta et al., 2006). All other examples tend to show nevertheless the opposite. The diatom *Sellaphora capitata* showed a geographically structured intraspecific diversity (Evans et al., 2009). Even though some deep-sea benthic foraminiferans were shown to have a bipolar distribution, suggesting gene flow between Arctic and Antarctic populations (Pawlowski et al., 2007), most cold-water species have restricted distributions. Among soil protists, euglyphid testate amoebae distributions appeared to be influenced by classical biogeographical barriers such as the desert belt around the Cancer tropic (Lara et al., 2016). The diversity of these organisms appeared to be influenced by the much documented classical macroecological latitudinal diversity gradient (Lara et al., 2016). Restricted distributions were also shown among lineages of the *Sphagnum* associated testate amoeba *Hyalosphenia papilio* (Heger et al., 2013) (fig. 2.2). Altogether, the evolution of methods of analysis to assess eukaryotic diversity is challenging the old "everything is everywhere" paradigm. However, assessing and predicting the causes and consequences of distribution patterns in microeukaryotes is still an expanding domain (Fontaneto, 2011).

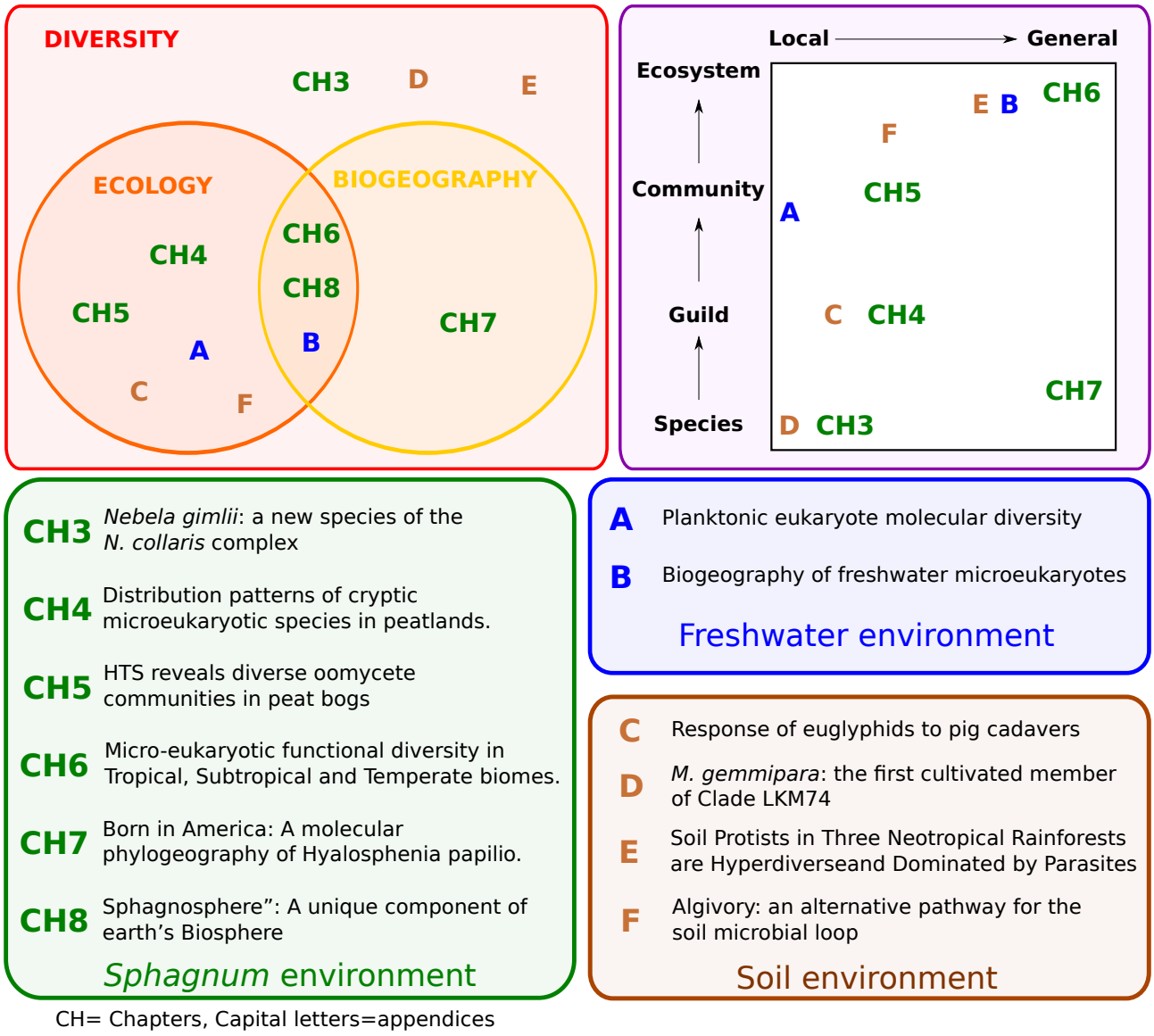

Figure 2.3 – Structure of the thesis showing the relation of the publication and the three major disciplines considered in this PhD (Diversity, Ecology and Biogeography). CH = main chapters of the manuscript (green = *Sphagnum* environment). Capital letters = Appendixes (blue = Freshwater environment and Brown = Soil environment)

Peatlands and *Sphagnum*

By definition, *Sphagnum*-dominated Peatlands is a wet emvironment colonised by a vegetation who growth on a soil with a low permability and formed by a layer of peat of a few tens of centimeters to several meters Manneville (1999). Even though they cover less than four percent of the land surface (Gorham, 1995), peatlands play an important role in the water and carbon biogeochemical cycles (Rezanezhad et al., 2016). These cycles are of considerable significance at a global scale as 30% of the world's pool of organic carbon is stored (Gorham, 1991) through slow accumulation of organic matter as peat (Limpens et al., 2008) and annually up to 12% of all anthropogenic emissions of CO_2 can be sequestered by peatlands (Moore, 2002). However, under a global warming scenario, it has been demonstrated that these carbon sinks revert into carbon sources (Dorrepaal et al., 2009). *Sphagnum* mosses are the main plant producer in high latitude peatlands. *Sphagnum* sp. are viewed as classical 'drought avoiders', avoiding desiccation by the ability to accumulate an important amount of water inside their dead hyaline cells (Hájek and Beckett, 2008). These mosses influence considerably the biological factors and create a homogenous and characteristic microenvironment, which is characterized by very low concentrations of mineral nutrients, low pH values (sometimes below 4.0) (Dedysh et al., 2006) and high concentration of humic acids, some of which are known to have biocidal properties, provoking strong allelopathic effects (Steinberg et al., 2006). These conditions influence microbial community composition, leading to highly specialized assemblages, as shown in prokaryotes (Dedysh et al., 2006), fungi (Thormann and Rice, 2007) and

micro-metazoa (Gilbert et al., 2006). Total microeukaryotic communities surveyed with a molecular approach revealed that this microenvironment has a high and specific diversity (Lara et al., 2011).

The microenvironment created by bryophyte cushions is very specific and is strongly influenced by the plant. This very particular zone of influence has been coined bryosphere (Lindo and Gonzalez, 2010). Because of the very specific physical and physiological particularities of *Sphagnum* mosses, the associated communities are even more constrained (Clymo and Hayward, 1982; Van Breemen, 1995) and the term sphagnosphere has been proposed (Jassey, 2011). The sphagnosphere can be considered as an excellent model to study the microeukaryotic diversity, ecology and biogeography.

Aims and Objectives of the thesis

My PhD research focuses on the diversity, ecology and biogeography of microeukaryotes associated to *Sphagnum* sp. mosses. My thesis has four main overarching aims: 1) increase the global knowledge about the diversity and community composition of microeukaryotes, 2) characterize the ecological preferences and determine which are the main variables that influence community composition, 3) understand the rules that govern protist geographical distribution and finally 4) apply the methods developed in the previous three points to the study of others environments (i.e. freshwater and soils). The logical scheme is (1) a characterization of diversity, which will build the bases for (2) ecological and (3) biogeographical studies. Each chapter of this PhD includes one or several of these aspects according to the subject of the study. The general plan can be viewed in fig. 2.3. The three first aims will be addressed in the five main chapters and in the final discussion, while the fourth will be treated in the six appendices (fig. 2.3). The main chapters include works that I have led (and which result in papers as the first author, published, in press or in preparation). I consider that a scientist should be judged both on his ability in leading projects and collaborating with others. For these reasons I have chosen to add in the appendices the different side projects in which I had the chance to be involved!

Summary of the different chapters and appendices

Chapters

CH3 *Nebela gimlii*: a new species of the *N. collaris* complex: We described here a new species belonging to the *Nebela collaris* species complex (now genus *Nebela* (Kosakyan et al., 2016)) from the Swiss Jura Mountains: *Nebela gimlii*. This species is the smallest representative of the genus and is morphologically and genetically very closely related to *Nebela guttata*. *Nebela gimlii* seems to be found preferentially in forest environment rather in lawns like *Nebela guttata*. This species could improve bioindication and palaeoecological reconstruction studies. *Published in European Journal of Protistology, DOI: 10.1016/j.ejop.2014.11.004*

CH4 Distribution patterns of cryptic microeukaryotic species in peatlands: We investigated the molecular diversity of genus *Nebela* among different peatland microhabitat. We demonstrated that community composition varied strongly across microhabitats and was associated to environmental gradients (nitrogen content and water table depth). Moreover we observed phylogenetic clustering in the most nitrogen-poor areas, supporting the hypothesis of phylogenetic inheritance of adaptations in the group of *Nebela guttata*. *Submitted in Molecular Ecology*

CH5 HTS reveals diverse oomycete communities in peat bog: We explored the molecular diversity of Oomycetes in *Sphagnum* peatlands with a high-throughput sequencing (HTS) approach (Illumina HiSeq) in different peat bogs microhabitats. We found an unexpected diversity of oomycetes which was strongly correlated to microhabitat, including unknown basal clades and obligate plant and animal parasites. Our approach proved effective in screening oomycete diversity and could be applied to other environments. *Published in Fungal Ecology, DOI: 10.1016/j.funeco.2016.05.009*

CH6 Micro-eukaryotic functional diversity in Tropical, Subtropical and Temperate climatic zones: We screened the microeukaryotic functional diversity of *Sphagnum* mosses in three different climatic zones: Tropical, subtropical and temperate with a HTS approach (Illumina HiSeq). We evaluated the effect of the latitudinal and altitudinal effect. We found that only 25 percent of the phylotypes are common to the three biomes, and that climate influences the community composition, OTUs related to mixotrophic protists decreasing in numbers with higher temperatures, thus suggesting a switch in ecosystemfunctioning. This switch may have consequences on peatland functioning in a global warming context. *Submitted in FEMS Microbiology Ecology*

CH7 Born in America: A molecular phylogeography of *Hyalosphenia papilio*: We assessed the diversity of the *Hyalosphenia papilio* complex based on the COI gene and the spatial distribution of the different lineages found across the Holarctic region. We found 13 different lineages; nine of which showed a narrow distribution while four of were broadly distributed all across the Holarctic realm. This evidence in addition to our phylogenetic reconstruction suggests a North American origin for *Hyalosphenia papilio*. *In preparation*

CH8 The "Sphagnosphere": A unique component of earth's Biosphere: In the discussion, I evaluate the state of the art of knowledge on microeukaryotes associated to *Sphagnum* ssp.. I show why the interstitial water of *Sphagnum* ssp. ("Sphagnosphere") is a powerful model to study the diversity, ecology, biogeography, and paleoecology of microeukaryotes. *In preparation*

Appendices

In order to clarify my personal involvement, I summarized my contribution to each study in (table 2.1), I considered here seven distinct steps to lead a project from the conception to the scientific publication: experimental design, sampling, laboratory work, data analysis, figures editing, writing and funding (i.e. use of the funds dedicated to my project). At each stage, different degrees of involvement can be envisaged: Not involved, slightly involved (like a consultant), in charge of one whole part of the project and main responsible of the project.

Table 2.1 – Involvement in the different side project realized during my PhD. - = Not involved, + = involved as consultant, ++ = in charge of one whole part of the project, +++ = in charge of the 100% of the project, *Point remove in the final manuscript

Appendices	Experimental Design	Sampling	Laboratory work	Data analysis	Figures editing	Writing	Funding
A Planktonic eukaryote molecular diversity	+	-	++	++	+	+	++
B Biogeography of freshwater microeukaryotes	-	-	++	++	++	+	++
C Response of euglyphids to pig cadaver	+	-	++	++	++	+	+
D *M. gemmipara*: the first cultivated member of Clade LKM74	-	-	-	++	++	+	+
E Neotropical Rainforests protists are hyyerdiverse	-	+*	+*	++	-	+	-
F Algivory: an alternative pathway for the soil microbial loop	+	++	-	+	+	+	-

A: Planktonic eukaryote molecular diversity: We investigated the composition of eukaryotic plankton communities of five pools located in the Rancho Hambre peat bog in Argentinean Tierra del Fuego with a HTS approach. We showed that autotrophic taxa were more common in minerotrophic environments, whereas mixotrophic taxa were typically best represented in ombotrophic water bodies. *Published in Journal of Plankton Research, DOI: 10.1093/plankt/fbv016*

B: Biogeography of freshwater microeukaryotes: We assessed the distribution patterns of microeukaryotic communities along a latitudinal gradient of water bodies from Argentinean Patagonia to Maritime Antarctica using two different approaches (DGGE and Illumina HiSeq). We observed a decreasing richness with latitude and significant differences between Patagonian and Antarctic lakes communities. Our results suggest the co-existence of a 'core biosphere' containing a reduced number of abundant/dominant OTUs on which classical ecological rules apply, together with a much larger seedbank of rare OTUs driven more by stochastic processes. *Published in Environmental Microbiology, DOI: 10.1111/1462-2920.13566*

C: Response of euglyphids to pig cadaver: We evaluated the euglyphid testate amoeba diversity over time in a decomposing cadaver experiment. Most of the OTUs decreased in abundance over time, but some of them responded positively after a time lag. These results suggest that metabarcoding of soil euglyphids could be used as a forensic tool to estimate the post-mortem interval. After sequencing the full length of SSU rRNA gene of this "dead pig lovers" we observed that these sequences did not belong to any known euglyphid family.*Published in International Journal of Legal Medicine, DOI: 10.1007/s00414-015-1149-7*

D: *Mycamoeba gemmipara*: the first cultivated member of Clade LKM74: We describe a new species belonging to the environmental Dermamoebidae clade LKM74. Environmental sequences of this clade were pervasively reported in soil and freshwater environment. Here we formally described morphologically and genetically *Mycamoeba gemmipara*. We also documented its unique and unusual life cycle. *Published in Journal of Eukaryotic Microbiology, DOI: 10.1111/jeu.12357*

E: Soil protists in three neotropical rainforests are hyperdiverse and dominated by parasites: We highlighted that protist inhabiting Neotropical rainforests are hyperdiverse and dominated by the parasitic Apicomplexa which use arthropods and other animals as hosts. These host-specific protist parasites potentially contribute to the high animal diversity in the forests by reducing their population growth in a density-dependent manner. Our data show that protists play a large role in shaping tropical terrestrial ecosystems long viewed as being dominated by macro-organisms. *Accepted in Nature Ecology and Evolution*

F: Algivory: an alternative pathway for the soil microbial loop: We investigated the taxonomic and functional diversities of soil micro-eukaryotes in three categories of land-use in Switzerland. We found that some phylotypes representing phagotrophic protist were significantly correlated to the total number of photroph sequences. Microscopic observations confirmed that these phylotypes are algal predators. These results suggest that microalga represent a functionally significant input of carbon in soils that should be taken into account when modelling soil nutrient cycling. *Accepted in Soil Biology and Biochemistry*

Chapter 3

Eight species in the *Nebela collaris* complex: *Nebela gimlii* (Arcellinida, Hyalospheniidae), a new species described from a Swiss raised bog

David Singer[1], Anush Kosakyan[1], Amandine Pillonel[1], Edward A.D. Mitchell[1,2], Enrique Lara[1]

[1]Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, Neuchâtel CH-2000, Switzerland
[2]Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

European Journal of Protistology, Vol.51(1), p.79–85, February 2015
DOI: 10.1016/j.ejop.2014.11.004

Abstract

We describe here a new species of sphagnicolous testate amoeba found abundantly in the forested part of the Le Cachot peatland (Jura Mountains, Neuchâtel, Switzerland) based on microscopical observations (LM, SEM). The new species, called *Nebela gimlii* was placed in a phylogenetic tree based on mitochondrial cytochrome oxidase sequences (COI), and branched robustly within the *N. collaris* complex next to the morphologically similar *N. guttata* and *N. tincta*. It is however genetically clearly distinct from these two species, and differs morphologically from them by its smaller size and stouter shape of the shell. This new species completes the phylogeny of the *Nebela collaris* species complex, with now eight species described, mostly from peatlands and acidic forest litter, and further demonstrates the existence of an unknown diversity within testate amoebae. Improving the taxonomy of testate amoebae in peatlands and clarifying the ecology of newly discovered species should make these organisms even more valuable as bioindicators and for palaeoecological reconstruction.

Keywords: Arcellinida; Amoebozoa; Cytochrome oxidase gene (COI); Peatland; Protist; Testate amoeba

Introduction

Arcellinid testate amoebae are common and diverse in peatlands, where they constitute a large part (typically 10–30%) of microbial biomass (Gilbert, 1998; Gilbert et al., 1998b; Mitchell et al., 2003). Their sensitivity to environmental changes and the good preservation of their shells in peat has led to their use as indicators of past environmental changes (Charman, 2001; Mitchell et al., 2008). However, their taxonomy is still far from being resolved in a satisfactory way, and recent studies have revealed a high diversity within individual morphospecies, sometimes referred to as cryptic or pseudocryptic (Kosakyan et al., 2012). A thorough morphological analysis and the application of a single-cell barcoding approach (based on the COI gene) revealed the existence of several morphologically and genetically distinct taxa within the *Nebela collaris* complex (Kosakyan et al., 2013).

Members of the *Nebela collaris* species complex (or *N. tincta* complex) are the second most common group of testate amoebae in peatlands. They were found to occur in 72.6% of all samples in a review of European and North American data and are the most dominant taxa in communities (10.8% of the community on average) (Gilbert et al., 2006). They are found through the Northern Hemisphere, but also in South America (Zapata et al., 2008).

Species discrimination has often been cited as problematic, and palaeoecologists have often lumped the different forms into a couple of species or varieties (Charman et al., 2000). This lumping may well have prevented ecologists from distinguishing forms that occupy different niches. The distribution of "*N. collaris*" along a wetness gradient indeed showed a multimodal distribution (Valiranta et al., 2012). This suggested the existence of several taxa differing in their ecological optima and several distinct species were indeed described based on COI gene sequences and morphology (Kosakyan et al., 2013). Further evidence that genetically closely-related and morphologically similar forms may occupy different ecological niches and/or have contrasted geographic distributions was found in a broad scale study of the mixotrophic (morpho)species *Hyalosphenia papilio* (Heger et al., 2013). If such closely related forms are proved to differ in their ecological preferences and if they can be securely identified, then testate amoeba-based bioindication and palaeoecological tools could potentially be improved.

In order to improve the taxonomic framework for the *N.collaris* group, and allow sound ecological work and subsequent application in bioindication, it is essential to clarify the true diversity of testate amoebae using a combination of molecular and morphological approaches. We therefore describe *Nebela gimlii*, a new species of the *Nebela collaris* group from the Le Cachot peatland in the Swiss Jura Mountains.

Material and Methods

Sample collection and identification

Sphagnum sp. mosses were collected from the forested area (*Pinus mugo uncinata*) of "le Cachot" peatland, in the Swiss Jura Mountains (47°00'15.23" N, 6°39'52.83" E). Samples were visualised under light microscopy and contained, besides *N. gimlii*, specimens of *N. collaris*, another species of the complex which cannot be mistaken morphologically as cells are almost twice as long. Cells were isolated individually with a narrow diameter pipette under an inverted microscope and rinsed several times with tap water. Measurements of 14 cells were taken under an inverted microscope (Olympus IX81) at magnifications of 100× and of 400×. Photographs were taken at magnification of 400× (fig. 3.1). We measured the following morphometric traits on the test: length, breadth, depth, and aperture width as described in (Kosakyan et al., 2013) and calculated the width/length ratio. The biovolume was calculated according to (Charriere et al., 2006).

Scanning electron microscopy

Three *Nebela gimlii* tests out of the 14 analysed cells were mounted on stubs and then kept during one week in a desiccator. The tests were coated with gold in a vacuum coating unit and then observed either with a JEOL JSM-5510 microscope (Tokyo, Japan) at 10 kV or with a Philips XL30 FEG microscope (Amsterdam, The Netherlands) at 3 kV (fig. 3.2).

DNA amplification

DNA from two single cells was extracted using a guanidine thiocyanate-based protocol (Chomczynski and Sacchi, 1987), adapted after (Gomaa et al., 2013). The COI sequences were obtained by polymerase chain reaction (PCR) using the broad spectrum primer LCO (Folmer et al., 1994) and a *Nebela collaris*-complex specific primer and PCR conditions as in (Kosakyan et al., 2013). The amplicons were cloned into a PCR2.1 Topo TA cloning vector and transformed into *E. coli* TOP10'. One Shots cells (Invitrogen kit) according to the manufacturer's instructions. Two inserts per PCR product were amplified with M13F, M13R primers. Sequencing was carried out using a BigDye197 Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and analysed with a ABI- 3130XL DNA sequencer (Applied Biosystems). Sequences were deposited in GenBank with the following accession numbers: KP083297 and KP083298. Light microscopy pictures of the two extracted cells are shown in fig. 3.1.

Phylogenetic analysis

We build an exhaustive reference database containing 31 different sequences of the COI of the *Nebela collaris* species complex present in the GenBank database (Heger et al., 2011a; Kosakyan et al., 2012, 2013) plus three sequences of *Nebela tubulosa* and one sequence of *Certesella martiali* used as out- group. We aligned the sequences manually using the BioEdit programme (Hall, 1999). The alignment is available from the authors upon request. Programmes and parameters used to build the trees are the same as described in (Kosakyan et al., 2013) (fig. 3.3).

Results

Description of the species

Arcellinida Kent 1880

Hyalospheniidae (Schulze) Kosakyan and Lara.

Nebela gimlii n.sp. Singer and Lara.

The test is wide pyriform or drop-shaped, laterally compressed, with a protruding narrow neck. Two lateral pores are present ca. 1/4 from the distance from the pseudostome to the fundus of the test (fig. 3.1). A variable number of small pores can be seen on SEM images in apparently random positions from near the aperture to the fundus of the test (fig. 3.2). The test is colourless or slightly brownish, composed of small particles (likely

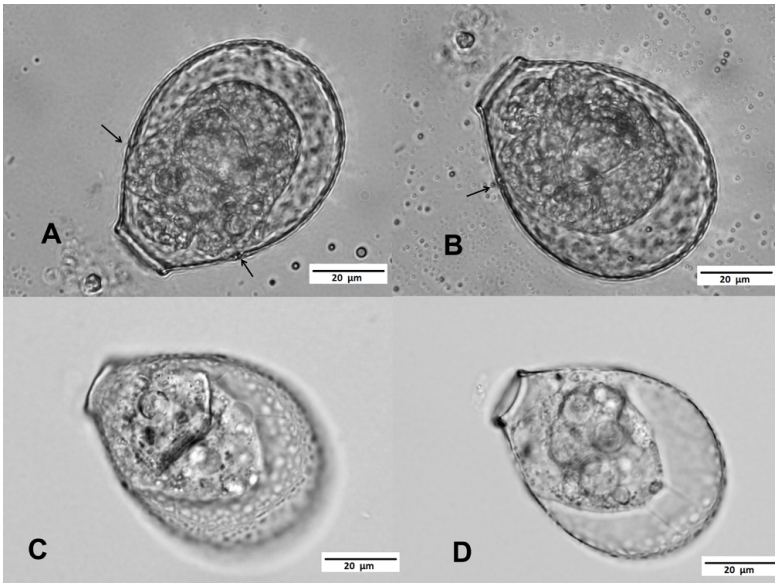

Figure 3.1 – (A–D) Light microscopy images of *Nebela gimlii* test. (A) test corresponding to Nebela_ gimlii_el3_1 (GenBank: KP083297) – note the presence of two lateral pores (arrows) ca. 1/4 from the distance from the pseudostome to the fundus of the test, magnification of 400× (B) test corresponding to Nebela_gimlii_el3_4 (GenBank: KP083298), magnification of 400×, (C) example of *Nebela gimlii* test showing the scales, magnification of 400×, (D) example of *Nebela gimlii* with focus on the aperture, magnification of 400×.

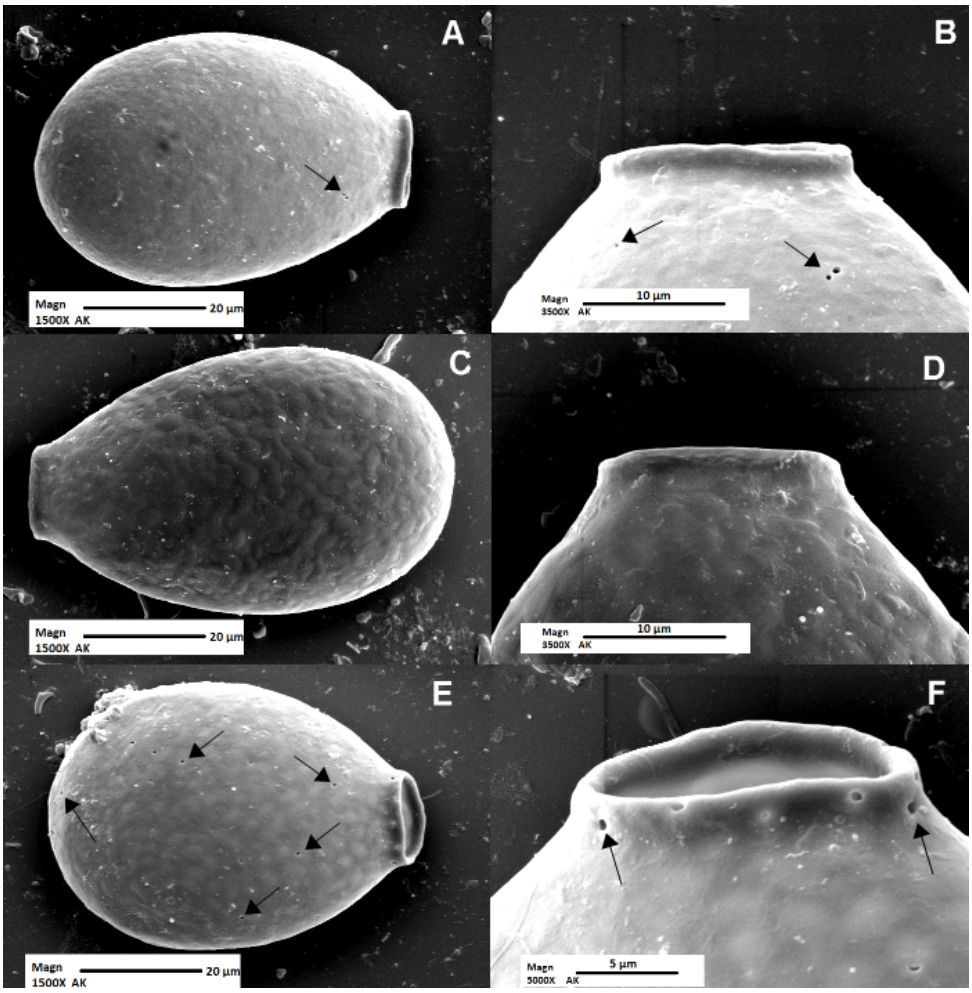

Figure 3.2 – (A–E) SEM micrographs of three individuals of *Nebela gimlii*. (A, C and E) – illustrating the shell general shape and composition; and (B, D and F) – detail view of the aperture of the same individuals. Arrows show some pores on the shell.

Table 3.1 – Biometrical characteristics of *Nebela gimlii*: M – median, SD – standard deviation, SE – standard error of the mean, CV – coefficient of variation, Min – minimum, Max – maximum, n – number of individuals examined (measurements in µm).

Characters	Mean	M	SD	SE	CV	Min	Max	n
Length	72.8	73.23	2.92	0.78	4	67.71	77.61	14
Breadth	53.9	53.12	3.67	0.98	6.81	49.68	61.64	14
Depth	31.4	30.7	1.4	0.37	4.45	30.5	35	14
Aperture width	19	18.71	1.03	0.28	5.45	17.5	21.06	14
Ratio (L/B)	1.4	1.35	0.07	0.02	5.45	1.22	1.5	14
Biovolume (104 µm^3)	8.25	7.82	1.08	0.28	13.12	6.94	10.47	14

obtained from preys, e.g. euglyphid testate amoebae), which are covered by a thin layer of organic cement. The aperture is oval in frontal view, slightly curved in lateral view, surrounded by a thin organic collar characteristic of family Hyalospheniidae (Kosakyan et al., 2012). Dimensions based on 14 individuals: length: 67.7–77.6 µm (mean = 72.8 µm), breadth: 49.7–61.64 µm (mean = 53.9 µm), width of aperture: 17.5–21 µm (mean = 19 µm). Illustrations and morphological data of the new species are given in fig. 3.1, fig. 3.2, table 3.1 and table 3.2.

Ecology

Nebela gimlii was found exclusively in the relatively dry *Sphagnum* mosses in the forested part of a peat bog (influenced by lateral drainage) and the drained peatland margin a few meters from the base of old peat extraction walls. These habitats are characterised by low pH, moderate moisture and low nutrient content.

Table 3.2 – Comparison of morphological characters between *Nebela gimlii* and *N. guttata*. Min – minimum, Max – maximum, n – number of individuals examined (measurements in µm).

Characters	*Nebela gimlii* (Min/Max) n=14	*Nebela guttata* (Min/Max) n=4
Length	68/78	80/89
Breadth	50/62	53/63
Aperture width	17.5/21	20/21
Ratio (L/B)	1.2/1.4	1.4/1.5

Hapantotype

The tests were collected from a *Sphagnum* sample in a peatland in Le Cachot, Jura Mountains, Switzerland (47°00'15.23"N, 6°39'52.83"E). One SEM stub with several specimens is deposited at the Natural History Museum of Neuchâtel (Ref. Nr.: UniNe-EM-5). COI gene sequences of *Nebela gimlii* el3_1 (499 bp) and *Nebela gimlii* el3_4 (499 bp) were deposited in GenBank (Nebela_gimlii_el3_1 KP083297 and, respectively, Nebela_gimlii_el3_4 KP083298).

Etymology

The name of this species refers to the name of Gimli, one of the dwarfs in J.R.R. Tolkien's masterpiece "The Lord of the Rings", because of its small size (the smallest known member of the *Nebela collaris* complex) and stout shape. In addition, it has been found abundantly in a forest, and Gimli was unique among his kind to have been travelling in the woods.

Note

Nebela gimlii can be confused with *Nebela guttata*, but clearly differs in the wider shape of the shell and the smaller dimensions of the test table 3.2. Both species are closely related and differ by 3.7% on the considered COI fragment. The two sequences of *N. gimlii* obtained from two different cells share 99.8% of similarity.

Discussion

Phylogeny of the *Nebela collaris* complex and position of *Nebela gimlii*

Our phylogenetic analysis places *N. gimlii* unsurprisingly within the *N. collaris* complex, as its morphology would predict. Within this group, it branches in a clade together with *N. guttata* and *N. tincta*, and (although less robustly) with the non-sphagnicolous *N. aliciae*. It shares a curved aperture with *N. guttata*.

N. gimlii resembles strongly *N. guttata*, but both forms do not overlap in size and have a slightly different shape. In addition, if we consider a barcoding gap of around 4% that has been suggested for hyalospheniids (Kosakyan et al., 2012, 2013), and also vannellid naked amoebae (Nassonova et al., 2010), *N. gimlii* can be considered as a new species.

Testate amoeba taxonomy has always been based on the observation of morphology. However, the degree to which arcellinids are phenotypically plastic is debated and has caused much confusion in taxonomy. Most Arcellinids do not secrete the mineral parts of their tests and therefore lack of the geometrically well-defined ornamented plates present in euglyphids. Several species have been shown to fall within a continuum of shapes (Lahr et al., 2008). This taxonomic confusion allows endless debates between "lumpers" and "splitters" and undermines the interpretation of species biogeography (Heger et al., 2011a). DNA barcoding is thus an invaluable tool that can be used to assess the true diversity within species complexes and more generally the taxonomic significance of seemingly minor morphological differences. A possible distinctive ecological niche would corroborate the specific status of these forms, as a triple species concept would then be applied: molecular, morphological and ecological. In the case of *N. gimlii*, it can be suggested that its optimum is located in the driest parts of the peatland. Indeed, we found it only in the forested part of the bog and the drained peatland margins where the water table was low, while it was apparently absent in the wetter parts. The presence of this small species in these comparatively dry habitats is consistent with the observation that smaller sized testate amoebae are favoured in drier environments (Jassey et al., 2011). However more work is clearly required to clarify the ecological optima of individual species within the *N. collaris* complex.

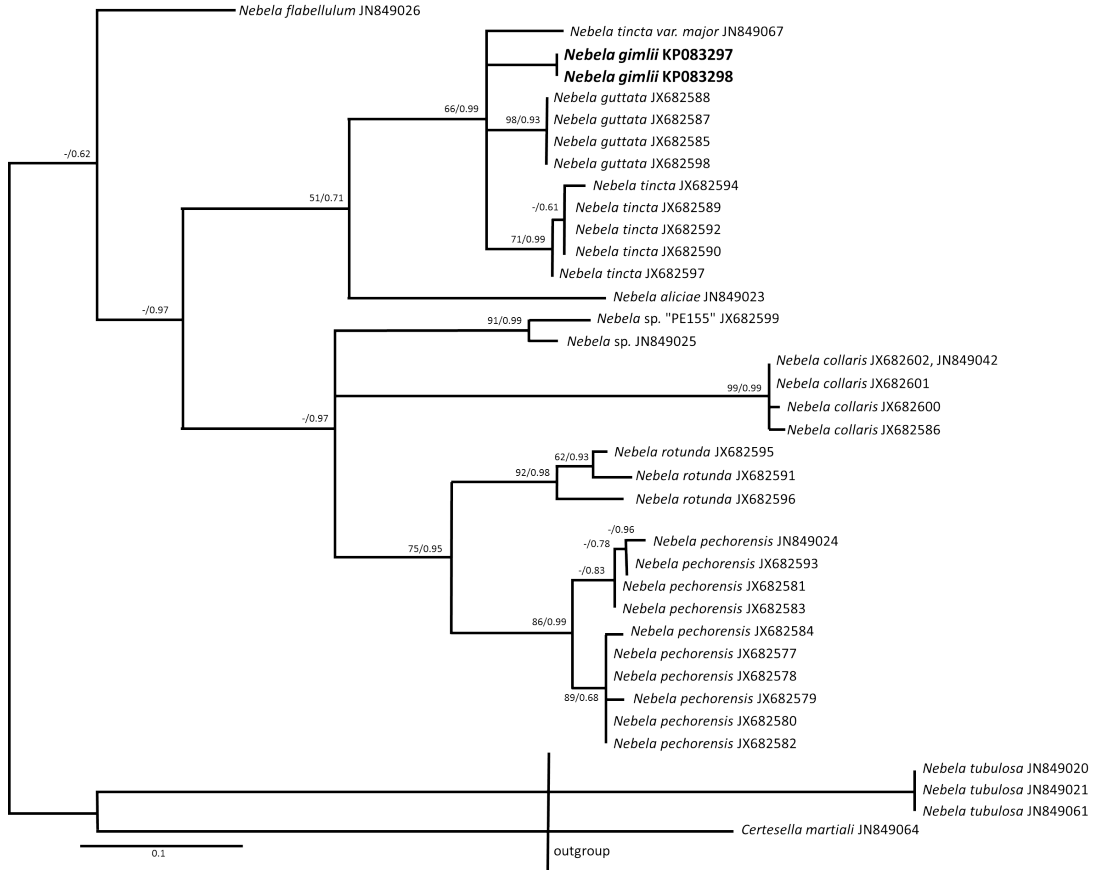

Figure 3.3 – Bayesian Markov Chain Monte Carlo consensus tree of 31 *Nebela collaris* s.I. testate amoeba. Based on mitochondrial cytochrome oxidase, subunit I (COI) gene. *Nebela gimlii* is denoted in bold. The numbers along the branches represent, respectively, the bootstraps obtained by maximum likelihood method and the posterior probabilities as calculated with Bayesian analyses. Only values above 50/0.50 are shown. The tree was rooted with outgroup *Certesella martiali* and *Nebela tubulosa*.

Diversity of the *N. collaris* complex

Generally, peatland ecologists usually consider three or four members of the *Nebela collaris* complex: a small species (*N. tincta*), one of two larger species (*N. collaris* and/or *N. tincta* var. *major*) and the conspicuous wider than long *N. flabellulum* (Charman, 2001). With this study, we raise to eight the number of barcoded species. But the picture is most likely still incomplete, as other sequences in GenBank derived from organisms that have not been documented morphologically yet (JN849067, JN849025, JX682599 GenBank entries) do not fit within these eight species. In addition, most sequences are originated from Europe, and it is likely that some species at least have a restricted geographical distribution, as shown in *Hyalosphenia papilio* (Heger et al., 2013). The case of the *Nebela collaris* complex therefore illustrates well the current under-estimation of protist diversity. The fact that so many new species can be found in well-studied ecosystems within relatively large and conspicuous groups of protists is in line with the idea that protists dominate eukaryotic taxonomic diversity (Pawlowski et al., 2012).

Acknowledgements

This work was funded by Swiss National Science Foundation projects no. 310003A 143960. SEM was made possible through the Swiss Centre of Electronic and Microtechnique (CSEM) in Neuchâtel. We thank two anonymous reviewers for commenting on the manuscript.

Chapter 4

Environmental filtering and phylogenetic clustering correlate with the distribution patterns of cryptic microeukaryotic species in peatlands

David Singer[1], Anush Kosakyan[1], Christophe V.W. Seppey[1], Amandine Pillonel[1], Leonardo D. Fernández[1,3], Diego Fontaneto[4], Edward A.D. Mitchell[1,2], Enrique Lara[1]

[1]Laboratory of Soil Biodiversity, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland
[2]Department of Zoology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, Brazil
[3]Laboratorio de Ecología Evolutiva y Filoinformática, Universidad de Concepción, Concepción, Chile
[4]National Research Council, Institute of Ecosystem Study, 28922 Verbania Pallanza, Italy
[5]Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

Molecular ecology, submitted

Abstract

Little is known about the diversity of protists and the factors that determine their diversity and community structure, as compared to other microbial groups, despite their ubiquity and known environmental roles. Environmental DNA studies are increasingly revealing the composition of natural protist communities in various environments. However, the mechanisms responsible for the assembly of these communities remain largely unknown, including the relative importance of resource partitioning, competition and environmental filtering and especially considering their phylogenetic context. To address these questions, we investigated the diversity of a complex of cryptic amoeba species in *Sphagnum*-dominated peatlands - the *Nebela collaris* group (Amoebozoa; Arcellininda) - across different microhabitats. We used an environmental DNA survey approach based on the mitochondrial cytochrome c oxidase 1 (COI) gene, allowing species-level resolution as well as phylogenetic reconstruction. Community composition varied strongly across microhabitats and associated environmental gradients (nitrogen and water table depth). Species showed little overlap in their realized niche, suggesting resource partitioning rather than competition, and a strong influence of environmental filtering in community composition. Furthermore, phylogenetic clustering (NTI analysis) could be observed in the most nitrogen-poor areas, supporting phylogenetic inheritance of adaptations in the group of *Nebela guttata*.

Keywords: Environmental filtering; Phylogenetic clustering/over-dispersion; Nearest Taxon Index (NTI); Niche partitioning; Protists; *Sphagnum*-dominated peatlands

Introduction

Molecular environmental diversity studies of microbial Eukaryotes have completely challenged our vision of Earth's biodiversity. High throughput sequencing (HTS) of aquatic and terrestrial habitats have revealed the existence of hundreds of thousands of phylotypes. A recent study shows that eukaryotic diversity in the sunlit part of the ocean is dominated by protists and exceeds that of Prokaryotes (De Vargas et al., 2015). Presently, a major challenge in protist ecology is to understand the rules that shape the community composition of these extremely diverse assemblages in natural systems, and notably if these differ from what is known in macroscopic organisms.

Historically, the coexistence of many protist species in a single homogeneous environment has been formulated by Hutchinson (Hutchinson, 1961) as the "paradox of the plankton". Several explanations have been given, most often based on chaotic oscillations of the different populations, which rely on fast population growth and high mortality rates (Scheffer et al., 2003). However, not all protists are fast growers; for example in soil testate amoebae, only 19 generations per year have been reported for *Phryganella acropodia* in the River Saale (Germany) (Schönborn, 1992). Although fluctuations in testate amoebae populations can be observed in temporal studies, these remain limited and do not reach an order of magnitude (Gilbert et al., 2003). These dynamics rule out the impact of chaotic effects in shaping testate amoeba community composition, and suggest rather that communities reach rapidly a state of equilibrium. Indeed, a comparative study conducted at 40 years interval in a *Sphagnum* peatland showed a remarkable stability when environmental conditions did not change significantly (Buttler et al., 1996; Lamentowicz et al., 2010). The composition of such communities can be rather seen as the classical product of deterministic forces due to both biotic (organisms interactions) and abiotic (environmental filters) processes, plus random drift. The relative importance of these drivers remains nevertheless to be determined in communities of slow growing protists. In this respect, niche breadths of protists as well as the characteristics (stability and selectivity) of the ecosystem under consideration can be expected to play an important role (Weiner et al., 2012). Ecological tolerance, which can be considered as a good proxy for niche breadth (Futuyma and Moreno, 1988) varies considerably amongst microbial eukaryotes, including with a single genus. For instance, species from genus *Dunaliella* are famous for thriving in different extreme environments (hypersaline, hyperacidic, high UV radiation (Pick, 1998), as well as under more mesophilic conditions (Pick, 1998). It has been shown in this particular group that adaptation to different environmental stresses was phylogenetically inherited (Gonzalez et al., 2001).

Sphagnum peatlands are characterised by stable but harsh conditions (oligotrophic, low pH, recalcitrant organic matter) (Dedysh et al., 2006; Steinberg et al., 2006), causing a strong selective pressure resulting in the existence of highly specific biotic communities. They are also characterised by high spatial heterogeneity, with chemical and physical conditions varying strongly in relation to microtopography, thus influencing vegetation (Okland et al., 2008). Members of the testate amoeba genus *Nebela* are especially common and diverse, and the *Nebela collaris* species complex is the most frequent Arcellinid testate amoeba taxon in *Sphagnum* dominated peatlands (Gilbert et al., 2006). This complex includes several "genetic" species (Singer et al., 2015) but, as the taxonomy of this group is confusing, ecologists and palaeoecologists usually handle them as a single "type" referred to as *Nebela tincta* (Charman et al., 2000). However, multimodal patterns of distribution observed along a moisture gradient (water table depth) in *Sphagnum* dominated peatlands (Valiranta et al., 2012) suggests that

this group includes species with diverging ecological optima. If this were the case, the question then is how far do these organisms truly co-exist in syntopy, and if so, what are the likely mechanisms involved that make this co-existence possible?

In order to answer these questions, we investigated species composition of members of the *Nebela collaris* s.l. in the whole range of microhabitats existing in peatlands of the Swiss Jura Mountains. In order to avoid identification mistakes which may occur with these closely-resembling species, we used a metabarcoding strategy based on the specific sequencing of *N. collaris* s.l. mitochondrial COI gene from environmental DNA. For each sampling point, we also measured relevant environmental variables (water table depth, nitrogen and phosphorus content, pH, brightness and microtopography). We determined how far community composition was correlated to the environmental variables characterising the different microhabitats, and if strong niche overlap occurred between species. A significant niche overlap would suggest interspecific competition rather than facilitation, given the close genetic relationship between species and trait similarity (Beltran et al., 2012). Finally, we determined taxon relatedness within the different microhabitats in order to detect if adaptations to a given environment were phylogenetically inherited or not. Phylogenetic inheritance would result in communities of closely related taxa (phylogenetic clustering) and that are organised by environmental filters; whereas lack of phylogenetic inheritance would result in communities of distantly related taxa (phylogenetic overdispersion) and that are organised by intraspecific competition (Emerson and Gillespie, 2008; Vamosi, 2014).

Peatlands are stable and highly selective environments which exhibit strong spatial variation due to microtopography, and subsequent drastic changes in abiotic conditions. As a consequence, plant communities also present high spatial structuration (Andersen et al., 2011). Accordingly, we expect *Nebela* species to be distributed in accordance with microtopography. In addition, given the low nutrient levels (and especially Nitrogen) in *Sphagnum* peatlands, we expect that Nitrogen deficiency could represent a strong environmental filter that will influence the composition of communities. Secondly, if *Nebela* species is associated to microtopography as we suppose, niche overlap should be minimal, suggesting low competition. In addition, as it has been observed in other protist groups, we expect that ecological preferences should be inherited in closely-related *Nebela* lineages.

Methods

Sampling and environmental variables

The study was carried out in two Swiss peatlands from the Jura Mountains: le Cachot (47°00'N 6°39'E) and Praz-Rodet (46°33'N 06°10'E). The peatlands were selected according to their similar characteristics of altitude (approximately 1000m a.s.l.) and geomorphological setting (Lotze and Hölder, 1964). We collected 21 samples containing each 30g of fresh *Sphagnum* sp. (le Cachot = 11, Praz-Rodet = 10) during the period spanning from September 2013 to April 2014 in five contrasted microhabitats present in the peatland (poor fen, bog margin, bog forest, hummock and lawn; Supp. table 1). We measured water table depth (WTD), brightness and microtopography as described in (Singer et al., 2016) at the time of sampling. We then measured total Nitrogen (%), total Phosphorus (%) and pH H_2O following standard protocols (Carter and Gregorich, 2007) (Supp. table 2). The five microhabitats had different vegetation (Singer et al., 2016) and environmental characteristics (Supp. fig 1). Bog margins (former peat exploitation area currently a contact zone with an adjacent hay field) were only found in le Cachot peatland.

Molecular analyses

Testate amoeba cells were extracted from 20g of fresh *Sphagnum* and concentrated by sieving at 150µm. We extracted DNA from the <150µm. using a MoBio Power Soil® DNA Isolation kit according to the manufacturer's instruction. We amplified a fragment of the mitochondrial COI gene by using the general primer LCO (Folmer et al., 1994) and a specific primer TINCOX (CCATTCKATAHCCHGGAAATTTC) following the protocol recommended for the amplification of *Nebela collaris* s.l. species (Kosakyan et al., 2013). PCR steps consisted of a 5 min initial denaturation step in a 40 cycles program of 15 s at 95°C, 15 s at 43°C, and 1 min and 30 s at 72°C with the final extension at 72°C for 10 min. The amplicons were cloned into pCR2.1 Topo TA cloning vectors and transformed into E. coli TOP10' One Shots cells (Invitrogen kit) according to the manufacturer's instructions. We amplified (M13f, M13R primers (Kosakyan et al., 2015)) and sequenced enough inserts per PCR product to reach the total diversity of *Nebela collaris* s.l. In order to evaluate if the total diversity was reached we performed rarefaction curves with visual check of asymptotic behaviour for each site independently and for the overall data set (sup Fig 2, 3) and we computed Chao estimates for the overall data set using the R vegan package (V.2.3-5 (Oksanen et al., 2015)), Sequencing was carried out using a BigDye197 Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and analysed with an ABI-3130XL DNA sequencer (Applied Biosystems).

Phylogenetic reconstructions

Phylogenetic analysis was used to determine the phylogenetic position of the sequences obtained. We built a reference database containing 32 COI sequences (obtained from single cells) of the *Nebela collaris* species complex from GenBank (Heger et al., 2011a; Kosakyan et al., 2012, 2013; Singer et al., 2015) plus three sequences of *Nebela tubulosa* (GenBank accession number: JN849020, JN849021, JN849061) used as outgroup. Sequences were aligned using the Clustal W algorithm (Thompson et al., 1994) (Supp. PDF 1). COI gene sequences of the *Nebela collaris* species complex contained some post-transcriptional mitochondrial editing resulting on the insertion of single nucleotides in the gene sequence (Kosakyan et al., 2013), these single nucleotides were further removed for all subsequent analyses. Trees were reconstructed using Bayesian and maximum likelihood (ML) approaches as in (Kosakyan et al., 2013; Singer et al., 2015). A Bayesian Markov Chain Monte Carlo (MCMC) analysis was performed using MrBayes v3.2 (Ronquist et al., 2012) with a general time reversible model of sequence evolution. A ML tree was build using the RAxML v7.7.1 algorithm (Stamatakis et al., 2008) using the GTR+G+I model. Model parameters were directly estimated in RAxML.

Phylotype identification

To evaluate the number of independent evolving units within *Nebela collaris* s.l., we used the Automatic Barcode Gap Discovery (ABGD) method (Puillandre et al., 2012). ABGD is an automated procedure that clusters sequences into candidate species based on pairwise distances by detecting differences between intra- and inter-specific variation without any a priori species hypothesis (Puillandre et al., 2012). We used the web-server of ABGD http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html with the Jukes-Cantor (JC69) model and the default parameters to run this analysis.

Niche partitioning and competition

In order to convert clone numbers to relative proportions of individuals from each species, we calculated the "corrected clone number" (Cx_corr) of each phylotype as described in (Kosakyan et al., 2015). Briefly, (Kosakyan et al., 2015) have demonstrated that there is a significant correlation between the Biovolume of a selected species and the number of clones obtained for the species within the community. Cx_corr is obtained by dividing the biovolume of the largest species (*Nebela collaris*) by the biovolume of the given species multiplied by the number of clones found in the community. For the phylotypes that could not be assigned to a described species, and thus could not be associated to a known biovolume, the value of the largest species (*Nebela collaris*) was selected (Supp. table 3) in order to avoid an over-representation of the phylotype within the community.

We assessed if species assemblages (relative proportions of each species for each sample) of the *Nebela collaris* group changed significantly among microhabitats using a redundancy analysis (RDA) with water table depth, pH, nitrogen content used as explanatory variables. These variables correspond to the main ecological gradients in *Sphagnum*-dominated peatlands (Bridgham et al., 1996, 2001).

Direct experimental evidence cannot be provided to prove if competition takes place or not given that *Nebela* spp. do not grow in culture; thus, we had to assess only indirect evidence to evaluate if competition may take place or not. Under this scope, we evaluated this possibility by computing (1) C-scores to evaluate if niche breadth is significantly different between species and (2) Pianka index, to estimate whether niches overlap.

Species co-existence can be impossible either because they occupy different niches (reduced fundamental – or grinnellian - niche overlap) or because they replace each other due to strong competition (high potential niche overlap but distinct realised – eltonian - niches) (Elton, 1927; Grinnell, 1917). We tested the null hypothesis that phylogenetically close species of the *Nebela collaris* group have the same niche breadth and thus, exhibit strong interspecific competition (i.e. they exhibit a strong niche overlap and cannot co-occur) using the C-score index (Stone and Roberts, 1990). This method requires a species x site matrix to calculate the species occurrence at a given site in the absence of another one. If a high number of species do not co-occur, then a checkerboard pattern arises in the matrix. The occurrence of this pattern is interpreted as an indication of exclusive competition between species (Diamond and Cody, 1975; Gotelli and Graves, 1996; Stone and Roberts, 1990). The statistical significance of this analysis was evaluated with a row-column null model (50,000 random matrices, $P < 0.05$), a conservative algorithm that minimises Type I errors (Gotelli, 2000). C-score values higher than expected supports the null hypothesis, while values lower than expected reject the null hypothesis. C-score and null model calculations were computed using the R package EcoSimR (v.0.1.0) (Gotelli and Ellison, 2013).

We calculated Pianka's niche overlap index (Pianka, 1973) in order to evaluate if species could be expected to compete with each other using another line of evidence. This index requires an incidence matrix with species as rows and niche categories as columns to analyse how much overlap do exist in the use of a given resource axis among species. Pianka's index ranges from 0 (no resource used in common between two given species) to 1 (complete overlap in resource use). The statistical significance of this analysis was evaluated

with a row-column null model (50,000 random matrices, $P < 0.05$) a conservative algorithm that minimizes Type I errors (Ulrich and Gotelli, 2010). Values lower than expected (below the 95% confident intervals of the null model expectations) support the occurrence of low niche overlap between taxa, while values higher than expected (above the 95% confident intervals of the null model expectations) reject this hypothesis. Pianka's index and null model calculations were conducted using EcoSimR (v.0.1.0) (Gotelli and Ellison, 2013).

We calculated Nearest Taxonomic Index (NTI) to evaluate the phylogenetic structure of species assemblages in each sample and compare the obtained values across microhabitats. The NTI index is more sensitive to patterns of evenness and clustering closer to the tips of the phylogeny than other index such as the Net Relatedness Index (NRI) (Kembel et al., 2010), and the use of clearly identified fine-scale taxonomic units such as phylotypes is known to improve the reliability of the method (Koeppel and Wu, 2014). Studies based on microbial communities suggested that a phylogenetic clustering (NTI>0) may indicate habitat filtering (Horner-Devine and Bohannan, 2006; Pontarp et al., 2012). To the opposite, phylogenetic overdispersion (NTI<0) was also detected in communities and potentially connected to interspecific competition (Koeppel and Wu, 2014). The NTI was calculated using Picante (V. 1.6-2) (Kembel et al., 2010) within R. Analysis was run with sequences abundance information and the parameters used for the phylogeny reconstruction were the same as described above. The Phylogeny was pruned in keeping the sequence of the more abundant haplotype of each phylotypes defines with the ABGD analysis. In order to understand the role of the environmental correlates on NTI and NRI, we build linear models with NTI or NRI values for each site as the response variable and a set of environmental measurements as explanatory variables. Most of the explanatory variables expressed habitat features (i.e. N, P, N/P, pH, Water table depth, conductivity), but we also included the potential confounding effect on NTI of observed species richness, measured as number of phylotypes. Some variables (i.e. water table depth and conductivity) were log transformed to improve model fit.

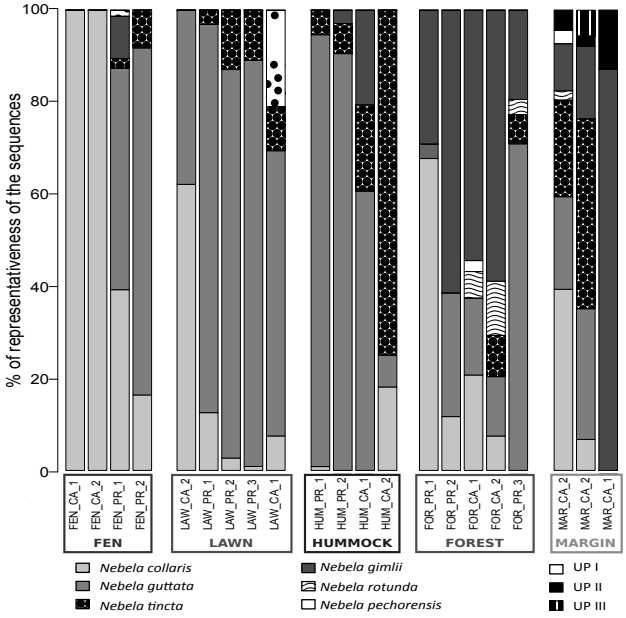

Figure 4.1 – (Communities profiles of testate amoebae from the *Nebela collaris* complex s.l. extracted from *Sphagnum* samples from two peatlands in the Swiss Jura Mountains. Each barplot represents the relative ratio of the clone sequences corrected by the cell biovolume (Kosakyan et al., 2015). UP = Unknown Phylotype.

We started by trying to minimise the effect of multicollinearity by checking for correlation between the explanatory variables: for each pair of continuous variables with Pearson correlation parameter higher than 0.75 we removed the one with the highest correlation values with the other variables. The only variables with > 0.75 correlation parameter were N and P (Supp. Table 4), and we retained N for the subsequent analyses.

After this preliminary step, we run a full model for NTI and a full model for NRI including all the remaining variables. We directly assessed model fit by: plotting the distribution of the model residuals to visually check if their distribution was approximately Gaussian; plotting the predicted versus fitted residuals to check for the Gauss-Markov assumptions; plotting the normal Q-Q plot to compare theoretical and observed quantiles; checking whether Cook's distances were all < 0.5 (Crawley, 2012). Moreover, in order to assess the relative importance of each predictor in the statistical models, we used a model averaging approach (Burnham and Anderson, 2002): the set of sub-models including all possible combinations of the explanatory variables was generated and the Relative Importance (RI) value of each variable was calculated, on a scale from 0 to 1, as the sum of the Akaike weights of the sub-models in which the variable appears; better models have larger Akaike weights, and a variable that contributes more to model fit will thus have a higher relative-importance value, closer to 1. Model averaging was performed with the R MuMIn package version 1.10.5 (Barton et al., 2013).

Results

The cumulative curves for the overall data set and for the single sites reached asymptotic values (sup Fig. 2,3), and the observed species richness (n=9) is within the confidence interval of the Chao estimates (9.9±2.1).

The ABGD analysis including our 1028 sequences (Supp. PDF 1) and the 32 references sequences available in GenBank for the *Nebela collaris* s.l. group identified a distinctive "barcode gap" centred around 7% of divergence

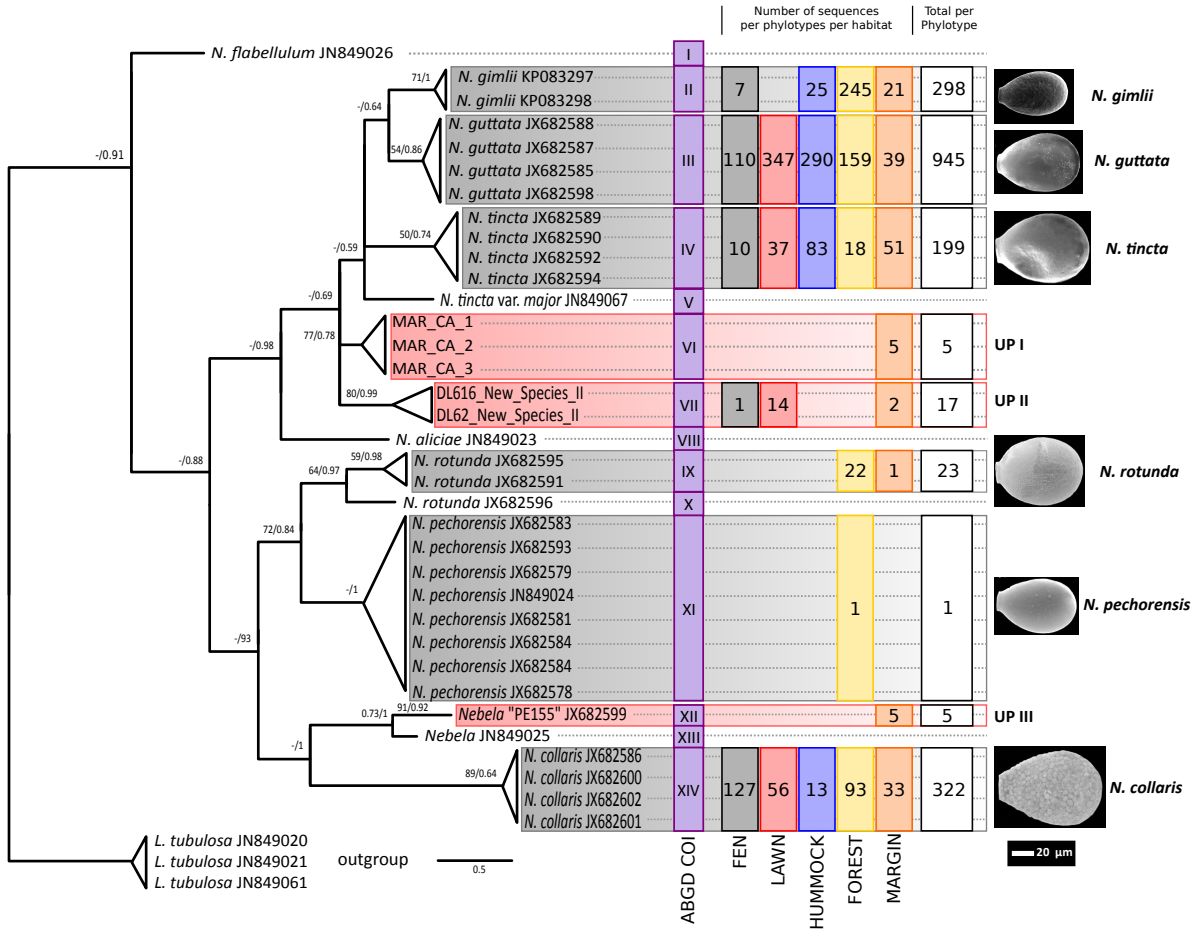

Figure 4.2 – (Bayesian Markov Chain Monte Carlo consensus tree of 32 single cells of *Nebela collaris* s.l. mitochondrial cytochrome oxidase, subunit I (COI) sequences and 1028 environmental sequences, plus 3 sequences of *Longinebela tubulosa* as outgroup. The numbers along the branches represent, respectively, bootstraps obtained by maximum likelihood method and posterior probabilities as calculated with Bayesian analyses. Only values above 50/0.50 are shown. Sample colours indicate the five different microhabitats (red = lawn, yellow = forest, blue = hummock, orange = margin, black = fen).

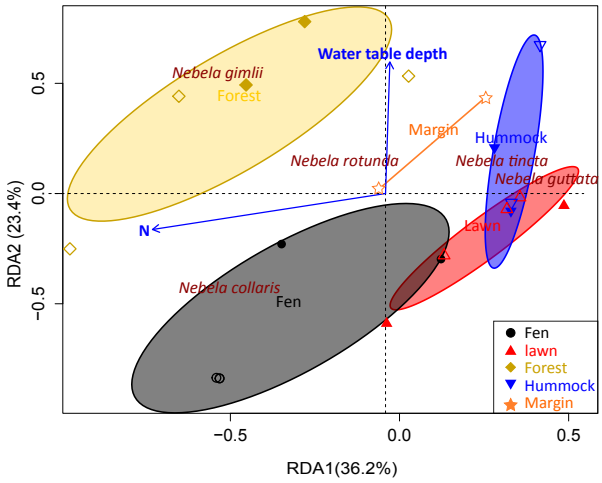

Figure 4.3 – (Redundancy analysis of *Nebela collaris* s.l. communities extracted from *Sphagnum* sample collected in Le Cachot bog (empty symbols) and Praz-Rodet bog (filled symbols). Environmental variables ($p < 0.05$) are represented by the arrows.

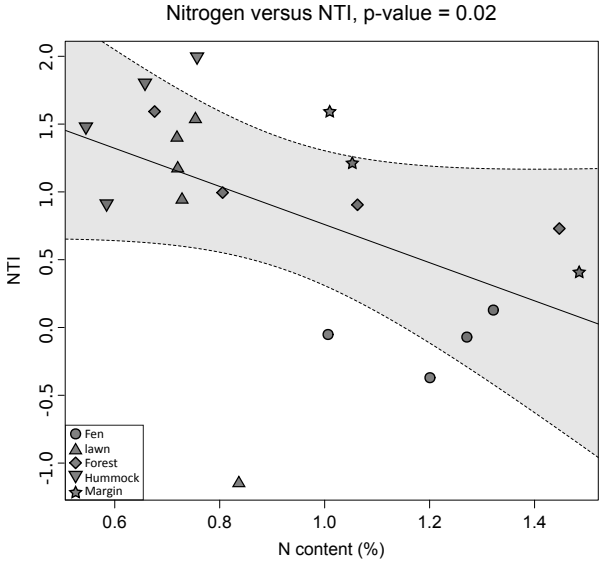

Figure 4.4 – (NTI values in function of Nitrogen content (%). Linear regression (p-value = 0.022), dotted lines correspond to the 95% confidence interval.

of the COI sequences and delimited 14 genetic clusters as candidate species (fig. 4.2). This result was consistent in all recursive partitions with prior intraspecific genetic divergence values between 0.28% and 0.77%, and we considered it more likely than the two other alternatives (defining 54 candidate species with intraspecific divergence values below 0.2% or considering only two species with an intraspecific divergence values higher than 1.29%).

According to the most conservative estimate in ABGD 14 phylotypes could be identified. Nine of the phylotypes were present in our samples, whereas four other were represented only by sequences from GenBank (fig. 4.2). Six of the phylotypes in our samples corresponded to known species (*N. collaris*, *N. guttata*, *N. tincta*, *N. rotunda*, *N. pechorensis* and *N. gimlii*) and the three other (referred to here as unknown phylotypes UP 1-3) represented probably undescribed species (fig. 4.1). UP3 corresponded to a sequence derived from a single cell documented in previous studies (Kosakyan et al., 2012, 2013) (fig. 4.2). The four more abundant phylotypes could be affiliated to the species *N. collaris, N. guttata, N. tincta* and *N. gimlii*. The less represented phylotypes, which never exceeded 20% of the total sequences within individual samples, were found mostly in the peatland margins (fig. 4.1). Numbers of COI sequences found for each phylotype are presented in the Supp. table 1. Obtained sequences were deposited in GenBank under the following accession numbers: (to be provided after acceptance of the manuscript).

The RDA analysis (fig. 4.3) of *Nebela collaris* s.l. communities in the different microhabitats from the two peatlands showed that species assemblages did not differ between the two peatlands ($p = 0.093$) and that the effect of sampling date was not significant ($p = 0.08$), but differed significantly among microhabitats ($p = 0.018$) with only limited overlap between lawns and hummocks in the first two axes of the ordination space (fig. 4.3). The first two axes together explained 59.6% of the total variance, with two variables, N (total nitrogen content) and WTD (water table depth) explaining respectively 24% and 10% ($p=0.003$ and 0.037, respectively).

The C-score index was significantly lower than expected (C-score Observed = 0.3611; expected = 5.5373, $\pm$ SD 0.078; $p = 0.01$), thus suggesting an absence of competitive exclusion between the different species of genus *Nebela* (i.e. they co-occur in the same site more often than expected). Furthermore, the Pianka's index was lower than expected by chance (Pianka's index; Observed = 0.369, expected = 0.773, $\pm$SD 0.00194, $p < 0.0001$), suggesting that overall *Nebela* species exhibit a low niche overlap and that therefore, they perform an effective partitioning along the resource axes measured (habitat breadth). Taken together, these two analyses suggest that current competition does not play an important role in shaping the assemblages of Nebela sp. species in the surveyed sites.

The regression coefficient estimated in the full linear model including (estimate, Standard error and p-value) together with the RI value obtain with the model averaging approach revealed N content as the most relevant environmental correlate for NTI (sup table 5) and for NRI (Supp. Table 6). The NTI values were closer to 0 (i.e. neither phylogenetic clustering nor over-dispersion) in microhabitats where the N content was high (e.g. fens), while NTI values increased where the N content decreased, showing phylogenetic clustering (fig. 4.4). Thus, when plotted against N content, NTI indices decreased linearly with increasing N content ($p = 0.0269$; fig. 4.4). Species composition of the most N-deprived environments (i.e. lawns and hummocks) indicated a shift in species assemblages towards communities dominated by *N. tincta and N. guttata* (figs. 4.1 to 4.3).

Discussion

Testate amoeba species are generally lumped into "morpho-types", in ecological and especially in palaeoecological studies (Charman et al., 2000; Mitchell et al., 2008). However, detailed DNA barcoding studies have revealed the existence of unsuspected cryptic diversity within numerous protist taxa, including the *Nebela collaris* group (Kosakyan et al., 2012, 2013; Singer et al., 2015). This raises the question whether this diversity corresponds to different ecological optima and/or functions, but so far this has not been explored in protists. This question is thus also of applied interest: if cryptic species do not differ in ecology or function then lumping them in morpho-taxa is a sound approach. If however they do differ in ecology and/or function then a lot of potentially useful information for biomonitoring or palaeoecological inference will be lost. Our environmental survey showed the existence of several new phylotypes (UP), in addition to the recently described taxa, thus suggesting that the total diversity in the *Nebela collaris* group has still not been uncovered. Most sequences from these unknown organisms were recovered from peatland margins, which represent clearly transitional environments between the peatland and, for instance, coniferous forest litter. Further investigations may therefore show that these unrelated new species have their ecological optima outside of peatlands. As the current sampling is strongly biased towards peatlands in Europe the true diversity of this species complex is likely much higher and more species are likely to be found in other habitats and/or regions of the World.

This study revealed that the different *Nebela* spp. were not randomly distributed in the studied peatlands, but were strongly correlated to microhabitats, thus following the microtopology present in the different microhabitat. Vascular plant and bryophyte communities also follow this trend, and are at least partly responsible for

changes in the local physicochemical conditions. Especially, *Sphagnum* spp. is well known to strongly influence key factors such as pH and ion content of their direct environment called bryosphere (Clymo, 1973; Lindo and Gonzalez, 2010; Van Breemen, 1995). Testate amoebae species distribution may be influenced by plants directly (e.g. through facilitation or allelopathy) or indirectly (e.g. through environmental modifications induced by the plants (Imparato et al., 2016; Philippot et al., 2013), but the possible existence of such mechanisms remains to be determined. Irrespective of plant community composition, the distribution of *Nebela* spp. was highly correlated with N content (figs. 4.3 and 4.4). Nutrient deficiency is a common trait in *Sphagnum*-dominated peatlands and especially in the hummock microhabitats (Bridgham et al., 1996). This lack of nutrients is known to shape prokaryotic communities, and peatlands have been cited as an example of extreme environment for this particular reason (Dedysh et al., 2006). It is, thus, likely that N deficiency acts as an environmental filter which influences species composition in *Nebela* communities. Other abiotic factors commonly measured in peatlands, like depth to water table, pH and P content were not correlated with community composition. A likely reason is that the range of variation in the microsites where *Nebela* species were found was too narrow to exert a detectable pressure on the communities. Altogether, these findings suggest that environmental filters do influence *Nebela* species distribution differently supporting the hypothesis that species differ in their respective ecological preferences.

Phylogenetic niche conservatism (Losos, 2008; Wiens et al., 2010) predicts that two closely-related species sharing the same niche should exclude each other. In our case, C-score values suggested a low likelihood that the presence or absence of one species in a given site was directly affected by the presence of other species. Likewise, the Pianka index suggested that, overall, species of *Nebela* exhibit low niche overlap, and that species perform an effective partitioning along the resource axis measured (habitat breadth). However, these assumptions are solely based on co-occurrence of species on the same sites and thus cannot be considered as strong evidence against species competition. Observations based on species traits can nevertheless corroborate this hypothesis. Indeed, *Nebela* species can be differentiated, amongst other features, based on the width of the shell aperture and shell size/biovolume (Kosakyan et al., 2013). These morphological traits have been correlated to prey size, larger species with broader apertures feeding on large preys (Jassey et al., 2014). In this study, *N. tincta* and *N. guttata* have been found very often in the same sample (fig. 4.1), and because *N. tincta* has a aperture on average 19% wider than *N. guttata* and a biovolume 44% larger, it can reasonably be expected that the two species have different food regimes (Kosakyan et al., 2013, 2015). These two species would therefore be expected not to compete, or at least not for preys, and thus be able to co-occur due to resource partitioning.

Also we demonstrate that phylogenetic clustering occurs in the most N-depleted areas of the peatland. This is illustrated by strongly positive NTI values in microhabitats where N concentrations were lowest (hummocks and lawns), meaning that organisms encountered in these microhabitats are phylogenetically more closely related than expected by chance (Webb et al., 2002). Typical species from N-poor environments are *N. guttata* and *N. tincta*, which are phylogenetically closely related (fig. 4.2). Therefore, it can be hypothesized that the most recent common ancestor of *N. tincta* and *N. guttata* underwent adaptations to exceptionally N-poor environments. These adaptations might be slow population growth and low rates of metabolism. By contrast, in microhabitats where N is less limiting (i.e. forest, fen and peatland margins), NTI values were close to 0, which means that communities tend to have a more heterogeneous phylogenetic composition. When NTI values are plotted against N content, this relationship appears even more clearly (fig. 4.4). We interpret this situation as a decrease in selective pressure in richer (or more mesophilic) environments, allowing higher numbers of species, not phylogenetically related, to co-occur.

Here, we show that the different species, as defined by COI (fig. 4.1) have contrasted distributions among microhabitats within *Sphagnum* peatlands (figs. 4.1 and 4.2), while each microhabitat type harboured very similar communities, independently of the studied peatland (fig. 4.3). Therefore, species distribution is not random across microhabitats and community patterns are indeed correlated to the main ecological gradients characterising *Sphagnum* peatland (fig. 4.3). These species also have different niche breadths and, where they co-occur, are more likely to be affected by resource partitioning than competition. Finally, our results suggest that niche conservatism takes place in genus *Nebela*, where only a few species can adapt to excessively low N contents. Altogether, we show that habitat selection and resource partitioning are effective in shaping *Nebela* communities, and might be possible drivers for diversification in the genus *Nebela* and other "slow-growing" protists in general.

Acknowledgements

This work was funded by Swiss National Science Foundation project no. 310003A 143960. We would like also thanks Thierry Heger for the help on the ABGD analysis and Quentin Blandenier for fruitful discussion on the manuscript. We also thank Sébastien Tschanz for giving us the official permission for sampling in the protected le Cachot peatland.

Chapter 5

High-throughput sequencing reveals diverse oomycete communities in oligotrophic peat bog micro-habitat

David Singer[1*], Enrique Lara[1*], Mónica M. Steciow[2], Christophe V.W. Seppey[1], Noelia Paredes[3], Amandine Pillonel[1], Tomasz Oszako[4], Lassâad Belbahri[1]

[1]Institute of Biology, Laboratory of Soil Biology, University of Neuchâtel, CH-2009 Neuchâtel, Switzerland
[2]Instituto de Botánica Spegazzini, Universidad Nacional de La Plata, 53 N°477, 1900 La Plata, Buenos Aires, Argentina
[3]CADIC, CONICET. Lab. Ecología Terrestre, Bernardo Houssay 200, Ushuaia, Tierra del Fuego, Argentina
[4]Forest Research Institute, Braci Lesnej St., Sekocin Stary, 05-090 Raszyn, Poland

[*]Both authors contributed equally to the work.

Fungal Ecology, Vol.23, p.42-47, October 2016
DOI: 10.1016/j.funeco.2016.05.009

Abstract

Oomycete diversity has been generally underestimated, despite their ecological and economic importance. Surveying unexplored natural ecosystems with up-to-date molecular diversity tools can reveal the existence of unsuspected organisms. Here, we have explored the molecular diversity of five microhabitats located in five different oligotrophic peat bogs in the Jura Mountains using a high-throughput sequencing approach (Illumina HiSeq 2500). We found a total of 34 different phylotypes distributed in all major oomycete clades, and comprising sequences affiliated to both well-known phylotypes and members of undescribed, basal clades. Parasitic species, including obligate forms were well-represented, and phylotypes related to highly damaging invasive pathogens (*Aphanomyces astaci*: X1100 and *Saprolegnia parasitica*: X1602) were retrieved. Microhabitats differed significantly in their community composition, and many phylotypes were strongly affiliated to free water habitats (pools). Our approach proved effective in screening oomycete diversity in the studied habitat, and could be applied systematically to other environments and other fungal and fungal-like groups.

Keywords: Oomycetes; Diversity; Metabarcoding; Illumina sequencing; V9 region of the SSU rRNA

Introduction

Oomycetes are a group of fungi-like stramenopiles which have a considerable impact on the economy through the pathogenic action of certain members. Indeed, they are responsible for major diseases in crops (Abad et al., 2014), forests (Duran et al., 2010; Rizzo et al., 2005) and aquaculture (Phillips et al., 2008). Some emergent pathogens have been recently detected and are currently causing considerable losses in both agriculture (Kamoun et al., 2015) and fish farms (de la Bastide et al., 2015). In addition, oomycetes have an important impact on food webs through regulation of host populations (Duffy et al., 2015). However, their environmental diversity has been rarely studied specifically in natural ecosystems (Willoughby, 1962, 1978), with the consequence that potential threats are detected only after disease outbreaks. There is, therefore, a need for environmental surveys, with a particular focus on understudied systems. Peat bogs and other acidic and oligotrophic environments have not been surveyed for oomycetes (Lara and Belbahri, 2011). Peatlands are heterogeneous ecosystems traditionally divided into five microhabitats (pools, lawns, hummocks, Fen/Transition and forest; (Rydin and Jeglum, 2013)). Indeed, the low nutrient amount present in these systems has been thought insufficient to support the growth of such osmotrophic organisms. However, an environmental survey of eukaryotic genetic diversity has revealed the presence of oomycetes in a pristine oligotrophic peatland located in the Swiss Jura Mountains (Lara et al., 2011). Sequences found in the course of that study appeared to be divergent and possibly represented novel clades at the genus level at least.

High-throughput sequencing approaches are increasingly applied to monitor environmental diversity of eukaryotic organisms (Amaral-Zettler et al., 2009; De Vargas et al., 2015; Lentendu et al., 2013). The immense amount of data generated allows drawing a more comprehensive picture of environmental diversity than earlier cloning and Sanger sequencing approaches (Lecroq et al., 2011). These approaches revealed novel deep and divergent lineages, as well as an unsuspected diversity within each established eukaryotic supergroup (De Vargas et al., 2015). Our knowledge of fungal diversity has also been challenged by the application of high-throughput sequencing approaches, revealing ancient groups that await description and characterization (Buée et al., 2009).

Oomycetes, although not belonging to Opisthokonta like true fungi (Ben Ali et al., 2002), share a lot of common traits with fungi. Most of the species described have been classified within two informal groups, the peronosporalean and the saprolegnialean 'galaxies' (Fuller and Jaworski, 1987). DNA based environmental surveys have shown that many more taxa, occupy basal positions within the oomycete tree (Lara et al., 2011; Steciow et al., 2013, 2014). The function of some of these organisms (obligate or facultative parasites; or strictly saprotrophic) has not been elucidated to date, and it is likely that they represent a novel diversity. By increasing the panel of surveyed environments, it is likely that novel organisms can be encountered. Peat bogs have, to date, never been investigated for oomycetes (Lara et al., 2011).

In this study, we applied a high-throughput sequencing approach (Illumina sequencing targeting the v9 region of the SSU rRNA gene) to characterize the environmental diversity of oomycetes in 65 samples from all microhabitats types across five peatlands located in the Jura Mountains (Switzerland and France). We surveyed their total diversity to evaluate how much is known of their environmental diversity, and also their presence and abundance in the different microhabitats to assess their preferred habitats.

Materials and methods

Studied areas and sampling strategy

The study was conducted in five *Sphagnum*-dominated peatlands located in the Jura Mountains (Switzerland and France). These sites are located at similar altitudes (900-1000 m a.s.l.), climatic conditions and geology. Five microhabitats were selected (fig. 5.1) with contrasting vegetation and microtopography and further characterized by chemical variables (nutrient content and pH). The vegetation composition was recorded at each sampling site in 1 m² plots. The central parts of the peatlands were ombrotrophic (i.e. rainwater fed) and open (i.e. no trees). Three of the microhabitats were located in the centre of the peatlands: hummocks (fig. 5.1A) are characterized by a mound microtopography dominated by *Sphagnum fuscum*, low nutrient content and pH. Lawns (fig. 5.1B), are flat habitats characterized by *Sphagnum magellanicum* and *Eriophorum vaginatum*. Pools (fig. 5.1C), are depressions usually with standing water and dominated by *Sphagnum cuspidatum*. These three microhabitats often lie close to each-other (fig. 5.1F) (Marcisz et al., 2014a; Pouliot et al., 2011). In addition, we studied two other habitats in the periphery and margin of the peatlands: i) tall pine forests (fig. 5.1D) grown on thick peat but with lower water table levels and dominated by pine (*Pinus mugo* subsp. *uncinata*) and *Vaccinium* spp.; ii) poor fens in some cases with scattered birch trees, representing the buffer zone between the peatlands and the surrounding environments (in most cases hay meadows) (fig. 5.1E). Nutrient content and pH are typically lower in the bog centre and pine forest than in the surrounding fen, but also decline from pool to hummock (Batzer and Baldwin, 2012) and this was indeed the case in our study table 5.1.

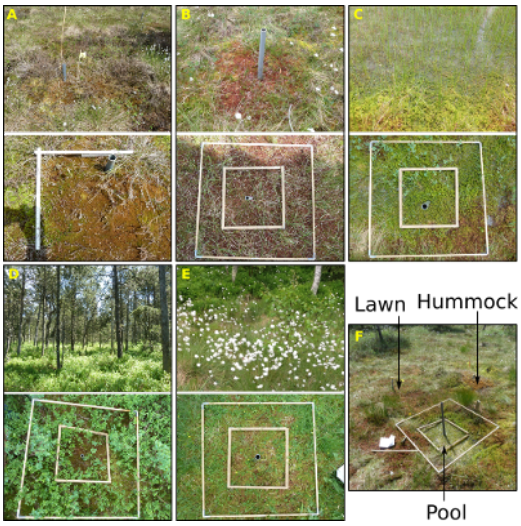

Figure 5.1 – The different peatland microhabitats. Each letter (except F) presents a general view of the habitat on the top, plus one detailed (small square=50 cm and large square=1 m) at the bottom. A: hummock, B: lawn, C: pool, D: forest, E: Fen/Transition and F: a general view of the center of the peatland with lawn, hummock and pool microhabitats. A to F from Edward A.D. Mitchell.

A total of 65 samples of *Sphagnum* spp. were collected using sterile equipment. To assess the intra-peat bog variation, five samples from each microhabitat were taken in 'Le Cachot' peat bog (coordinate: 47°0'19.99"N, 6°39'53.13"E). To test the variability among peatlands two samples of each microhabitat were taken in four other peatlands ('Praz-Rodet' bog 46°33'54.37"N, 6°10'20.37"E, 'Pontins' bog 47°7'38.41"N,6°59' 20.63"E, 'Etang de la Gruyere' peat bog 47°14'22.46"N, 7°2'58. 13"E and 'Frasne' bog 46°49'51. 45"N, 6°9'34.58"E).

PCR and sequencing

Environmental DNA was extracted using a MoBio Power Soil™ DNA isolation kit (Carlsbad, CA USA) following the manufacturer's instructions. PCR protocols targeting the SSU rRNA gene V9 region (amplicon size about 180bp) of eukaryotes was done according to (Amaral-Zettler et al., 2009). Sequencing was performed with an Illumina's HiSeq 2500, using V3 chemistry (Fasteris, Geneva, Switzerland).

Sequences processing

Quality check (Phred score filtering, elimination of reads without perfect forward and reverse primers, and chimera removal) of the sequences was performed following the pipeline developed by (De Vargas et al., 2015). Sequences were grouped in phylotypes using the clustering algorithm swarm (Mahé et al., 2014). The curated and updated PR² database (Guillou et al., 2013) was used to taxonomically assign the phylotypes after a selection of the V9 regions according to the primers used for the sequencing. Phylotypes were then assigned by aligning the dominant sequence of the phylotype to the database using GGSearch script in the FASTA package (Pearson, 2014). GGSearch is a script based on the global alignment algorithm (Needleman and Wunsch, 1970). Phylotypes of oomycetes were extracted from the total response matrix. Phylotypes present in more than 10% of total samples and with more than 10 sequences were kept for further analysis. The affiliation of each phylotype was determined by comparison to the GenBank database using BLAST; quick neighbour joining tree analyses were performed to place phylogenetically ambiguous taxa.

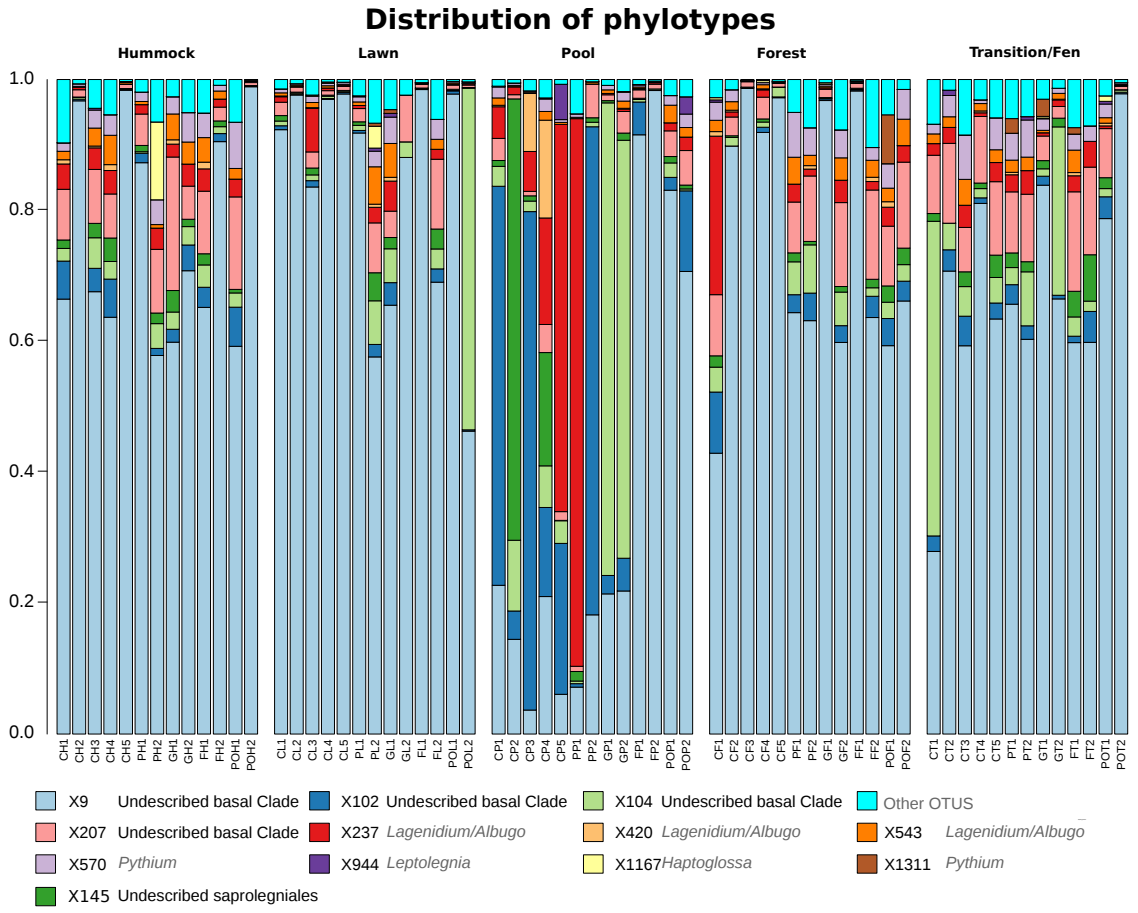

Figure 5.2 – Distribution of oomycete phylotypes (% of total) in the different microhabitats of the studied peatlands.

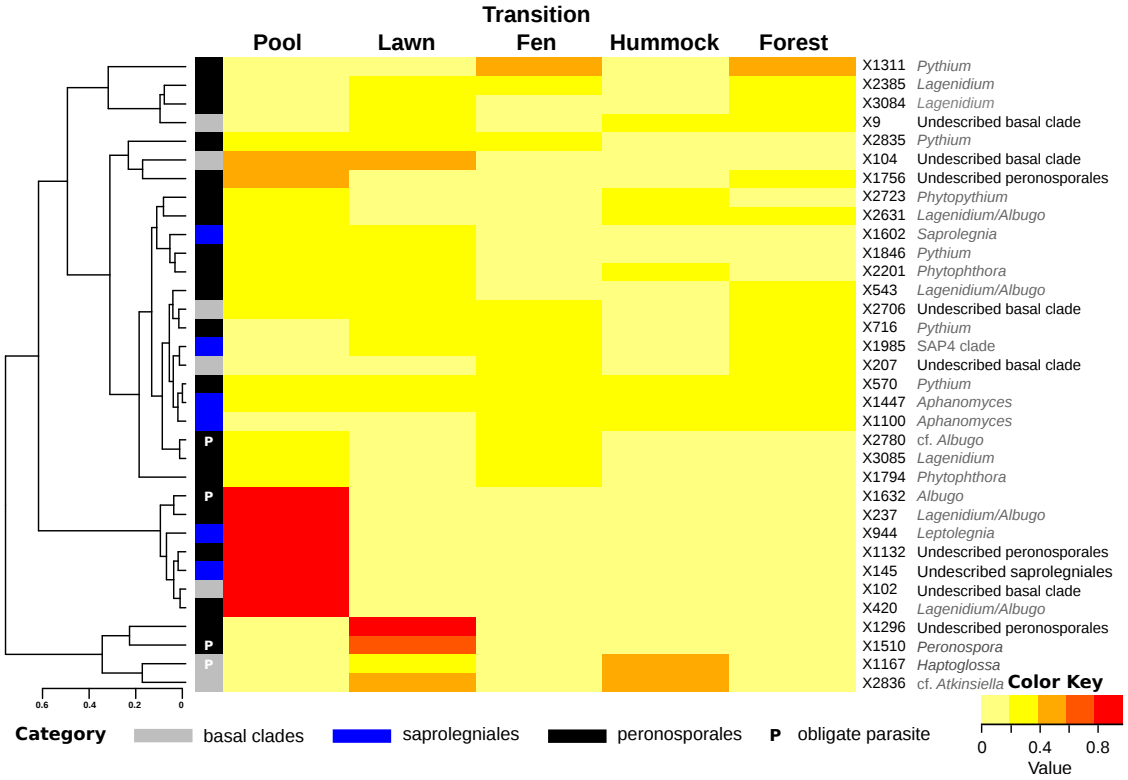

Figure 5.3 – Heat map depicting oomycete diversity and relative abundance in the microhabitats of the studied peatlands. The dendrogram was built using a complete agglomerative method on a Bray-Curtis distance matrix.

Data analysis

To determine if there is a significant effect of the microhabitat type or the sampling site on the oomycete community composition, we performed a permutation test using the 'envift' function on a NMDS of the R programs (Dixon, 2003). Hierarchical clustering analysis was made with a Bray-Curtis distance matrix of a Hellinger transformation matrix of the oomycete community composition. A heatmap was computed using the 'heat-map.2' function of the gplots package (Warnes et al., 2015).

Table 5.1 – Principal physico-chemical parameters and most characteristic bryophytes and plants of each micohabitat.

	N [%] (min/max/ average)	pH (min/max/ average)	water table depth [cm] (min/max/ average)	Tree cover [%]	Micro-topography (0=flat, 2=mound)
Hummock	0.3/0.7/0.5	3.8/4.1/4	26/73/42	0	2
Lawn	0.5/1.2/0.8	3.8/4.2/4	8/30/24	0	1
Pool	0.4/1.1/0.8	4/4.2/4.1	1/6/4	0	0
Forest	0.7/1.5/1	3.9/4/4	19/77/48	70-100	1
Transition/Fen	0.7/1.4/1	3.9/4.7/4.2	10/51/35	ca. 20	1

	Characteristic bryophytes	Characteristic plants
Hummock	*Sphagnum fuscum, Polytrichum strictum*	*Eriophorum vaginatum, Vaccinium oxycoccos*
Lawn	*Sphagnum magellanicum*	*Eriophorum vaginatum, Trichophorum caespitosum, Andromeda polifolia, Drosera rotundifolia*
Pool	*Sphagnum* cf. *cuspidatum*	*Scheuchzeria palustris*
Forest	*Sphagnum* cf. *angustifolium, Sphagnum* cf. *capillifolium Pleurozium schreberi*	*Vaccinium uliginosum, Vaccinium vitis-idea*
Transition/ Fen	*Sphagnum* cf. *capillifolium, Sphagnum* cf. *fallax*	*Betula pubescens, Potentilla erecta, Viola palustris, Juncus* sp., *Carex* sp.

Results and discussion

Oomycete diversity in Swiss Jura peatlands

Oomycetes were present in all 65 samples, and represented about 3% (107932 sequences) of all microbial eukaryote sequences (4223719 sequences) obtained in the whole study. We detected a high diversity with 34 phylotypes (of 5652 in total) of oomycetes, and samples varying from 4 to 21 phylotypes (average 16). Culture-based studies targeting soils mentioned up to about 10 species for individual samples (Bahramisharif et al., 2014), which suggests similar amounts of diversity. The total diversity reached here was in the same order of magnitude as the 69 oomycete phylotypes found in the TARA project, which encompassed 334 size fractionated samples from marine plankton originating from tropical to temperate oceans worldwide (De Vargas et al., 2015). Obtained phylotypes were often highly divergent: we detected only 3 phylotypes that matched with 100% identity to barcoded taxa (namely *Peronospora schachtii*; X1510, *Phytophthora* sp.; X1794 and *Saprolegnia parasitica*; X1602). In contrast, up to 70% of all phylotypes had less than 98% similarity and, therefore, represented putative new genera, or deeper clades. The most abundant phylotype, X9, could not be placed with confidence within Saprolegniales or Peronosporales, and belongs to one of the deep basal lineages collectively designated as oomycete basal clades; known species from these basal lineages are animal and brown alga parasites (Lara et al., 2011). However, the great diversity of basal oomycetes may possibly host other lifestyles that have still not been investigated. This illustrates the potential for the discovery of a novel diversity in peatlands.

Phylotypes were not distributed randomly in all microhabitats (fig. 5.2 and fig. 5.3). While communities were not significantly different between peatlands ($p > 0.05$), we found that communities were significantly separated in function of their original microhabitat ($p = 0.001$). Specific richness did not differ significantly between microhabitats (pairwise Tukey test $p = 0.97$). However, communities of all microhabitats except pools were largely dominated by phylotype X9 (fig. 5.2), an unidentified, basal branching organism. In contrast, highly represented phylotypes in pools included X102 and X145 (an unidentified Saprolegniales).

Peat bogs host a high number of pathogenic oomycetes. We detected 5 phylotypes out of 34 belonging to obligate parasitic genera (namely *Haptoglossa*; X1167) for animal parasites and *Albugo* (X1632) and *Peronospora* (X1510) for plants. These obligate plant parasites are probably infecting the vascular plants present in the peat bogs. Phyloptype X1510 corresponded 100% to *Peronospora schachtii* (GenBank affiliation: KF888598), a sugar beet pathogen. However, the v9 region of the SSU rRNA gene is a conserved marker and the corresponding sequence can be discriminated only by 1 bp from *P. effusa*, a spinach pathogen which is not closely related in multigene phylogenies (Choi et al., 2015). While the organism from which this sequence derived belongs to a diverse cluster that infects, as far as it is known, only plants from the order Caryophyllales (Choi et al., 2015; Thines and Choi, 2015), further analyses will be required to determine the exact host. *Drosera rotundifolia*, is a member of the Caryophyllales and is common in the studied peat bogs; it could be a good candidate as a host for phylotype X1510. Moreover, these plants are most commonly found in the lawn microhabitat where X1510 was most common fig. 5.3. On the other hand, the *Albugo* sequence (X1632) is not closely related to any barcoded species; we assume that it may derive from any vascular plant of the bog. To our knowledge, bog plant oomycete parasites have never been surveyed. Furthermore, other members of the *Lagenidium/Albugo* clade, as well as basal phylotypes may also be obligate (See Supplementary material). Facultative parasitic organisms are potentially very common, as several genera have been detected (*Phytophthora* (X1794 and X2201), *Lagenidium* (X2385, X3084, X3085), *Saprolegnia* (X1602), *Aphanomyces* (X1100, X1447) and *Atkinsiella* (X2836)). The phylotype X1100 is closely related to both *Aphanomyces astaci* and a copepod parasite from the same genus (Wolinska et al., 2009). It is, therefore, most likely infecting the crustaceans that belong to the peatlands natural fauna (cladocerans, cyclopoid and harpacticoid copepods). Peatland inhabiting organisms may act as reservoirs for economically relevant pathogens such as members of genus *Aphanomyces*. Other oomycete genera are likely pathogens of the peatland microfauna; *Lagenidium* and *Leptolegnia* infect arthropods. Phylotype X2836 is closely-related (98% similarity) to the marine crustacean parasite *Atkinsiella dubia* (AB284575), and may also infect the crustaceans present in the peatland; however, it may belong to another genus. *Haptoglossa* (X1167) infects nematodes, also very common in peat bogs (Glockling and Serpell, 2010). *Saprolegnia* (X1602) is a wide spectrum animal parasite (Sarowar et al., 2014). Genus *Pythium* (X570, X716, X1311, X1846 and X2835) comprises both animal and plant parasites (Lara et al., 2011; Liu et al., 2014), fungal parasites (Benhamou et al., 2012; Horner et al., 2012) and also saprotrophic members (Kwasna et al., 2010).

We have demonstrated here that acidic, oligotrophic environments such as peat bogs host a wide diversity of oomycetes, some of them representing well studied taxa while others are basal branching and remain to be characterized morphologically and placed phylogenetically. Our approach using 'universal' eukaryotic primers may have been key to the detection of these basal clades which would not have been detected by using an oomycete targeted approach. These basal groups are crucial for building comprehensive systematic and phylogenetic concepts about oomycetes and towards understanding their evolution as a group. Furthermore, understanding their life history will lead to the evaluation of their impact on foodwebs and regulation of their hosts' population if they are pathogens. For this purpose, in situ hybridization methods can be applied to characterize morphologically organisms from which only SSU rRNA genes are available, as already applied to the parasitic group Cryptomycota (Jones et al., 2011).

Acknowledgments

Monica Steciow wishes to thank CONICET (PIP 0661CO; Fondos IBOL 2012) and UNLP (Project N11/696) for financial support. This project was funded by the European Union's Seventh Framework Programme under grant agreement 245268 (ISEFOR; LB). Further support came from the SwissBOL project, financed by the Swiss Federal Office for the Environment (LB) and the SciexeScientific Exchange Programme NMS. CH (LB).This work was funded by Swiss National Science Foundation project no. 310003A 143960. We thank Professor Edward A.D. Mitchell for the insightful comments about peatland ecology and for the permission to use his pictures to illustrate the habitats. We thank also an anonymous reviewer for the complete and detailed comments on the manuscript.

Chapter 6

Micro-eukaryotic functional diversity associated with *Sphagnum* mosses in Tropical, Subtropical and Temperate climatic zones

David Singer[1], Sebastian Metz[2], Fernando Unrein[2],
Edward A.D. Mitchell1[1,3], Enrique Lara[1]

[1]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland.
[2]Instituto de Investigaciones Biotecnológicas-Instituto Tecnológico de Chascomús (IIB-INTECH), UNSAM-CONICET, Av. Intendente Marino Km 8.200, (B 7130 IWA) Chascomús, Buenos Aires, Argentina
[3]Jardin Botanique de Neuchâtel, Chemin du Pertuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

FEMS, Microbiology Ecology, submitted

Summary

Microbial eukaryotes are highly diverse, and are key actors in major nutrient cycles. However, the drivers and functional consequence of their diversity and distribution patterns among climatic zones and environmental gradients remain poorly documented. We characterized the taxonomic and functional diversity of microbial eukaryotes in Sphagnum mosses (low pH and nutrients) along elevational gradients in temperate, subtropical and tropical climatic zones through an environmental DNA diversity survey. We hypothesized that community structure would vary among climatic zones and along elevational gradients; the proportion of decomposers and microbial predators and overall phylotype diversity being positively correlated to temperature while the proportion of mixotrophic organisms would decline. Microeukaryotic communities clearly differed among the three climatic zones; only 25% of phylotypes were shared. Richness was higher in the tropical climatic zones than in the temperate region. Parasitic organisms and heterotrophs were most abundant in the (sub)tropical environments, while photosynthetic organisms were most abundant in temperate samples. Mixotroph abundance decreased significantly with an elevation of temperature suggesting a decrease in photosynthetic activity. The contrasted patterns of micro-eukaryotic communities in relation to climate suggest that climate influences nutrient cycling in Sphagnum-dominated ecosystem, which could potentially have implications for carbon storage under a global warming scenario.

Keywords: sphagnosphere, V9 region of the SSU rRNA gene, altitudinal and latitudinal gradient, bryosphere, mixotrophy, protist

Introduction

Documenting the distribution patterns of diversity, its drivers and consequences for ecosystem functioning is a major topic in ecological research. Until recently however it was all but impossible to conduct such research for the multitude of poorly known microscopic eukaryotes living in the soil environment. Therefore, it was unclear to what extent ecological theory applied equally to macroscopic and microscopic organisms (Lara et al., 2016). Understanding the latter is arguably more crucial than the former owing to the dominant roles played by microorganisms in the functioning of all ecosystems (Wilkinson, 1998).

The developments of sequencing technology, such as High Throughput Sequencing (HTS) and the associated bioinformatics and biostatistics developments are currently revealing the extent of microeukaryotic diversity and allowing a reassessment of microbial biogeography and macroecology. Indeed, hundreds of thousands of sequences possibly representing millions of biological species are discovered, including previously overlooked forms such as symbionts, parasites or small heterotrophs (De Vargas et al., 2015; Geisen, 2016). These are extremely abundant in oceans, and probably play major roles in ecosystem functioning (De Vargas et al., 2015). Beyond the concept of microbial loop, where heterotrophic protists are considered mainly as bacterial eaters (Azam et al., 1983; Clarholm, 1985) the role of microbial eukaryotes as parasites and predators of other eukaryotes (eukaryovores) is becoming increasingly acknowledged (De Vargas et al., 2015; Dumack et al., 2016b). Eukaryovory, an ancient phenomenon (Porter, 2016) has been shown to be widespread and significant in soils where phagotrophic protists also play a key role as fungal grazers (Geisen et al., 2015b). HTS is increasingly used to reveal not only the diversity of soil protists but also functional relationships. Furthermore, the number of sequences obtained in environmental DNA surveys was shown to be proportional to the biomass of organisms, be it with mitochondrial (Kosakyan et al., 2015) or ribosomal markers (Giner et al., 2016). Even if such estimates are not absolutely quantitative, it is thus possible to estimate with reasonable confidence the relative abundance and thus is likely importance of a given functional group of organisms in the environment. Ecosystem functioning and the relative importance of the different functional types changes with environmental conditions. Planktonic eukaryotic communities shift from mixotrophy dominated to strict phototrophy with increasing minerotrophy (Jones, 2000; Lara et al., 2015); mixotrophy is also favoured in oligotrophic freshwater bodies (Izaguirre et al., 2016; Saad et al., 2013; Tittel et al., 2003) and in marine systems (Unrein et al., 2007, 2014; Zubkov and Tarran, 2008). High temperatures have also been shown theoretically to favour phototrophic organisms over mixotrophs (Crane and Grover, 2010). In practice, mixotrophs are common in cold areas such as Antarctica, also because the winter ice cap selects organisms that can rely at least partially on heterotrophy (Laybourn-Parry and Pearce, 2007). The micro-environment created by *Sphagnum* mosses is constantly wet (Gorham, 1957) and plants stock water up to 16-25 times their dry mass (Bold, 1973). Indeed, these mosses act like sponges in retaining a considerable amount of water in specialised, hollow cells (hyalocysts) of their leaves and stems. *Sphagnum* mosses are sensitive to drought (Bu et al., 2013), and thus, *Sphagnum*-associated organisms are supposed to live in an almost permanently moist environment. In line with observations in aquatic environments, warming was shown to modify the structure of microbial communities in Sphagnum, causing a decrease in mixotrophs and an increase in decomposers (Jassey et al., 2011, 2015). This effect has however never been studied using high throughput sequencing approaches and existing data are based solely on direct microscopic observations.

A confounding factor of observational studies along elevation and latitudinal gradients is that many factors co-vary, including climate, vegetation and soil type. One way to overcome this limitation is to focus on a very specific habitat. This approach has been used for the study of testate amoeba diversity along an elevational gradient by studying the litter of a dominant plant (*Calluna vulgaris*) and experimentally planted cushion plant (*Minuartia sedoides*) revealing a decline in diversity with elevation (Heger et al., 2016). If the chosen habitat is very specific, then confounding factors will be reduced to a minimum because ecological filtering will be maximal. Here, we focused on *Sphagnum* mosses, an environment characterized by its stability and particularly selective conditions, including low pH and nutrient content (Dedysh et al., 2006) and high concentration of humic acids which are known to have biocidal properties (Steinberg et al., 2006). These conditions strongly shape the composition of microbial communities, leading to highly specialized assemblages, as shown for prokaryotes (Dedysh et al., 2006), fungi (Thormann and Rice, 2007) and micro-metazoa (Gilbert et al., 2006). Overall eukaryotic communities are not only different from those of other environments, they are also highly diverse (Lara et al., 2011). The association between mosses and these organisms coined bryosphere (Lindo and Gonzalez, 2010) is therefore particularly specific in the case of *Sphagnum* mosses and referred to as the sphagnosphere (Jassey et al., 2011). Therefore, because of these particular characteristics, *Sphagnum* mosses are an excellent model to study (macro)ecological questions such as the impact of climate on microbial eukaryote diversity and ecosystem functioning.

Sphagnum mosses are the main components of high latitude peatlands especially in the Northern Hemisphere. These particular wetlands are well-known carbon sinks, sequestering 12% of all anthropogenic emissions of CO2 (Moore, 2002). In spite of covering only 3% of the Earth's emerged surface (Kivinen and Pakarinen, 1981), at the global scale, this process is of considerable significance: 30% of the world's pool of organic carbon is stored in peat bogs (Gorham, 1991). However, under a global warming scenario, it has been demonstrated that these ecosystems revert into carbon sources (Dorrepaal et al., 2009). We surveyed the microbial eukaryotic communities in the Sphagnosphere and their changes across gradients of latitude and altitude, in order to characterize the patterns of diversity and community composition in relation to climate. We used three altitudinal gradients representative of three different climatic zones, temperate (Europe), subtropical (Southern Japan) and tropical (Costa Rica). We hypothesized that species richness and diversity decrease in function of latitude and altitude, according to the species energy theory (Currie et al., 2004). As decomposition rates should be enhanced with increasing temperatures, the proportion of decomposers (fungi and bacteria) (Jassey et al., 2011), and likely also of bacterivorous eukaryotes should increase under warmer while the proportion of mixotrophic organisms would decline (Jassey et al., 2015). The effect on higher tropic levels may however be less marked as the magnitude of treatments effects typically declines from one trophic level to the next (Wardle, 2002). Therefore, community composition – and especially the relative proportion of organisms belonging to different functional categories such as decomposers, predators (phagotrophs) and mixotrophs should be strongly correlated to climate and primarily to temperature both among biomes and, within each biome, along elevational gradients.

Materials and methods

Sampling, PCR and sequencing

The study was carried out in three contrasted biomes: Temperate (Switzerland-Italy 46°50'N 8°10'E), Subtropical (Japan, 30°20'N, 130°30'E), and Tropical (Costa-Rica, 9°30'N, 83°28'W). We collected 21 samples (8 tropical, 7 subtropical and 6 temperate) of the top 3 cm of *Sphagnum* mosses under sterile conditions (sup table 1). These samples were stored into LifeGuard® buffer to fix and preserve the DNA prior DNA extraction. DNA was extracted directly from *Sphagnum* stems using a MoBio Power SoilTM DNA extraction kit® (Carlsbad, CA, USA) according to the manufacturer's instructions. The SSU rRNA V9 regions was amplified using the broad spectrum eukaryotic primers 1380F/1510R described in (Amaral-Zettler et al., 2009). Sequencing was performed with an Illumina's HiSeq 2500, using V3 chemistry (Fasteris, Geneva, Switzerland).

Sequences processing

Quality check (Phred score filtering, elimination of reads without perfect forward and reverse primers, and chimera removal of the sequences was performed following the pipeline developed by de (De Vargas et al., 2015) and applied in (Singer et al., 2016). Sequences were grouped in phylotypes using the SWARM clustering algorithm (Mahé et al., 2014). We retained phylotypes that comprised more than 10 sequences in at least 3 samples, as rare sequences tend to blur ecological signal (Schiaffino et al., 2016). Taxonomic assignations of the phylotypes were performed by aligning the dominant sequences of each phylotypes to the reference PR2 database (Guillou et al., 2013) using GGSearch script in the FASTA package (Pearson, 2014). We filtered out sequences belonging to groups that were not assigned to microbial eukaryotes (Metazoa, Embryophyceae, Archaea, Bacteria, and Streptophyta), an important step as we needed to remove sequences of *Sphagnum*, which

could represent up to 50% of the number of reads per sample. In order to homogenize the number of sequences and to be able to compare samples richness, we normalized the numbers of reads by reducing it to smallest sample (here, 8070 sequences). Extra sequences were removed randomly from the dataset by using the function "rrarfy" from the R vegan package (V. 2.3-5) (Oksanen et al., 2015). In order to verify if the total diversity were reached we compute rarefaction curves with a visual check of the asymptotic behaviour for each sample (Supp fig 1) and for each biomes (Supp fig 2).

Environmental data

We used the coordinates for each sample to extract the biologically relevant bioclimatic variables from the finest resolution grids (30 arc-second) of the WorldClim project (Hijmans et al., 2005). These 19 variables were extrapolated from data on monthly temperatures and rainfall. They have been used for species distribution modelling in two soil protist groups, protosteloid amoebae (Aguilar and Lado, 2012) and euglyphid testate amoebae (Lara et al., 2016). In order to avoid excessive correlation among the explanatory variables, we performed a stepwise selection based on the variance inflation factors (VIF) with the recommended (and most stringent) threshold of five. Stepwise selection was performed using the customs script (https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/) on R (V.3.1.2) (R Core Team, 2013). The following variables were selected for further analysis: Mean Temperature of Wettest Quarter (MTWeQ), Mean Diurnal Range (MDR), Mean Temperature of Driest Quarter (MTDQ), Mean Temperature of Warmest Quarter (MTWaQ) and Precipitations of Coldest Quarter (PCQ) (sup table 2).

Functional assignation

We assigned all well represented taxa to major functional groups, based on their trophic mode: Osmotrophic (Ascomycota, Basidiomycota, Chytridiomycota, Glomeromycota, and Mucoromycota), Phototrophic (Archaeplastida, Bacillariophyta), Parasitic (Apicomplexa, Entomophtoromycota, Ichtyosporea, and Oomycota), Phagotrophic (Ciliophora, Cercozoa) and Mixotrophic (Cryptophyta, Dinophyta, and Haptophyta) (Supp table 3). Mixotrophic organisms are defined here as those organisms that use both phagotrophy and photosynthesis as a carbon source (Sanders, 1991). We considered as mixotrophic all cryptophytes with the exception of the Goniomonadea and related early branching taxa (Shiratori and Ishida, 2016). Although Dinophyta include many heterotrophic species in marine environments, these are anecdotic in freshwater systems (Gomez, 2012), therefore we did take them into account. Chrysophyceae (sensu (Adl et al., 2012)) are notorious for evolving quickly trophic strategies from mixotrophic to heterotrophic (Boenigk et al., 2005b). For these organisms (as well as Euglenida), we verified manually the precise phylogenetic affiliation of all phylotypes in order to infer their lifestyle, i.e. when the phylotype was included within a strictly mixotrophic group, we deduced that it was indeed mixotrophic (Supp table 4); all taxa that could not be assigned to functional category were labelled "unknown". We combined all phylotypes within the different mentioned functional categories (Supp table 5).

Data analyses

We assessed the relationships between the phylotypes found in the 21 samples and environmental and functional variables using a partial redundancy analysis (RDA) on Hellinger transformed data (Legendre and Gallagher, 2001). The significance of the variables and ordination axes (first and second) were assessed using a One-way Analysis of Variance (ANOVA; 1000 permutations, p-value threshold = 0.05) (Chambers and Hastie, 1991). These analyses were performed with R V.3.0.1 (R Core Team, 2013) and the R vegan package (V. 2.3-5) (Oksanen et al., 2015).

We tested if there was a significant correlation between altitude and the specific richness in each biome with a linear model (LM; p-value threshold = 0.05) (Chambers and Hastie, 1991). We also tested the correlation between the significant variables and the functional groups for the whole dataset using the same approach. We then investigated if phylotype richness and functional types were differently represented in the respective biomes using a Pairwise Test for Multiple Comparisons of Mean Rank Sums (Nemenyi-Tests) with the R package PMCMR V4.1. (Sachs, 2013).

In order to identify characteristic organisms for each biome, we performed an indicator species analysis. In this purpose, we used a multi-level pattern analysis with the function "multipatt" from the R package indicspecies V1.7.5 (de Caceres and Legendre, 2009) to find the significant indicator phylotypes in our dataset. Then a heatmap was computed using the 'heatmap.2' function of the gplots package (Warnes et al., 2015). We considered only phylotypes that occurred at least 100 times in the dataset in order to avoid noise and bias due to rare species for the indicator species analysis (Schiaffino et al., 2016).

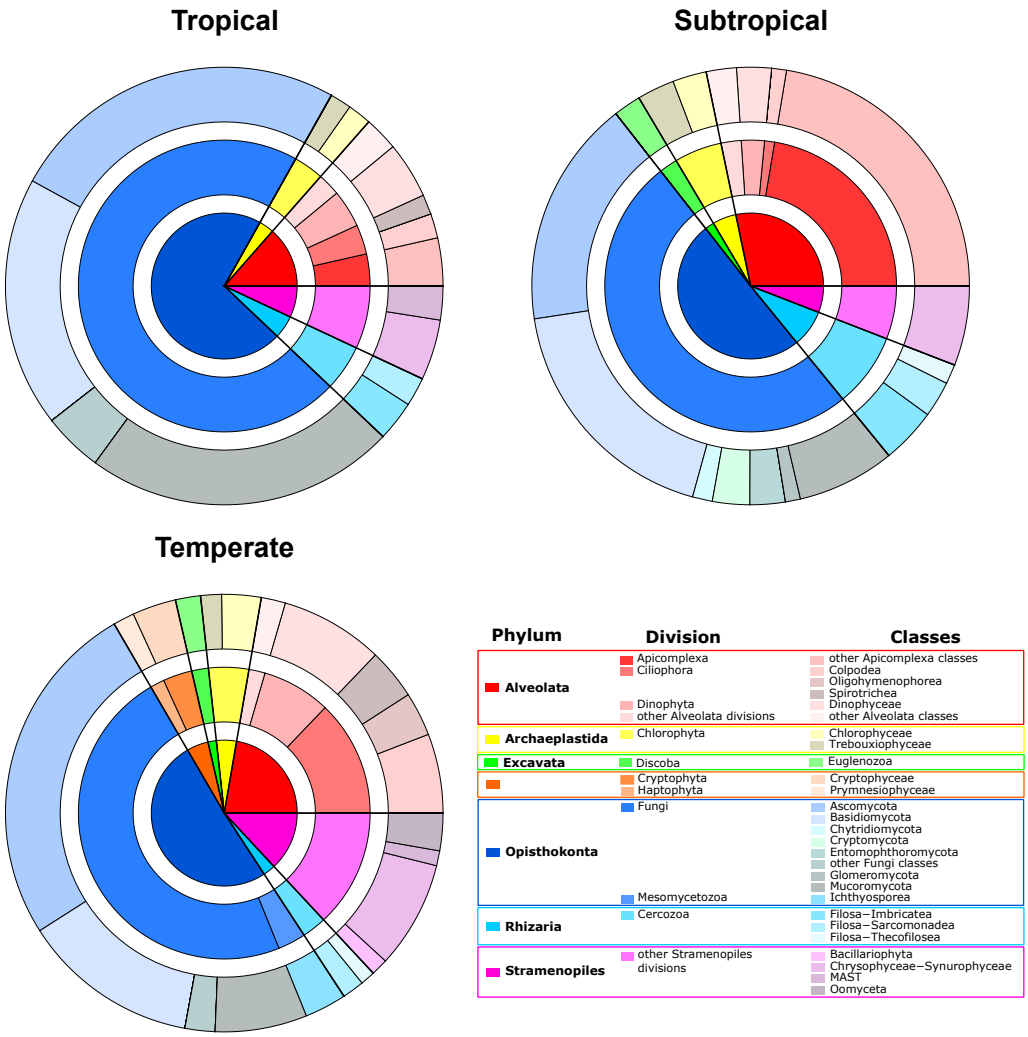

Figure 6.1 – Pie Chart representing the community composition of microeukaryotes living in *Sphagnum* sp. for three different taxonomic levels (phylum, division and classes) in three different climatic zones (tropical, subtropical and temperate). Taxa representing less than 1% of the total of a given level/taxa are represented in the "other" sections.

Results

After removing metazoan and embryophyte phylotypes, as well as low quality and chimeric sequences, we obtained a total of 169'470 sequences and a total of 3035 phylotypes (Tropical = 2236, Subtropical = 2105 and temperate = 1441) (fig. 6.2D), representing all eukaryotic major clades (fig. 6.1) (raw data and a file containing all phylotypes will be sent to GenBank upon acceptance of the manuscript). The rarefaction curves computed for each biome and for each sample reached a plateau, showing that the major part of diversity had been covered (sup fig 1 and sup fig 2). Phylotypes present in all three biomes represented only 24.8% (i.e. 753 phylotypes) of the total diversity. Tropical and subtropical biomes shared 62.5% of their phylotypes while samples of the subtropical and tropical biomes had respectively 33.6 and 34.2% of their phylotypes in common with samples from the temperate biome (fig. 6.2D). Phylotype richness decreased along the longitudinal gradient. Phylotype richness was significantly higher in the tropical biome as compared to the temperate biome ($p = 0.0075$), and was intermediate in the subtropical biome and not significantly different from either of the other two biomes (fig. 6.2C). Phylotype richness was not correlated to altitude (fig. 6.2A).

Tropical and subtropical biomes were characterized by high abundance of sequences related to the parasitic apicomplexa, a group that was marginal in the temperate climatic zone (fig. 6.1). In the temperate biome, the most abundant parasitic group was the Mesomycetozoa (2.73%), which represented less than 1% of all sequences in the other climatic zones (fig. 6.1). Tropical samples were dominated by Opisthokonta, which represented 67.8% of the total abundance of the sequences, these percentages dropped to 49.6 and 48.4% respectively in subtropical and temperate environments. Cryptophyta and Haptophyta phylotypes were well best represented (3% and 1.8%, respectively) in the temperate biome but insignificant in the other two climatic zones (fig. 6.1).The partial redundancy analysis (fig. 6.2B) of community composition from the three climatic zones showed that

phylotypes assemblages differed significantly among the three climatic zones. The first two axes (F1 p = 0.001 and F2 p = 0.001) together explained 27.4% of the total variance with six significant variables: three functional variables (Mixotrophs (p = 0.004), Parasites (p = 0.004), Osmotrophs (p = 0.001)) and three environmental variables (MDR (p = 0.013), MTWeQ (p = 0.001) and MTWaQ (p = 0.01)). The temperate biome was associated with mixotrophs, the subtropical biome with parasites and the tropical biome with osmotrophs.

The amount of osmotrophs, phagotrophs and phototrophs did not vary significantly among biomes (fig. 6.4). However, parasites were significantly more abundant in the subtropical biome than in both the tropical and the temperate climatic zones (p = 0.04 and p = 0.0004, respectively). Mixotrophs were more abundant in the temperate biome than in the subtropical climatic zones (p = 0.006) (fig. 6.4). The abundance of mixotrophs was significantly inversely correlated to MTWaQ (p = 0.006) (fig. 6.5).

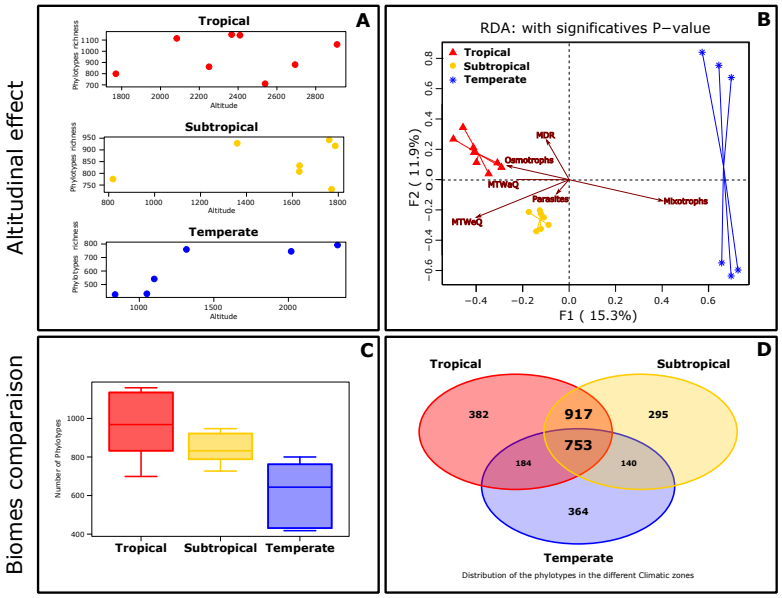

Figure 6.2 – Four graphs representing the effect of the altitudes and the effect of the climatic zones (tropical in red, subtropical in yellow and temperate in blue) on the phylotypes richness of microeukaryotes living in *Sphagnum* sp. A: Biplot representing the altitude (m a.s.l.) versus the phylotypes richness for each climatic zones. B: Redundancy analysis (RDA) builds with the 6 significant variables ($p < 0.05$): Mean Diurnal Range (MDR), Mean Temperature of Wettest Quarter (MTWeQ), Mean Temperature of Warmest Quarter (MTWaQ), mixotrophs, parasites and osmotrophs. C: Boxplot of phylotypes richness per climatic zones. D: venn diagram representing the phylotypes find in each climatic zone, intersection represents common species of the considered biome.

The indicator analysis identified 89 phylotypes significantly indicating either one biome or two biomes (tropical = 16, subtropical and tropical = 47, subtropical = 12, temperate = 12, tropical and temperate = 1, and tropical and temperate = 1) (fig. 6.3). Gregarina (parasite) phylotypes indicates subtropical and tropical climatic zones. Opisthokonta (osmotroph) are a good bioindicators for subtropical and tropical climatic zones. Temperate climatic zones are characterized by the presence of two Cryptista indicators (fig. 6.3).

Discussion

Phylotype richness

We observed significantly higher phylotype richness in tropical and subtropical locations than in the samples originating from temperate climatic zones (fig. 6.2C). These results are in phase with the latitudinal gradient of diversity (Hillebrand, 2004). This gradient, which was first documented for macroscopic organisms has also been observed in several protist taxa such as Dictyostelida (Perrigo et al., 2013) Euglyphida (Lara et al., 2016), radiolarians (Boltovskoy et al., 2010), marine diatoms (Salinas et al., 2015) and freshwater phytoplankton in general (Izaguirre et al., 2016; Stomp et al., 2011). However, some other protist groups do not seem to follow this trend, including marine benthic ciliates (Azovsky and Mazei, 2013) and freshwater diatoms (Passy, 2010; Soininen et al., 2016). If indeed the patterns of diversity differ among taxa, then the drivers of these patterns are also likely to vary. When entire communities of soil protists were surveyed in a large latitudinal gradient with an environmental DNA sequencing approach, no consistent pattern related to temperature appeared (Bates et al., 2013). Given the above-mentioned discrepancies, this could possibly be due to the presence of contrasted patterns among the groups.

Even though our results show higher diversity in the (sub)tropics, we didn't find diversity patterns consistently correlated with elevation in any of the three climatic zones. This suggests that locally, factors other than temperature drive community composition, all other parameters being relatively equivalent. Historical reasons, related to the individual evolutionary patters of all concerned taxa may, at least in part, explain these patterns, perhaps superposed with the effects of temperature.

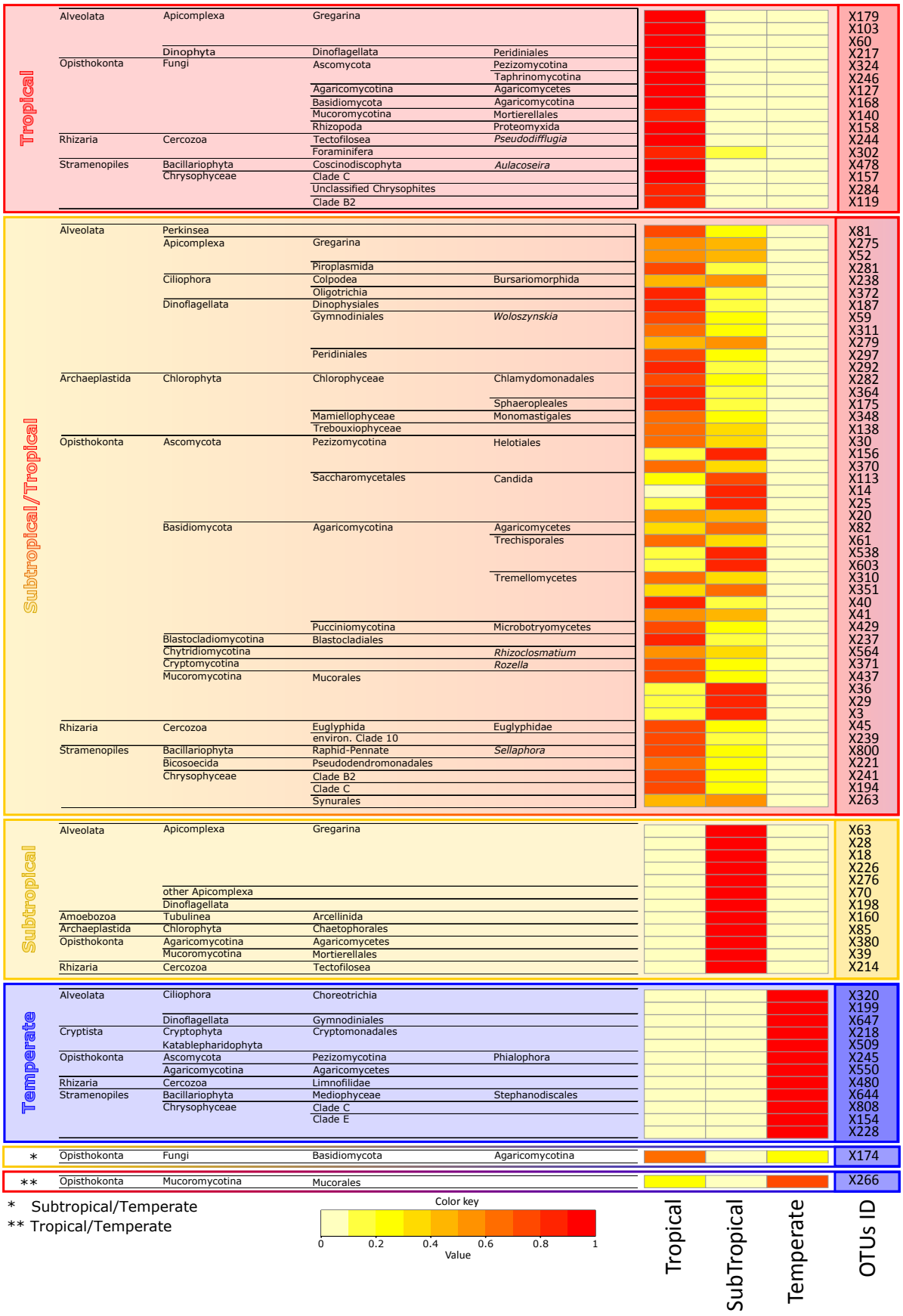

Figure 6.3 – Heat map representing the abundance of the bioindicator phylotypes occurring at least 100 sequences) found in the dataset. Taxonomic affiliations were manually check verified against the NCBI database using BLAST algorithm.

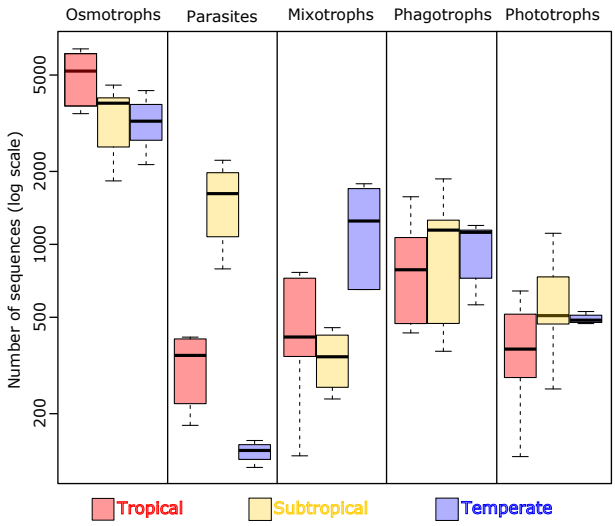

Figure 6.4 – Boxplot representing the abundance of sequences assigned to each functional groups (Parasites, Osmotrophs, Phagotrophs, Photorophs and Mixotrophs) for the three climatic zones (Tropical in red, Subtropical in yellow and Temperate in blue).

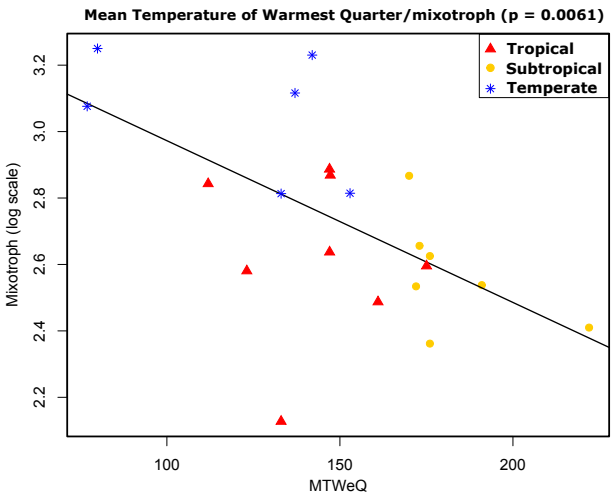

Figure 6.5 – Biplot representing the mixotrophs (abundance, log scale) in function of the Mean Temperature of Wettest Quarter (MTWeQ). Linear regression (p = 0.0061), blue star = temperate climatic zone, yellow circle = subtropical climatic zone and red triangle = tropical climatic zone.

Because we found a higher diversity in the tropics, the main factors correlated to community richness were environmental variables related to temperature (MDR, MTWeQ, and MTWaQ). The selected significant variable related to precipitations were not correlated to richness (even if the biomes were considered separately), most probably because of the ability of Sphagnum sp. to keep their relative humidity content high by taking up water (passively, by capillarity) from the water table depth (Ketcheson and Price, 2014) and thus compensate the water lost by evaporation (Kettridge and Waddington, 2014).

Functional diversity

While the number of sequences related to phagotrophs and phototrophs did not differ significantly amongst climatic zones, osmotrophs, parasites and mixotrophs were most abundant in, respectively, tropical, subtropical and temperate climatic zones. These differences may reveal contrasted modes of functioning and interactions webs among climatic zones.

Sequences related to parasitic organisms are especially numerous in the subtropical, but also in tropical samples. Most belong to the Apicomplexa, a group of protist parasitizing exclusively Metazoa (Grabda, 1991). This high abundance in warm climates is consistent with findings in tropical soils, and has been associated to higher metazoan diversity in tropical forest litter (Mahe et al., 2017). These organisms are rare in the temperate biome; by contrast, sequences related to Mesomycetozoa (absent in other climatic zones) can be found. Mesomycetozoa is also composed of Metazoa parasites to a large extent (Mendoza et al., 2002). They could, possibly, replace Apicomplexa in colder environments.

Fungi was the most diverse group in the tropical samples, and among them Mucoromycotina, a group of mostly saprotrophic fungi with a few exceptions (Field et al., 2015) were especially well represented (31% of fungal phylotypes). Mucoromycotina were far less represented in other climatic zones. Six bioindicators for tropical samples out of 16 were also fungi, including yeasts, saprotrophs and one possible ectomycorrhizal form (fig. 6.3). Saprotrophy implies heterotrophic utilization of carbon obtained through the degradation of organic matter.

Temperate samples are characterized by the highest amount of sequences corresponding to mixotrophic organisms, i.e. organisms able to use both CO_2 (owing to the presence of chloroplast or pigmented symbionts) and organic carbon (through phagocytosis) for carbon acquisition. This concerns members of the Cryptophyta, Haptophyta, Dinophyta and some Chrysophyceae. The total number of sequences of mixotrophic organisms is strongly correlated with several variables related to temperature, including the temperature of the coldest month (fig. 6.5). It has been demonstrated that mixotrophic species tend to use more their phagocytosis ability under low temperatures (Wilken et al., 2013). It can therefore be hypothesized that, under warm climates, selection acts against mixotrophs and favours heterotrophic relatives, as the photosynthetic function is not needed anymore. In line with these theoretical considerations, a decrease in the number of mixotrophic testate amoebae has been demonstrated in response to in situ experimental warming in a temperate *Sphagnum*-

dominated peatland (Jassey et al., 2013). This reduced the microbial photosynthesizing population considerably and as a consequence the carbon-sink function of the peatland (Jassey et al., 2015).

The carbon sequestration of peatlands results from the balance between decomposition and net primary production. Our results show that under warmer climates the proportion of decomposers is higher and that of mixotrophs lower. It has been suggested that mixotrophic protist abundance has a significant effect on C fixation in Northern Hemisphere peatlands (Jassey et al., 2015); our result would then suggest that *Sphagnum*-dominated ecosystems would fix more carbon under cold climates. Warm climates would rather favour decomposition (and hence, respiration), and therefore, one could expect a different carbon balance shifted towards C release. Direct measurements of CO_2 fluxes on tropical and temperate *Sphagnum*-dominated ecosystems, as well as experimental studies under different climates including tropical regions should be undertaken in order to further test this. These results would be even more relevant under a global warming context, knowing that boreal *Sphagnum*-dominated peatlands contain one third of global soil carbon pool (Post et al., 1982; Yu et al., 2011).

Acknowledgements

We would like thanks Christophe V.W. Seppey for the help on the statistical part and for the fruitful discussion. Then this work was funded by Swiss National Science Foundation project no. 310003A 143960 given to EL. SM received a short-term fellowship from the European Molecular Biology Organization (EMBO). SM was also supported by a CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina) doctoral fellowship and the Argentinean project FONCyT PICT-2014-1290. FU is a CONICET researcher.

Chapter 7

Born in America: A molecular phylogeography of *Hyalosphenia papilio*

David Singer[1], Quentin Blandenier[1], Clément Duckert[1], Edward A.D. Mitchell1[1,2], Enrique Lara[1]

[1]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland.
[2]Jardin Botanique de Neuchâtel, Chemin du Pertuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

In prep

Abstract

The increasing use of molecular tool to assess the diversity of free-living microorganisms has revealed the existence of hitherto unsuspected diversity. The next question is now to explore the geographical distribution and phylogeographical history of this diversity and compare it to that of better documented macroscopic organisms.

We investigated the spatial distribution of mitochondrial lineages (based on COI, i.e. partial cytochrome oxidase subunit 1 sequences) within *Hyalosphenia papilio*, a testate amoeba with Holarctic distribution in *Sphagnum*-dominated peatlands. This species represents a complex of cryptic species with contrasted geographical distribution. We expanded the inventory of mitochondrial haplotypes, clarified their distribution in the Northern Hemisphere, and inferred the geographical origin of this taxon.

Of the 13 identified lineages, nine showed restricted geographical distributions and were absent from Europe while four were well distributed across the Holarctic realm and these were the only lineages found in Europe. This evidence, in addition to our ancestral reconstruction, suggests a North American origin for *Hyalosphenia papilio*, and four colonisation events of the Palaearctic realm.

Our results show the existence of significant unknown cryptic diversity in free-living protists corresponding to dispersal and speciation events similar to what is observed in macroscopic organisms. The phylogeography of *Hyalosphenia papilio* is congruent with range contractions and expansions related to Quaternary glacial - inter-glacial cycles.

Introduction

Microbial eukaryotes are a key component of all ecosystems, as primary producers (Falkowski, 2002), decomposers (van der Wal et al., 2013), and microbial grazers (Azam et al., 1983; Bonkowski, 2004; Clarholm, 1985) and are taxonomically (Adl et al., 2012) and functionally (Wilkinson and Mitchell, 2010) extremely diverse. While local environmental factors including pH (Bates et al., 2013), nutrient and water availability (Adl and Gupta, 2006; Marcisz et al., 2014b), or anthropogenic disturbances such as pollution (Lara et al., 2007b) strongly influence micro-eukaryotic diversity, the patterns and causes of geographical distribution are still matter to debate. Recent developments in molecular methods have led to questioning the once dominant view on microbial biogeography summarised by the tenet "everything is everywhere, but, the environments selects" (Baas-Becking, 1934).

Free-living microorganisms were thought to have a cosmopolitan distribution owing to the fact that they are small, have the ability to form huge populations and build resistant cysts that can be dispersed over large distances (Finlay, 2002). The alternative view was that a proportion of species have limited distributions (i.e. "moderate endemicity" model) (Foissner, 2006). Most protists species were described based on morphology only ("morphospecies"), which were often disputable and hence did not allow sound biogeographical interpretation (Heger et al., 2010; Mitchell and Meisterfeld, 2005). Methodological developments such as single-cell DNA barcoding of protists (Pawlowski et al., 2012) now allow identifying biogeographically-relevant - but often morphologically cryptic - taxonomic entities, as illustrated e.g. by arcellinid testate amoebae (Kosakyan et al., 2013; Singer et al., 2015).

Hyalosphenia papilio is a testate amoeba morphospecies, restricted to Northern Hemisphere *Sphagnum* peatlands (Meisterfeld, 2002) where it may dominate microbial communities (Jassey et al., 2013). DNA barcoding using cytochrome oxidase I (COI) showed that it represents a complex of genetically distinct species (Heger et al., 2013; Kosakyan et al., 2012). This morphospecies is relatively large (ca. 100 µm) and thus unlikely to being transported over long distance passively (Wilkinson et al., 2012). In addition, given its ecological requirements, the likelihood that propagules find suitable environments is low; it can therefore be hypothesized that *H. papilio* colonizes new environments slowly and lineages within *H. papilio* should show restricted distributions and geographically structured populations. A recent study (Heger et al., 2013) showed that mitochondrial lineages are not distributed randomly, some have limited geographical distribution, and that lineages distribution are best explained by climate (Heger et al., 2013). However, the phylogenetic and geographical coverage of this study were uneven: sampling in eastern North America and Asia were limited and some phylogenetic lineages were represented by only few sequences. Our goal was therefore to expand these data, geographically and phylogenetically, and to infer the likely origin of *H. papilio* and patterns of diversification and dispersal.

We compiled all available mitochondrial COI data from single cells of *H. papilio* and isolated and barcoded 58 additional individuals from *Sphagnum* samples targeting under-sampled regions North America and Asia. Based on this, we clarified the distribution areas of the different lineages and haplotypes encountered and used phylogenetic data to infer the probable origin of the species complex.

Materials and methods

Dataset preparation

We retrieved all existing sequences of COI gene of *Hyalosphenia papilio* in GenBank, together with precise information on the origin of the cells; this information was retrieved from four different studies (Gomaa et al., 2014; Heger et al., 2013; Kosakyan et al., 2012; Oliverio et al., 2015). This dataset consisted in 360 sequences from 50 sites (Supp mat 1). In order to complete the dataset by adding more sites and sequences, we added the sequences of 58 new cells taken from 13 sites (seven in the Eastern North America, four in Asia and two in Europe) (sup mat 1) (fig. 7.5).

Single Cell isolation, PCR amplification and DNA sequencing

Single cells of *Hyalosphenia papilio* were obtained by sieving the top five centimetres of fresh *Sphagnum* mosses using a 100 µm mesh, and isolated individually by using a narrow diameter pipette under an inverted microscope (Olympus IX81). The cells were washed three times in autoclaved distilled water before DNA extraction, which was performed using a guanidine thiocyanate-base protocol (Chomczynski and Sacchi, 1987) adapted after (Gomaa et al., 2013). Amplifications of COI gene fragments (525 bp) were performed in two steps following a nested PCR protocol: a first PCR was achieved with the general COI primers LCO1490 and HCO2198 (Folmer et al., 1994) in the first reaction. These products were used as templates for a second, *H. papilio*-specific PCR using the couple HPcoxR: 5'-ATA CAA AAT AGG ATC ACC TCC ACC-3' and HPcoxF: 5'-GTT ATT GTT ACT GCT CAT GCC-3'. The first DNA amplification profiles consisted in an initial denaturation steps for 3 min at 95°C, followed by 39 cycles of 15 sec. of denaturation at 95°C, 15 sec. of annealing at 43°C and 1 min of elongation at 72°C with an additional final elongation step at 72°C for 10 min. The second PCR profile was the same as the first one except that the annealing temperature was increased to 55 °C. Sequencing was carried out using a BigDye197 Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and analysed with an ABI-3130XL DNA sequencer (Applied Biosystems). COI sequences were manually aligned and edited using Bioedit (V.7.2.3 (Hall, 1999)). (Sequences will be provided after acceptance of the manuscript).

Lineage delineation strategies

In order to be coherent with previous studies, we delimited genetic lineages following the strategy described in (Heger et al., 2013). Briefly, we used three different methods: ≥1% sequence divergence (Kimura, 1980), Automatic Barcode Gap Discovery (ABGD) (Puillandre et al., 2012) and the General Mixed Yule-Coalescent (GMYC) (Fontaneto et al., 2007, 2015; Fujisawa and Barraclough, 2013) model. We computed the sequences divergences using the Kimura 2-parameter (Kimura, 1980), with the recommended threshold of ≥1% (Nassonova et al., 2010) using the "ape" (V3.2. (Paradis et al., 2004)) package in R (V3.0.1. (R Core Team, 2013)). ABGD analysis was computed using the web-server of ABGD http://wwwabi.snv.jussieu.fr/ public/abgd/abgdweb.html with the Jukes-Cantor (JC69) model and parameters set to default values. We applied the GMYC model based on an ultrametric tree obtained using BEAST 1.8.3 computed with parameters proposed in the Beast manual: Tamura Nei + Gamma + Invariant model identified as above, with the Yule speciation model, and the Lognormal Uncorrelated Relaxed Clock model (Drummond et al., 2012). The GMYC analysis per se was performed with the SPLITS package, version 1.0-19 (Fujisawa and Barraclough, 2013) coded in R, version 3.1.2 (R Core Team, 2013)).

Haplotype and lineage network

Haplotypes were assigned to the previously determined lineages. Haplotype networks were constructed using the minimum spanning network analysis as implemented in the software package in PopART (V1.7 (Leigh and Bryant, 2015)). "Traits" used in the analysis corresponded to rough geographical localisations of the sequences in the Northern Hemisphere (NW America, NE America, Europe (from the Atlantic Ocean to Ural Mountains) and Asia (from The Ural Mountains to the Pacific Ocean; Supp table 1). The same analysis was conducted with all haplotypes collapsed in their respective lineage and represented by a pie-chart representing the proportion of cells found in each respective geographical localisations.

Phylogenetic reconstructions

Phylogenetic reconstructions were based on a matrix of 76 unique sequences among the 418 considered in this study. We construct both a Maximum Likelihood (ML) and a Bayesian tree with the RAxML algorithm (Stamatakis et al., 2008) and MrBayes V. 3.2.6 (Ronquist and Huelsenbeck, 2003) respectively. TrN + I + G

was selected as the best model using the Modeltest program (Posada and Crandall, 1998). The ML analysis was performed with 1000 bootstraps, which are represented on the nodes. Bayesian MCMC analysis was carried out with two simultaneous chains, and 50,000,000 generations were performed. Trees were sampled every 1000 generation and the burning was done at 25%. The trees were not rooted because the position of genus *Hyalosphenia* was clear among Hyalosphenidae.

Ancestral reconstruction

We used the Bayesian binary Markov chain Monte Carlo (BBM) analysis implemented in RASP v3.1 (Yu et al., 2015) (but originally developed for MrBayes (Ronquist and Huelsenbeck, 2003) and maximum parsimony analysis as implemented in Mesquite v3.11 (Maddison and Maddison, 2001) to infer past distribution of the lineages of *H. papilio* from their current distribution. We used the BEAST generated tree for these analyses. We used the ultrametric tree obtained previously with BEAST (Drummond et al., 2012) as an input for both RASP and Mesquite analyses. The BBM analysis was run with 10 MCMC chains run for 5'000'000 generations and sampled every 100 generations for a total of 50'000 samples, of which 12'500 were discarded as the burn-in. The states frequency was set with the Fixed Jukes Cantor model and among site variation was set as equal. Nodes were set to be possibly distributed in more than one area. For the maximum parsimony analysis, the number of possible areas was set to one.

Results

Lineage delineation

The Kimura test with a threshold of $\geq$1% sequence divergence inferred 14 distinct lineages (fig. 7.3). The GMYC analysis was based on the single threshold methods and revealed 13 independent entities (nine clusters and four singletons). The confidence interval ranged between 10 and 29 and the threshold was significantly supported (P = 0.046). The ABGD analysis yielded a distinctive barcoding gap of 7% and defined 13 lineages. A more restricted threshold (e.g. <5% of divergence) would overvalue the number of lineages (>20) whereas a broader threshold (e.g. >35% of divergence) would underestimate it (6). Both GMYC and ABGD are in accordance with (Heger et al., 2013). Our new sequences allowed us the discovery of a new lineage which was named lineage M. This lineage was supported by all analysis even if the Kimura test would suggest the division in two different lineages (fig. 7.3). The lineages E and D were considered differently depending on the analyses: they were dissociated by the ABGD analysis and the Kimura test but were regrouped by GMYC analysis. The opposite case occurs for the lineage F: it is confirmed by the ABGD analysis and Kimura test but the GMYC propose a separation in two distinct lineages. For the other lineages, all the three analyses were in adequacy and in agreement with (Heger et al., 2013).

Phylogenetic reconstruction

The ML and Bayesian tree (fig. 7.3) showed a deep phylogenetic division between two robust clades. The smallest one was composed of four lineages (J, K, L and M) all strongly supported with posterior probability and bootstraps values (i.e. 1 and >80 BS respectively). This clade had especially long branches (e.g. lineage L and M had more than 10% divergences). The second clade was the most diverse with eight lineages (A-I) and most of its nodes were poorly supported. Indeed, except for the node between lineages A and B (75 BS and 99 PP), all nodes had lower support values (BS<75 and PP<80) (Fig 4). This second clade possesses a particularly high number of haplotypes (fig. 7.1). The sequences from lineage M were very divergent from the others and branched in a basal position relative to lineages J and K (fig. 7.3). Only a single haplotype was retrieved from Lineage E (sensus (Heger et al., 2013)) and five from lineage D (sensus (Heger et al., 2013)). There again the genetic divergence was very small (i.e. at most six nucleotides difference between the sequence of lineage E and the five sequences of lineage D, all of which were separated by a single nucleotide (fig. 7.1)).

Haplotype network

The haplotype network (fig. 7.1) shows that some lineages (B, H, and L) are composed of only one haplotype, whereas some are made of several haplotypes. Some haplotypes remain under-sampled (e.g. L, B and M with two, seven and seven specimens respectively) whilst others were largely retrieved (e.g. lineage A was retrieved more than a hundred times). fig. 7.1 shown that some lineages are genetically close, e.g. A and B which possess at most five mutation divergences, while others are much more divergent, e.g. L possesses more than 30 mutation dissimilarity compare to all other lineages. The repartition of the different lineages (fig. 7.5, fig. 7.1 and fig. 7.2) showed that several haplotypes are endemic to a geographical area (e.g. B, D, E, H, I and L occur

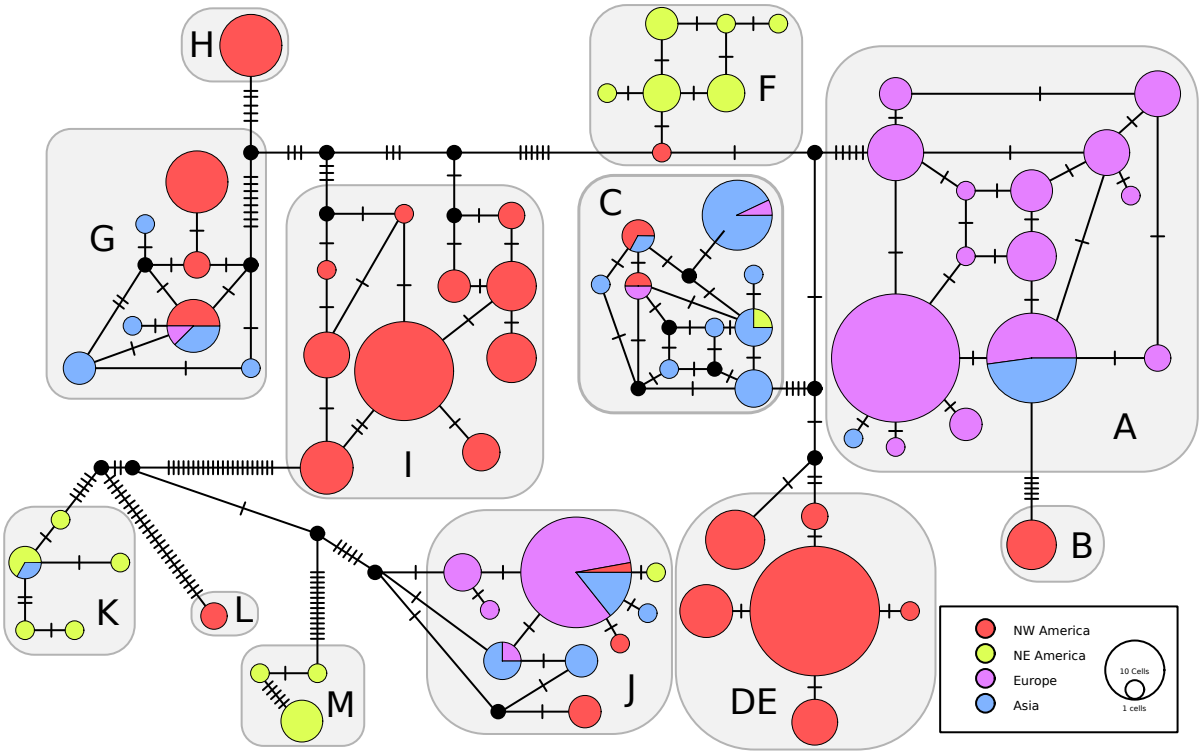

Figure 7.1 – A median joining haplotype network of cytochrome oxidase subunit 1 (COI) gene of *Hyalosphenia papilio*. The grey boxes represent the separation of the different lineages. Color of the pie chart: red =North West of America, yellow = North East of America, Purple = Europe and blue = Asia. Circle size is proportional to the number of single cells of *H. papilio* within a given haplotype and lines between haplotypes represent mutational steps within alleles.

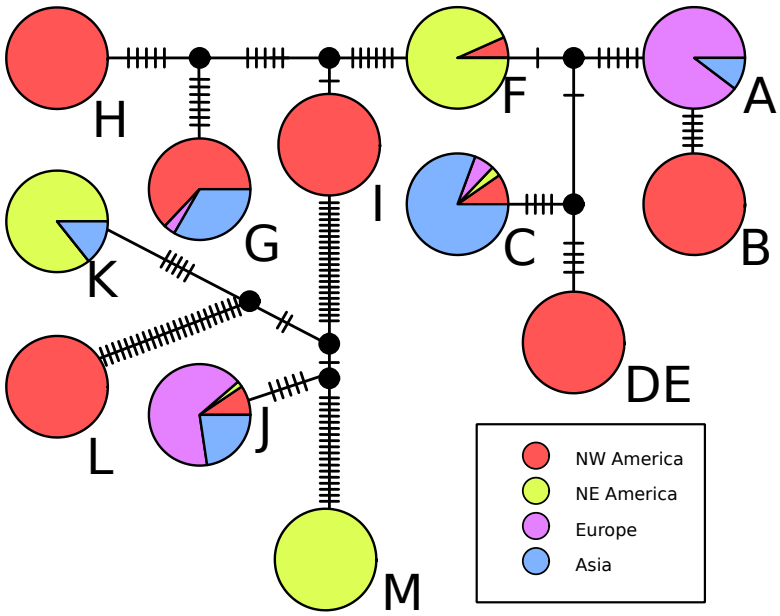

Figure 7.2 – A median joining lineages network of cytochrome oxidase subunit 1 (COI) gene of *Hyalosphenia papilio*. Each pie chart represents the proportion of the haplotype distribution between the four zones. Colour of the pie chart: red =North West of America, yellow = North East of America, Purple = Europe and blue = Asia.

only in North-West America and M only in North-East America), whereas others are cosmopolitan (e.g. C and J are found in all the North hemisphere).

Ancestral areas reconstruction

The BBM analysis suggests that the origin of *Hyalosphenia papilio* is situated in western North America (fig. 7.4). This fact/hypothesis is also supported by the maximum parsimony analysis. From there, it spread independently to the eastern North America, Europe and Asia. Furthermore, our data show that *H. papilio* spread from western to eastern North America at least twice in unrelated events. Clade J has a cosmopolitan distribution and the BBM analysis could not reliably infer the ancestral distribution of its most recent common

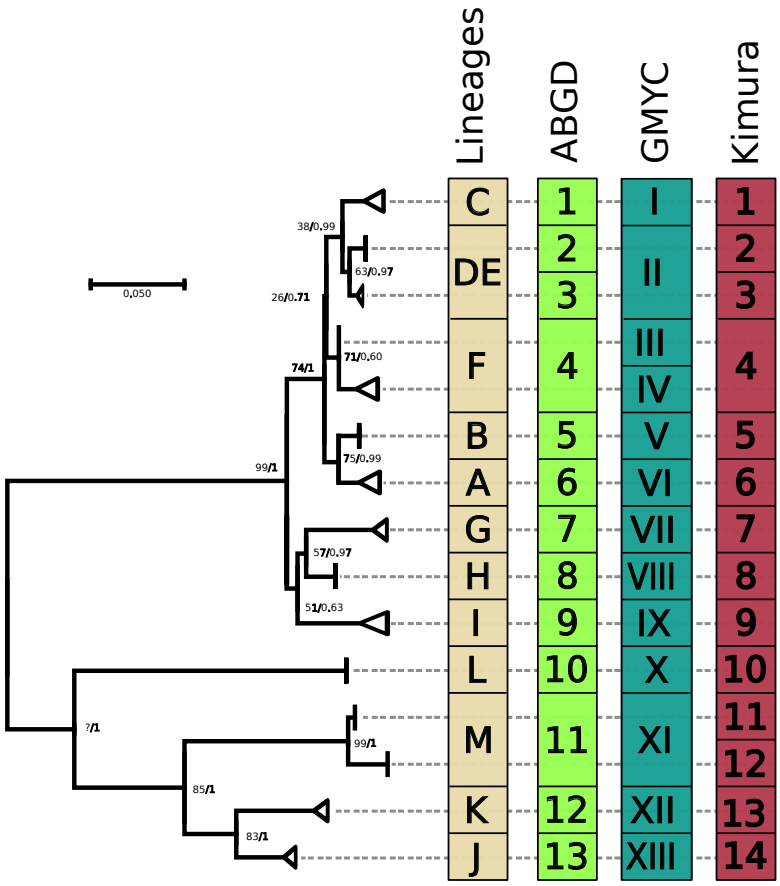

Figure 7.3 – Maximum Likelihood (ML) and Bayesian concatenated phylogenetic tree of 525 nucleotides from 76 unique sequences of *Hyalosphenia papilio*. Numbers along the branches represent respectively, bootstraps obtained by ML method and posterior probabilities as calculated with Bayesian analyses. The tree also represents the different lineages associated to the tips of the branches and the analyses used for their delineation (i.e. ABGD and GMYC analyses as well as Kimura test)

ancestor, which could have been living in either Europe or Asia. This lineage may have then propagated from there to the whole Northern Hemisphere.

Discussion

Our extended inventory of *Hyalosphenia papilio* COI diversity from Holarctic *Sphagnum* peatlands revealed the existence of at least, twelve lineages. In comparison with a previous study (Heger et al., 2013), we found a new lineage in Southern USA (Alabama), but in our analysis two lineages (namely D and E) that were presented as different in (Heger et al., 2013) were merged, probably due to the incorporation of new intermediate sequences in our study. These two taxa were already very closely related in (Heger et al., 2013), but it is unlikely that our newly defined lineages will fuse in a similar way, as all remaining lineages are separated now with higher genetic distances (i.e. at least 3%). These genetic distances, as determined by our species delineating approaches, coincide with the barcoding gap used to discriminate species in other Hyalospheniidae (Kosakyan et al., 2012, 2013). This suggests that they may be considered as true species, although more observations (morphology, ecological preferences) are needed to confirm this.

Our study suggest the idea that *Hyalosphenia papilio* lineages were not randomly distributed across the globe. Instead, most (6) were present in only one of the defined zones, and some presented narrow distributions. For example, lineage I was encountered so far only on the Western Coast of North America. On the other hand, four lineages showed wide distributions, covering sometimes all four geographic zones (e.g. lineage A). This case is particularly relevant, as it is found in regions with relatively mild climate and long growing season (e.g. NW Spain) as well as under the subarctic climate of Hudson Bay (Canada/CA). Likewise, the equally widespread lineages C and J can be found both under benign and harsh climates (fig. 7.5). Therefore, widespread lineages can thrive under very different climatic conditions. This suggests that climate alone is not the main driver of diversity in the *H. papilio* complex. The existence of lineages with drastically different distribution ranges suggests that certain lineages may have physiological adaptations that facilitate their fast transportation through long distances or establishment in contrasted climates. Resistant cysts can be the key to long distance transportation (Kristiansen, 1996). Alternatively, these strains may exhibit higher tolerance to environmental stress (e.g. cold temperatures), which increases their chances to colonize new suitable environments. In any case, these physiological/lifestyle adaptations seem to have appeared at least four times independently in the history of the *H. papilio* complex (fig. 7.4). Furthermore, it can be observed that out of 76 total haplotypes,

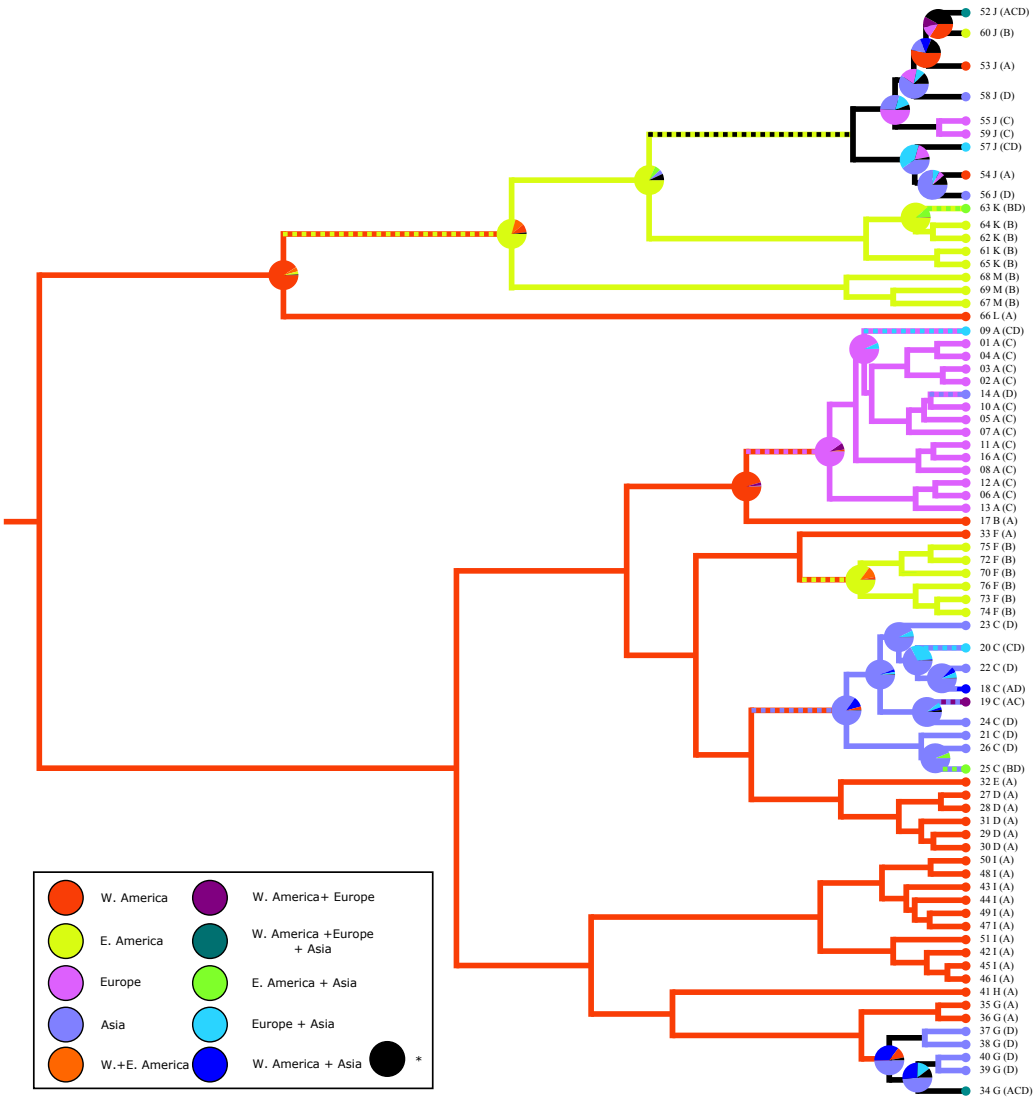

Figure 7.4 – Reconstruction of the past distributions of all known lineages of *Hyalosphenia papilio* using the BBM analysis as implemented in RASP. The tree was generated using BEAST v1.8.4. The tips represent extant haplotypes and their current distribution. The colours of the branches and the pie charts on the nodes represent ancestral distribution (colour code shown at the bottom of the figure). The probability of an ancestral distribution is indicated by the proportion of the pie chart covered by its colour. Pie charts are not shown when the most probable ancestral distribution, then indicated by the colour of the branches downstream, was supported at more than 95%. Dashed branches represent transitions between areas. Black branches represent uncertain ancestral distribution. The black colour in the pie charts (noted as "*" in the legend) represents other possible distributions lowly supported.

only seven (9.2%) were present in two zones, two in three zones (2.6%) and no single haplotype were found in all four zones. This shows that even widespread lineages show high infraspecific genetic structuring, which suggests limited gene flow among sites (Lara et al., 2011), and thus, geographical isolation.

Our phylogenetic reconstruction of the lineages evolution suggests Western North America as the most probable origin for the *H. papilio* complex. Ancestral trait reconstructions indicate with good confidence (98.5% of probability) that all most basal groups are to be found in this region. This type of approach has been used to identify the origin of multicellular taxa like hyacinthioid monocots (Ali et al., 2012) and chameleons (Tolley et al., 2013) and has been used here, to our knowledge, for the first time in eukaryotic micro-organisms. Although many papers now illustrate the importance of geographical distance on protist distributions (Heger et al., 2013; Pawlowski et al., 2007), there are only few phylogeographic studies documenting historical processes. Examples can be given in Foraminifera (Darling et al., 2007; Weiner et al., 2014), diatoms (Kooistra et al., 2005), and, in soils, Mycetozoa (Aguilar et al., 2013).

All lineages of *H. papilio* are strictly associated to peatlands, and appeared simultaneously with *Sphagnum* bogs during the Miocene, 5-20 Mya ago (Fiz-Palacios et al., 2014). A precise dating of lineage differentiation is impossible, because morphologies are subject to important phenotypic plasticity (Mulot et al., submitted.) and

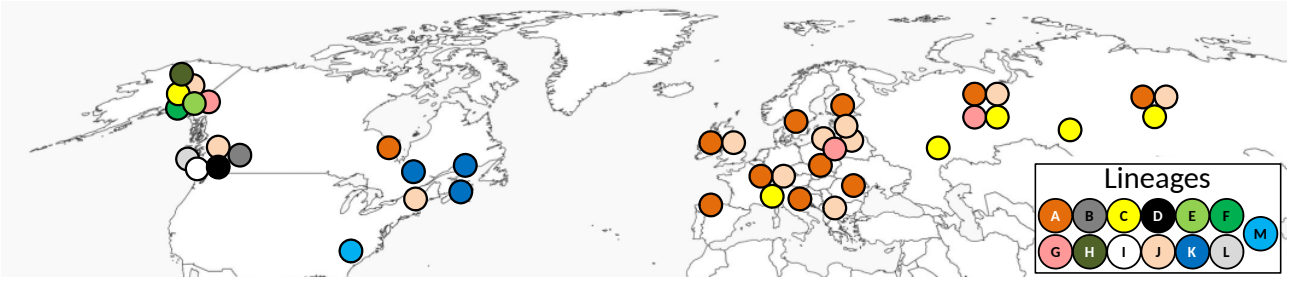

Figure 7.5 – Representation of *Hyalosphenia papilio* lineages distribution among the Northern hemisphere. Each circle corresponds to a site where the lineage has been detected, 9 lineages (B,D,E,F,H,I,K,L,M) showed restricted geographical distributions and were absent from Europe while 4 (A,C,G,J) were distributed all across the Holarctic realm

fossils tests, if encountered, cannot be assigned to the different lineages. However, it is noteworthy that important glacial refuges have existed in the Western North America during the Pliocene maximum, when the Cordilleran Ice Sheet was at its maximum extension (Hidy et al., 2013). The importance of glacial refugia in the historical processes leading to present day distributions has still not been evaluated with molecular methods. The fact that free-living protists show high cryptic diversity and that this diversity shows a geographical structuration opens the door to an in-depth investigation of the role of historical events such as Quaternary glaciation cycles in shaping current biogeographical patterns of microorganisms. Our results thus clearly show that the future of protistan phylogeography is bright and that these organisms can be invaluable witnesses of the factors that contributed in shaping today's biodiversity patterns. These evidences also increasingly demonstrate that it is highly likely that microbial diversity may also be threatened when regionally rare ecosystems such as *Sphagnum* peatlands in the southern margin of their distribution range are destroyed. Free-living microorganisms therefore also deserve to be considered in biodiversity conservation strategies.

Acknowledgement

This work was funded by the Swiss National Science Foundation projects no.310003A_143960. We would like also thanks Gustaf Granath and Hakan Rydin for the collaboration with the *Sphagnum* productivity project. Boris Droz, Christopher Niewoehner, Luca Bragazza, Richard J. Payne, Kosykh N. Pavlovna, Leonardo Fernandez, Koronatova N. Gennadievna, Irina Goia, Kai Vellak, Mariusz Lamentowicz, Mariusz Gałka, Katarzyna Kajukało for the field work. Maria Noskova for the Russian samples management. We would like to thanks also Amandine Pillonel for the laboratory work.

Chapter 8

Discussion: The "Sphagnosphere": A unique component of earth's Biosphere to assess microeukaryotes diversity, ecology and biogeography

Abstract

This discussion will take the shape of a review, where we will focus on the diversity, ecology and biogeography of microeukaryotes associated to the interstitial water of *Sphagnum* spp., the key builders of high latitude peatlands. The "sphagnosphere" is characterized by its stability (permanent high water content) and particularly selective conditions (low pH and nutrient content, biocidal effects of moss secondary compounds). These conditions strongly shape the composition of microbial communities, leading to highly specialized assemblages. They are also relatively homogeneous across different *Sphagnum* peatlands around the world. At the local scale however, the microtopography influences important factors such as water table depth and nitrogen content. Microeukaryotic communities are influenced at both global and local levels by, respectively, climate and geographic distances and biotic and abiotic factors. We propose the sphagnosphere as a powerful model to test ecological and biogeographical hypotheses.

Background

Because they play crucial roles in most of the world's ecosystems (see Introduction), the study of diversity and the interactions among microeukaryotes and their relationships with their environment is crucial to understand the functioning of natural systems. Plankton is certainly the best-studied system, because of organisms are easy to observe directly. The paradigm of the microbial loop (Azam et al., 1983), provided a framework to evaluate the nutrient fluxes among organism towards the higher trophic levels and also to ocean productivity. Subsequently, the diversity of planktonic protist communities became a fashionable research topic with the development of DNA-based environmental diversity surveys, and marine plankton was frame of the first pioneering studies (Diez et al., 2001; López-García et al., 2001; Moon-van der Staay et al., 2001). Other systems like soils (Lesaulnier et al., 2008) or *Sphagnum*-dominated peatlands (Lara et al., 2011) started to be studied much later, and presented also high levels of molecular diversity. With the advent of High-Throughput Sequencing (HTS) technology, the difficulties encountered in early studies to get rid of the many sequences corresponding to fungi which masked eukaryotic diversity (Lesaulnier et al., 2008) were not an issue anymore because of the huge amount of sequences retrieved. While extensive studies still revealed an unsuspected diversity in marine plankton (De Vargas et al., 2015), an equally important or perhaps even higher richness was also found in soils (Appendix E (Mahe et al., 2017) (Grossmann et al., 2016)). This technology should be promising and powerful to define bioindicators and develop biomonitoring tools (i.e. (Pawlowski et al., 2016) Appendix C (Seppey et al., 2016)). Nevertheless to our knowledge, the diversity associated to *Sphagnum* spp. had never been studied with HTS approach before this thesis.

Diversity of microeukaryotes in the Sphagnosphere

Microeukaryotes seem to be more diverse than their prokaryotic counterparts, at least in certain environments (De Vargas et al., 2015; Grossmann et al., 2016). However, while the number of OTUs retrieved in different studies can be more or less compared, functional diversity still remains difficult to evaluate. Indeed, new species with non-canonical life cycles are commonly isolated and described (Appendix D, (Blandenier et al., 2016; Tice et al., 2016)). New taxa are very often discovered in understudied and extreme habitats (Appendix C (Heesch et al., 2016; Seppey et al., 2016; Yubuki et al., 2009)) The sphagnosphere (Cheema et al.,

2012), due to its particular and very selective conditions can be considered as an extreme environment for microeukaryotes (Lara et al., 2011) and hosts also particular species which are found in other environment. For example, testate amoebae encountered in these environments comprise a majority of forms which are restricted to the sphagnosphere, such as *Hyalosphenia papilio, Planocarina carinata, Amphitrema wrightianum, Archerella flavum, Euglypha penardi, Placocista spinosa* or the species described in this thesis, *Nebela gimlii* (Chapter 3 (Singer et al., 2015)), plus many others. From the functional point of view, they comprise several mixotrophic and entirely heterotrophic species, small bacterivores and large eukaryvorous forms. In spite of the strongly selective conditions of the sphagnosphere, these organisms are highly diversified, both from the phylogenetic and the functional points of view. At the specific level, the diversity is even higher as most testate amoeba "morphospecies" actually hide a high level of cryptic diversity, i.e. species that look superficially like each other but are genetically different. Almost every morphologically described testate amoeba species investigated to date in detail hosts a large cryptic diversity, both in Arcellinida (Kosakyan et al., 2012) and in Euglyphida (Heger et al., 2011a). Investigations combining Cytochrome oxidase subunit I data (COI) from barcoded single cells and detailed microscopic pictures (light and electronic) allowed the delimitation of eight different species in the complex of *Nebela collaris* ((Kosakyan et al., 2013), now called genus Nebela (Kosakyan et al., 2016) Chapter 3: (Singer et al., 2015)). Testate amoebae are large eukaryotes that are particularly easy to observe; other groups of conspicuous protists with members infeodated to peatlands, and which can be particularly diverse are the ciliates, centrohelids and euglenids (Kreutz and Foissner, 2006). However, for eukaryotes that need to be cultured to be properly identified, like small heterotrophic flagellates and naked amoebae for instance, diversity estimations based on microscopic observations is impossible. Indeed, very few peatland protists can be cultured (Ralf Meisterfeld, person. comm.). The reason is most probably the particular chemical composition of the water present in the sphagnosphere, which is impossible to recreate in vitro. For all inconspicuous forms, only DNA-based studies can provide the key to their diversity. This does not mean that *Sphagnum* associated species are not diversified; we found in (Chapter 5 (Singer et al., 2016)) many previously undetected basal oomycetes (a group of fungi-like stramenopiles containing many parasites and mostly inconspicuous forms because of their mycelial growth). It is interesting to notice that we observed more oomycete diversity in five peat bogs than in the whole TARA study that covered dozens of sites and even more plankton samples (De Vargas et al., 2015).

Factors influencing micro-eukaryotic diversity at the local scale

As documented in Chapter 3: (Singer et al., 2015), *Nebela gimlii* seemed to be more abundant in peatland forest or margins than its sister species *Nebela guttata*, which occurred preferentially in the most oligotrophic part of the peatland (hummocks and lawns). In Chapter 4, we demonstrate that the different members of genus *Nebela* are not homogeneously distributed across peatland microhabitats but are strongly associated with the two principal environmental gradients existing within a peatland: water table depth (Marcisz et al., 2014b) and nitrogen content (Lin et al., 2012). Moreover, in this genus, we observed phylogenetic clustering in the most oligotrophic microhabitats of the peatland, suggesting a phylogenetic inheritance of adaptation to nitrogen-poor conditions. Therefore, differences in the physicochemical parameters which characterize the different microhabitats of the peatlands change community composition. A side effect of these conclusions is that that closely related species of testate amoebae with similar morphology may indicate different environmental conditions. The application of DNA-based methods to bioindication would circumvent the problems inherent with morphological species identification, as models can be expected to lose explanatory power when superficially similar species are systematically pooled (Amesbury et al., 2016; Charman et al., 2000).

HTS approaches also show inhomogeneous distribution in oomycete communities, pools hosting the highest diversity (Chapter 5: (Singer et al., 2016)). The most diverse microhabitats are thus different than for genus *Nebela*. Several causes can be implied here. Possibly different taxa may be influenced by selective pressures such as the two above-mentioned variables (water table depth and nitrogen content). Secondly, their flagellated propagules ("zoospores") are most easily propagated within an aquatic environment, and probably fully aquatic environments are less selective to oomycetes. Finally, as many species are parasites of plants, animals and even fungi (Lara and Belbahri, 2011), they may follow the distribution of their host species. Indeed, some OTUs representing exclusively parasitic organisms have been shown in Chapter 4. For different reasons, which depend on the lifestyle and physiology of the different groups of eukaryotes considered, communities change along the peatlands microhabitats. Abiotic factors such as water table depth and nitrogen content influence the communities in the two studied groups. Biotic factors are also strongly suspected, like parasitic relationships with different organisms as they seem to happen in oomycetes, but also prey availability or competition, even though this interaction seems to be marginal at most in genus *Nebela* (Chapter 3).

Factors influencing micro-eukaryotic diversity at the global scale

At a global scale, climate plays an important role in shaping the composition of microeukaryotic communities. Temperature has been shown to influence deeply not only the taxonomic composition but also the functional types present. A study under controlled conditions in a peatland (Open Top Chamber experiment) has demonstrated that an increase of temperatures will increase the decomposition rates (Jassey, 2011). Moreover, the proportion of heterotrophic testate amoebae increases under warmer conditions while the proportion of mixotrophic species declines (Jassey et al., 2015). If mixotrophy is not a winning strategy under high temperatures, it can be expected that, in natural communities, the proportion of mixotrophs decreases. We surveyed the diversity of microeukaryotes in *Sphagnum* shoots sampled under three different climatic zones (temperate, subtropical and temperate) using an environmental DNA survey approach (Chapter 6). We determined the proportion of mixotrophs by identifying taxonomically OTUs and assigning them a functional role based on the literature. As expected, the relative amount of sequences attributed to mixotrophs was higher in the temperate systems (and was logically inversely correlated with temperature in whole study), whereas osmotrophs (=decomposers) were more abundant in the tropical zone (Chapter 6). These results were in phase with observations on testate amoebae under controlled conditions, and corroborate the hypothesis that mixotrophy is favoured under cold climates. Carbon fixation through mixotrophic micro-organisms is significant at the ecosystem level (Jassey et al., 2016). If the increased amount of decomposers observed under warm climates indeed corresponds to increased decomposition rates, it can be expected that temperature plays a capital role in the carbon balance in peat bogs. This hypothesis still needs to be tested experimentally, but can have deep implications for the world's climate in a global warming context, as *Sphagnum*-dominated peatlands contain one third of global soil carbon pool (Yu et al., 2011). Climate therefore plays a very important role in the composition of the sphagnosphere's micro-eukaryotic communities, and influences not only community composition in terms of species (or OTUs) but also in terms of functions, which has probably implications on the functioning of the whole ecosystem.

This latter study has shown also that only 25 percent of the OTUs were shared between the three considered climatic zones. As hypothesized previously, climate is supposed to play a large role in the structure of the communities, but we suspect that geographical distance plays also a role in the observed beta-diversity. Increasing evidence has been disproving the "everything is everywhere" hypothesis in vogue at the turn of the century (see (Finlay, 2002), and Intruduction), although it got some support in some marine organisms that spread by tiny planktonic propagules (Pawlowski et al., 2007; Šlapeta et al., 2006). However, in the majority of the cases, evidence from different clades and environments tended to show the opposite (Evans et al., 2009; Lara et al., 2016).

The relative physicochemical homogeneity of the sphagnosphere (in comparison with soils or freshwaters for instance) and its strong environmental filters renders it an excellent model to test the effect of geographical distance on microeukaryotic communities. Heger and coauthors (Heger et al., 2013) defined twelve different lineages of *Hyalosphenia papilio*, a key species of testate amoeba in Northern Hemisphere peatlands and analyzed their respective patterns of distribution. The study demonstrated that climate explained a major part of the distribution of these lineages, but also showed that geographic distance explained a significant part. Inspired by this study, we completed their sampling effort (Chapter 7) and confirmed that genetic diversity was not randomly distributed in the northern hemisphere, both at infraspecific (Haplotypes) and lineage levels. Nine lineages showed narrow geographical distributions while four of them were present across the entire Holartic realm, suggesting divergent dispersal capacities in the different lineages. This fit well the "moderate endemicity" model proposed as an alternative to the "everything is everywhere" hypothesis (Foissner, 2006). In support to a limited dispersal capacity, we showed using ancestral character reconstructions that the *Hyalosphenia papilio* complex probably appeared somewhere in the West Coast of North America (Chapter 7).

Conclusions

The living part of *Sphagnum* mosses (i.e. upper part of the stems, aerobic) represents the interface between the atmosphere and the zone where peat accumulates (dead organic matter) (Titus and Wagner, 1984). The incorporation of carbon through photosynthetic activity of mosses and micro-organisms will happen in this region; thus, organisms living there will contribute actively to the stability of the system (Jassey, 2011). This balance will be strongly affected by the environmental conditions at both local and global scale. At a local scale peat harvesting or the drainage for agriculture (Chapman et al., 2003) will reduce significantly water table depth resulting in changes in the microeukaryotic communities composition as detailed previously. This is correlated with an increase in the decomposition rate and thus in the release of CO_2 (Joosten, 2009; Silvola et al., 1996). Peatlands are also exposed to indirect human impacts such as atmospheric deposition of nitrogen, which will also be correlated with carbon loss (Bragazza et al., 2006) and, at long term, promote the tree colonization

(Hedwall et al., 2017). Further experimentations including diversity surveys of eukaryotes but also bacteria and archaea should be planned to demonstrate a causality between these observations and suggest possible explanatory mechanisms. At a global scale, increase in temperatures do influence peatlands functioning, and it has been proposed that these ecosystems may shift from carbon sinks to sources in a global warming context (Dise, 2009; Moore and Dalva, 1993). A possible explanation may be found at the microeukaryotic communities levels as suggested in this thesis. And, finally, the differences in community composition and diversity in different sphagnospheres in the world (due to dispersal limitation) may mirror different webs of biotic interactions, and thus, different sensitivities of the ecosystems. It can be expected therefore that peatlands have less resilience in some parts of the world than in others.

The communities of the sphagnosphere play an immense role in the maintenance of the biosphere because of their direct and indirect involvement in peatland carbon uptake. The study of the causes and consequences of these shifts in diversity need to include analyses of the entire communities. We have now the technical tools to characterize this diversity, but need to acquire more information from classical protistology (strain isolation and description, plus barcoding) to interpret HTS data. We need also to sample regions that remain unexplored, especially the Southern Hemisphere, which has been neglected so far, in order to generalize our observations based mainly on Holarctic systems. And, finally, we need more experimental research in order to validate our hypotheses on the functioning of peatlands.

Chapter 9

Acknowledgments

A PhD thesis is not the work of a single person! With this respect, I would like first to thank the 45 scientists who contributed actively to the elaboration of the present manuscript!

J'aimerais en premier lieu remercier Enrique Lara sans qui toute cette aventure n'aurait pas été possible. Tu m'as fait confiance et introduit dans cet univers merveilleux alors que je n'y connaissais strictement rien! J'ai appris énormément avec toi, et j'espère que nous pourrons encore collaborer ensemble pendant de longues années! Merci pour ta précieuse amitié!

Je souhaite ensuite remercier Edward Mitchell! Tu as réussi à réunir dans ton labo un panel de gens hétéroclites bossant sur des sujets aussi variés tel que les traits des amibes, les bio-indicateurs de cadavres, et les microorganismes vivant dans les Sphaignes! Et ça marche! Merci pour cette émulsion scientifique au jour le jour et pour tous les conseils que tu m'as donnés pendant ces quatre années!

Je souhaite aussi remercier chaleureusement l'université de Neuchâtel et le fonds national pour la recherche suisse! Et particulièrement tous les membres du laboratoire de Biodiversité des sols! Christophe pour toutes ces discussions de stats et de Bioinfo. Amandine pour toute l'aide de labo. Anush for showing me the molecular works. Quentin et Clément, pour m'avoir aidé à la rédaction de cette thèse (surtout sur la fin!). Isa, Matthieu, Léo, Maj, Thierry, Lassaâd, Alex, Nico et Ildiko et tous les autres qui sont passé dans le labo pour tous les moments incroyables qu'on a vécu ensemble! Merci aussi à tous les collègues de Neuchâtel! A toute l'équipe du LEF et notamment Claire pour l'aide apportée tout au long de cette thèse. Le service informatique avec Romain Bessire, le service d'impression, le secrétariat ainsi que Christiane Bobillier sans qui la machine n'aurait pas été aussi bien « huilée » !

I would like also thanks my colleagues from other universities! We had some great time together during meetings and conferences and the different projects that we realized! The polish team with Mariusz Lamentowicz, Anna Basińska, Katarzyna Marcisz, Katarzyna Kajukało, Monika Katarzyna Reczuga, and also Jan Pawlowski, Valentyna Krashevska, Kenneth Dumack, Stefan Geisen and finally Daniel Lahr (I am looking forward working with you soon!).

Pour pouvoir réaliser un travail tel que cette thèse il est primordial d'avoir des amis sur lesquels on puisse compter afin de prendre du recul par rapport à la science!

Je tiens tout particulièrement à remercier Camille et Loïk pour m'avoir fait voyager toutes ces années! L'équipe du Greenfield (mais pas que) pour toutes les noces que l'on a faite ensemble : Romain, Stéphanie, Alex, John et Jen ! Les amis du Lycée : Valérie et Pierre ! Les Fishs Loïc, Jonas, Johan, Mirco, Damien, Jenny, Emilien : cette thèse n'aurait pas été la même sans vous et tous ces merveilleux concerts! Je souhaite aussi remercier sincèrement Mélissa pour m'avoir fait entrer dans un monde inconnu et incroyable! L'association INSIEME, Delphine Vaucher, et ma co-responsable préférée Mélanie! Et le top des monos avec, Aurélie, Damien, Chloé et Léa et tous les autres!

Then, I would also like to thank the members of the jury Thomas Posch, Diego Fontaneto, Micah Dunthorn and Sergio Rasmann for reading this huge manuscript!

Pour finir je souhaite dédicacer cette thèse à mes parents et mon frère ! Ils m'ont toujours soutenu et poussé dans mes choix ! Être scientifique n'est pas une vie des plus faciles, donc si j'en suis là aujourd'hui c'est grâce à vous ! Merci !


Jan Pawlowski
Emilien
Mirco
Alexandre
Kenneth Dumack
Daniel Lahr
Matthieu
Frédéric Mahé
Delphine
Vaucher
Elianne Egge
Ildikò
Colomban de Vargas
Micah Dunthorn
Lassâad
Katarzyna M
Jeni
Leonardo
Gabriela González Garraza
Thierry
Rainer Wirth
Gabriela Mataloni
Diego Fontaneto
Johnathan
Christelle
Gabriel Trueba
Chloé et Léa
Anaële
Isabelle
Valentyna Krashevska
Mariusz
Edward
Michèle Vimant
Sebastian
Romain
Valérie
MAJ
Johan
Damien
Fernando Unrein
Maman
Thomas Posch
Alexandros Stamatakis
Maria Victoria Quiroga
Irina Izaguirre
Ramon Massana
Jonas
Quentin
Camille
Luc
Monika
Alexey Kozlov
Isabelle Trautmann
Sarah Romac
Lucas Czech
Pierre
Tomasz Oszako
Tobias Siemensmeyer
Jordan Mayor
Enrique
Clément
Bertrand
Schiaffino Romina
Kenneth
Loïk
Mélanie
Cédric Berney
Loïc
Nicolas
Mélissa
Anna
Insieme
Amandine
Noelia Paredes
Guillaume Lentendu
Claire
David Bass
Papa
Alex
Katarzyna K
John Bunge
Romain Bessire
Lara
Christophe
Sarah Sernaker
Mónica Steciow
Christiane Bobtllier
Sréphanie
Anush
Vanessa Balagué

Appendix A

Planktonic eukaryote molecular diversity: discrimination of minerotrophic and ombrotrophic peatland pools in Tierra del Fuego (Argentina)

Enrique Lara[1], Christophe V.W. Seppey[1], Gabriela González Garraza[2], David Singer[1], Maria Victoria Quiroga[3], Gabriela Mataloni[2]

[1]Laboratoire de Biologie du sol, Département de Biologie, Université de Neuchâtel, Neuchâtel, Switzerland
[2]Grupo de Biodiversidad, Limnología y Biología de la Conservación, 3IA – Instituto de Investigación e Ingeniería Ambiental, Universidad Nacional de San Martín, Buenos Aires, Argentina
[3]Laboratorio de Ecología y Fotobiología Acuática, Instituto de Investigaciones Biotecnológicas–Instituto Tecnológico de Chascomús (IIB-INTECH), Unsam-Conicet, Argentina

Journal of Plankton Research, Vol. 37(3), p.1-11, May 2015
doi: 10.1093/plankt/fbv016

Abstract

We investigated the composition of the smallest size fraction (>3 µm) of eukaryotic plankton communities of five pools located in the Rancho Hambre peat bog in Argentinean Tierra del Fuego with an Illumina HiSeq massive sequencing approach applied to the v9 region of the eukaryotic SSU rRNA gene. Communities were generally dominated by chrysophytes, with a good representation of Perkinsea and Cercozoa clade NC-10. A community composition analysis performed using GUniFraC separated minerotrophic and ombrotrophic sites, reflecting perfectly the classification of the sites based on environmental data. However, this separation disappeared when more weight was given to abundant phylotypes, suggesting that subordinate phylotypes were responsible for site discrimination. The 5% best indicators for, respectively, minerotrophic and ombrotrophic environments were searched using an IndVal analysis. Among these, autotrophic taxa were more common in minerotrophic environments, whereas mixotrophic taxa represented best ombrotrophic water bodies. However, the ecological traits of many taxa have still not been determined, and still needs to be investigated for a better understanding of freshwater systems ecology.

Keywords: unknown diversity; next-generation sequencing; community; mixotroph; protist; algae; parasitoid

Introduction

Traditional limnological studies on microbial eukaryotic plankton were first restricted to net-sized taxa, and then to those that were identifiable with light microscopy, typically algae and ciliates. Heterotrophic nanoflagellates and small autotrophs/mixotrophs are known to exhibit an immense diversity, yet they were typically pooled together into a small number of genera, as these organisms lack morphologically discriminating traits. However, it has been demonstrated that genera such as *Chlorella* (Huss et al., 1999; Krienitz et al., 2004)), *Spumella* and *Ochromonas* (Boenigk et al., 2005a; Cavalier-Smith and Chao, 2006) include genetically and functionally diverging forms. Also parasitic and parasitoid taxa can only be observed when they protrude outside their hosts, like some life stages in chytrids. Therefore, diversity of planktonic communities has been largely underestimated.

The last decade saw the advent of DNA-based studies to characterize environmental eukaryotic diversity. These approaches revealed an immense diversity in freshwater systems (Richards et al., 2005; Slapeta et al., 2005), showing also the existence of many previously unknown deep clades. In the last 5 years, the development of massive sequencing technologies such as pyrosequencing or Illumina allowed a more in-depth picture of existing diversity by providing thousands to millions of reads representing arguably the whole extent of eukaryotic diversity present in one sample (Amaral-Zettler et al., 2009; Behnke et al., 2011). Indeed, the application of so-called next-generation sequencing (NGS) technologies indeed brought a better insight on diversity, bridging sequence and observation data is still not straightforward (Bachy et al., 2013; Stoeck et al., 2014). Beside the intrinsic interest of estimating total environmental diversity, there is a need to relate sequence data to a specific niche or role in the food web. Consequently, as large amounts of unknown sequences are revealed, it can be expected that our vision of freshwater ecosystem functioning may be challenged. For instance, the discovery of a wealth of sequences related to parasitoids in lakes suggests the importance of these organisms in regulating populations and nutrient cycling (Brate et al., 2010; Mangot et al., 2011). Beyond this example, trophic strategies of organisms (i.e. autotrophy/heterotrophy etc.) can be inferred in some cases by a careful examination of the taxonomic position of the organisms from which sequences derived. Also, successful adaptive strategies can be deducted from indicator OTUs for different sets of environmental conditions. These facts highlight not only the existence of unknown organisms, but also of unsuspected mechanisms ruling nutrient cycling in freshwaters (Sime-Ngando and Niquil, 2011).

Peatlands are wetland ecosystems characterized by the accumulation of slowly decomposing organic matter (peat) mostly under cold, wet and anoxic conditions. The areas where peat is actively formed, termed "peat bogs" are frequently dominated by the moss *Sphagnum magellanicum* in Tierra del Fuego and host often shallow, acidic, humic pools, which in turn can display a range of abiotic features. Five pools with different morphometric characteristics from Rancho Hambre peat bog were thoroughly studied over two consecutive ice-free periods (October–April) between 2008 and 2010 regarding their physical and chemical features (Gonzalez Garraza et al., 2012) as well as the variations in abundance, biomass and composition of their plankton communities (Quiroga et al., 2013). Due to the impossibility of identifying morphologically the smallest fraction of organisms, their composition remained unknown until now. Here, we explored the environmental molecular diversity of these small (i.e. >3 µm) eukaryotes. We sequenced the v9 region of the SSU rRNA gene using Illumina's HiSeq technology, and related community composition with physico-chemical parameters. We hypothesized that the minerotrophic (i.e. influenced by water of subterranean origin) versus ombrotrophic (i.e. exclusively fed by precipitation) character of water bodies would largely determine the composition of the smallest eukaryotes, as

this factor proved crucial for structuring trophic webs in these environments (Quiroga et al., 2013). Furthermore, we characterized taxonomically and determined the trophic strategies of the best indicator organisms for each water body type.

Method

Sampling

Five pools located within Rancho Hambre peat bog (RH; 54°44'52.87"S 67°49'29.44"W), undergoing a long-term ecological survey (fig. A.1) were sampled on 1 November 2012 during the austral late spring. They have been labeled RH1-5 and include two minerotrophic pools connected to the underlying aquifer and three ombrotrophic ones (fed only by rainwater); they have been described in detail in (González Garraza, 2012). On account of the homogeneous physical and chemical properties of each pool, one series of samples was taken from one point at the shore in each of them. Samples for chemical and biological analysis were collected using 2-L acid-washed plastic bottles pre-rinsed with pool water and then transported to the laboratory under cold, dark conditions. Temperature, pH and conductivity were measured *in situ* with a multiparametric probe, HachSension 156 (Hach, USA).

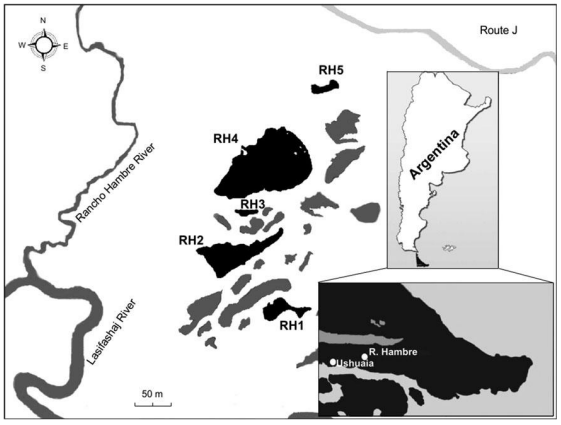

Figure A.1 – Map representing the respective positions of the different water bodies and their geographic location (Mataloni et al., 2015).

Lab methods

Dissolved reactive phosphorus (DRP), total phosphorus (TP) ammonia (NH_4-N), nitrate (NO_3-N), absorbance at 440 nm (A440) and total hardness (TH) were determined as described in (González Garraza, 2012). Dissolved inorganic nitrogen was the sum of NH_4-H + NO_3-N. Dissolved organic carbon (DOC) was determined from filtered water with the high-temperature Pt catalyst oxidation method (TOC-L, Shimadzu) following (Sharp et al., 1993).

DNA extraction, PCR and sequencing

In the laboratory, 20-µm-net-filtered plankton samples were prefiltered through 3-µm pore-size polycarbonate filters to exclude as many larger organisms as possible. The filtrate was deposited onto 0.2-µm nitrocellulose filters. These were placed in cryovials and stored at -80°C until analysis. Environmental DNA was extracted by cutting nitrocellulose filters into ∼1 mm^2 pieces which were subsequently transferred into the columns of a MoBio (Carlsbad, USA) Power Soil^{tm} extraction kit directly as recommended in Lara et al. (Lara et al., 2009b). A PCR protocol aimed at amplifying the v9 region of the SSU rRNA gene of eukaryotes present in the samples after the recommendations of (Amaral-Zettler et al., 2009), using Promega's GoTaq polymerase (without proofreading activity). Sequencing was performed with Illumina's HiSeq technology, using V3 chemistry (Fasteris, Geneva, Switzerland).

Sequences treatment

Sequences were sorted for quality by keeping only sequences without ambiguous nucleotides based on a phred score threshold of <28 (with a custom script). Sequences were then clustered into OTUs using the dbc454 program (Pagni et al., 2013). The following setups were used: minimum number of sequence by OTU: 5; distance cutoff: 1 – 7 and step: 0.2. OTUs were aligned against sequences from the PR^2 database as downloaded on 24 June 2014 (Guillou et al., 2013) to determine their taxonomic affiliation using SWIPE (Rognes, 2011) with the following parameters: reward: 1; penalty: 23; gap open: 2 and gap extend: 3. The OTUs assigned to Metazoa and Embryophyceae were discarded for further analysis. In order to remove rare clusters in an objective way, we determined the inflection point of the OTUs rank abundance curve from the combined data and removed all sequences that were less abundant; that threshold was found with a piecewise linear regression (Chiu et al., 2006; Toms and Lesperance, 2003). We verified that the diversity reached saturation in each sample with the function "rarecurve" as implemented in the vegan package (Oksanen et al., 2012) in R.

Once assigned, the frequency of each OTU was established for every sample. The taxonomic assignations of OTUs were then verified individually against the NCBI database. A functional category (i.e. trophic mode) was assigned to OTUs whose sequences branched within a group in which all members share the aforementioned trophic mode with a custom script.

Numerical analysis

Pie charts were built to show the gross taxonomic composition encountered in each sample (fig. A.5). Non-correlated physicochemical parameters were selected for the remaining analyses. The parameters were selected based on correlation tests done for every pair of two parameters. We removed iteratively the most significant parameter until the lower P-value was higher than 0.05. These were pH, conductivity, DRP, TP, TH and DOC. An unsupervised random forest analysis (1 million trees) was used to classify the five water bodies according to the selected physico-chemical parameters (Breiman, 2001). A dendrogram was computed on the basis of the distance between samples (fig. A.2).

In parallel, samples were also classified according to the composition of their communities using GUniFrac v.1.0 (Chen et al., 2012). The parameter α was set to vary between 0 and 1. This parameter varies between a presence – absence analysis of communities ($\alpha = 0$) to giving each OTU the weight corresponding to the number of occurrences in the set of sequences, which corresponds to taking into account only the most common sequences. The phylogenetic tree required for the GUniFrac analysis was built with ExaMLv 2.0.4 (Stamatakis, 2014) based on an alignment obtained from the program Clustal Omega (Sievers et al., 2011), with a gamma distribution of rate heterogeneities. A dendrogram was built for eight values of α, namely 0, 0.14, 0.29, 0.43, 0.57, 0.71, 0.86 and 1.

For each group of samples discriminated by dendrograms, based on both physicochemical parameters and community composition, we calculated indicator values for each OTU using the IndVal method (Dufrene and Legendre, 1997). This method determines the most characteristic organisms for each environment and evaluates their specificity with a score ranging from 0 to 1, 1 being an organism present only in a given environment and totally absent from the others. The 5% most characteristic organisms (39 OTUs), i.e. the OTUs that had the highest indicator score were selected and their abundance was represented in a heat map. The color represents the proportion of these OTUs in the different samples.

Results

Physical and chemical characterization of water bodies

Field observations showed that among the three large water bodies, only RH1 and RH4 had inflows and/or outflows (i.e. natural surface channels), while RH2 as well as shallow RH3 and RH5, did not and received water only through rainfall and snow. In November 2012 sampling, the water temperature of the five pools showed intermediate values regarding their respective ranges, reflecting late spring-early summer conditions (table A.1). Values of conductivity, TH, nutrient concentrations and organic matter were within the same range in all water bodies, and fell within the previously established variation range for these environments. On the other hand, pH and DOC clustered the pools into two groups: less acid and humic (RH1 and RH4, pH = 6.7 and 6.8, DOC = 0.42 and 0.33, respectively) and more acid and humic (RH2, RH3 and RH5, pH = 5.9, 5.1 and 5.8, DOC = 0.56, 0.63 and 0.61, respectively), showing the minerotrophic character of the first group of pools and the ombrotrophy of the second.

Taxonomic composition of communities

Environmental DNA survey gave a total of 732 109 sequence reads for all five (RH1-5) samples. From these, 205 274 sequences were kept after quality check. Sequence clustering resulted in 3291 different OTUs. From these, we removed 94 which belonged to Embryophyceae or Metazoa. After removing the rare sequences, we kept 783 OTUs for statistical analysis. Sequences related to Chrysophyceae were the most abundant in all samples with the exception of RH2 and, overall, represented the most common group. The parasitoid clade Perkinsea dominated RH2 and was present in all waterbodies, although in lower abundances. Phagotrophic taxa were represented by ciliates and bicosoecids. Cercozoa, which belonged almost exclusively to the environmental clade 10 (Bass et al., 2009a) were also well represented, reaching high proportions in the total number of reads in samples RH1, RH4 and RH5. Autotrophs were mostly represented by chlorophytes and bolidophyceae and many sequences of osmotrophic organisms (mostly Fungi) were also present (fig. A.5).

Table A.1 – Morphometric and physicochemical features of the five pools from Rancho Hambre peat bog (Tierra del Fuego)

Pools	RH1	RH2	RH3	RH4	RH5
Latitude (S)	54°44"52.87'	54°44'48.61"	54°44'46.75"	54°44'41.51"	54°44'39.35"
Longitude (W)	67°49'29.44"	67°49'31.66"	67°49'32.21"	67°49'31.69"	67°49'26.70"
Maximum length (m)	81.9	162.9	50.7	195.7	34.5
Maximum width (m)	28.5	66.2	10.5	122.9	12.7
Maximum depth (cm)	127	165	35	150	33
Perimeter (m)	238	445	115	555	162
Area (m^2)	1824	5976	137	16190	542
SDI	1.6	1.6	2.1	1.2	2
Temperature (°C)	10.4 (2.3-17.5)	9.5 (1.1-15.9)	9.5 (3.2-24.9)	9.6 (3.3-14.9)	10.4 (1.7-19.7)
pH	6.7 (5.0-7.1)	5.9 (3.8-5.9)	5.1 (3.6-5.4)	6.8 (5.2-7.0)	5.8 (4.1-5.8)
Conductivity (µS cm^{-1})	16 (14-50)	21 (9-40)	13 (10-82)	20 (16-60)	16 (5-50)
TH (mg equiv. $CaCO_3$ L^{-1})	38 (7-41)	32 (7-46)	48 (8-48)	26 (11-43)	35 (11-36)
DIN (µM)	1.6 (0.5-7.3)	1.6 (0.5-17.1)	1.6 (0.7-7.4)	1.5 (1.4-7.6)	2.4 (0.0-5.2)
TP (µM)	9.9 (3.7-9.9)	9.2 (3.0-10.7)	11 (2.9-11.0)	11 (2.9-11.0)	13.9 (2.5-13.9)
DRP (µM)	1.3 (0.8-2.7)	1 (0.7-2.5)	0.7 (0.7-4.2)	0.7 (0.5-1.9)	1.9 (0.6-1.9)
DOC (µM)	0.42 (0.42-0.80)	0.56 (0.42-0.75)	0.63 (0.23-1.22)	0.33 (0.33-0.50)	0.61 (0.32-1.04)

November 2012 values are given, with minimum and maximum values recorded from October 2008 to April 2010 in parentheses (González Garraza, 2012) TH, total hardness (Ca^{2+} + Mg^{2+}); TN, total nitrogen; DIN, dissolved inorganic nitrogen; TP, total phosphorus; DRP, dissolved reactive phosphorus; DOC, dissolved organic carbon

Numerical analysis

Random forest analysis based on environmental data grouped on one hand ombrotrophic samples RH2, 3 and 5, RH3 and 5 being the most similar and, on the other hand, minerotrophic RH1 and RH4. The GuniFrac analysis on community data and resulting dendrogram showed the same pattern when parameter α varied from 0 to ∼0.4, corresponding to a situation where a relatively lower weight is given to abundant OTUs. Higher values showed a completely different topology where ombrotrophic and minerotrophic communities are intermixed, α=1 corresponding to the situation where weight of OTUs corresponds exactly to numbers of reads (fig. A.2).

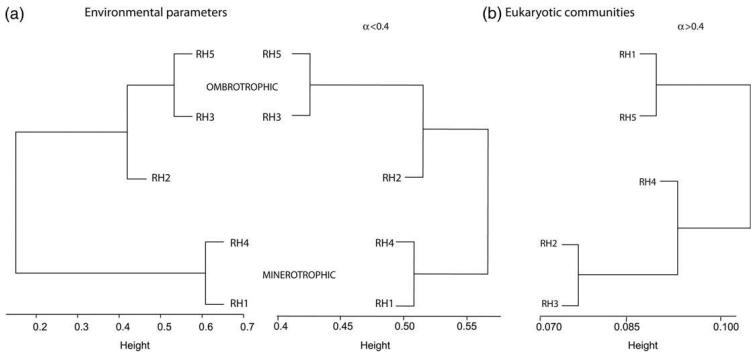

Figure A.2 – Correspondence between dendrograms based on (a) Random forest analysis of environmental parameters (DRP, TP, TH and A440 nm) and (b) communities as determined by GUniFrac analysis, with respectively a low (<0.4) and a high (>0.4) a parameter. When a is low, the dendrogram based on environmental patterns corresponds perfectly to the one based on communities.

The 5% most characteristic OTUs comprised all trophic types (fig. A.3 and fig. A.4). Sequences related to osmotrophic organisms where rare, occurred in the ombrotrophic pools and were represented by a basidiomycetous yeast related to *Rhodotorula* and a Mucorale (previously zygomycete) that belongs to genus *Mortierella*. Phagotrophic organisms were equally represented in both minerotrophic and ombrotrophic environments, with respectively 13 versus 11 OTUs. However, the taxonomic composition of this trophic group changed drastically between the two types of environment; small bacterivores were mainly represented by chrysophytes from genus *Paraphysomonas* in minerotrophic environments and free-swimming bicosoecids (i.e. Pseudodendromonadales) in ombrotrophic pools. Parasitoids of planktonic protists included Cryptomycota and Perkinsea in both environments, and some parasites of plants and invertebrates appeared also (respectively, a Plasmodiophorid and an Entomophthoromycota), possibly infecting organisms from the zooplankton. Pigmented organisms were divided into mixotrophic and strictly phototrophic organisms. Minerotrophic water body indicators included five autotrophic OTUs versus one in ombrotrophic systems. In contrast, mixotrophic indicator OTUs were far more represented in ombrotrophic pools than in minerotrophic, with respectively 3 and 15 OTUs. Still, the trophic mode of 13 OTUs could not be determined, which represents ∼17% of all sequence reads.

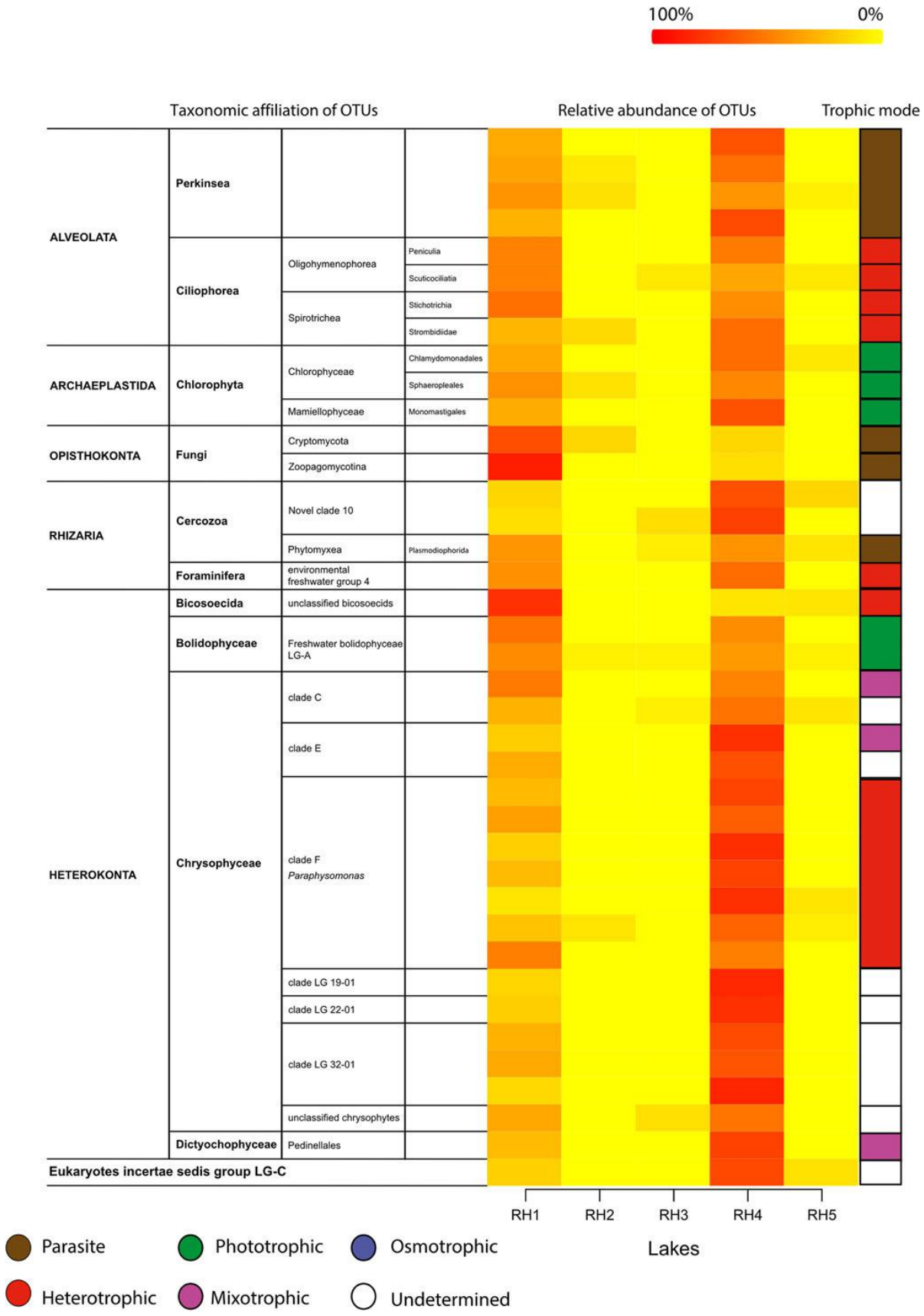

Figure A.3 – Taxonomical affiliation of the OTUs which have been found to be the 5% best indicators for minerotrophy. The heat map indicates relative proportions of each read, from rare to frequent (see reference bar above). Inferred trophic types are shown on the right, next to the corresponding OTU.

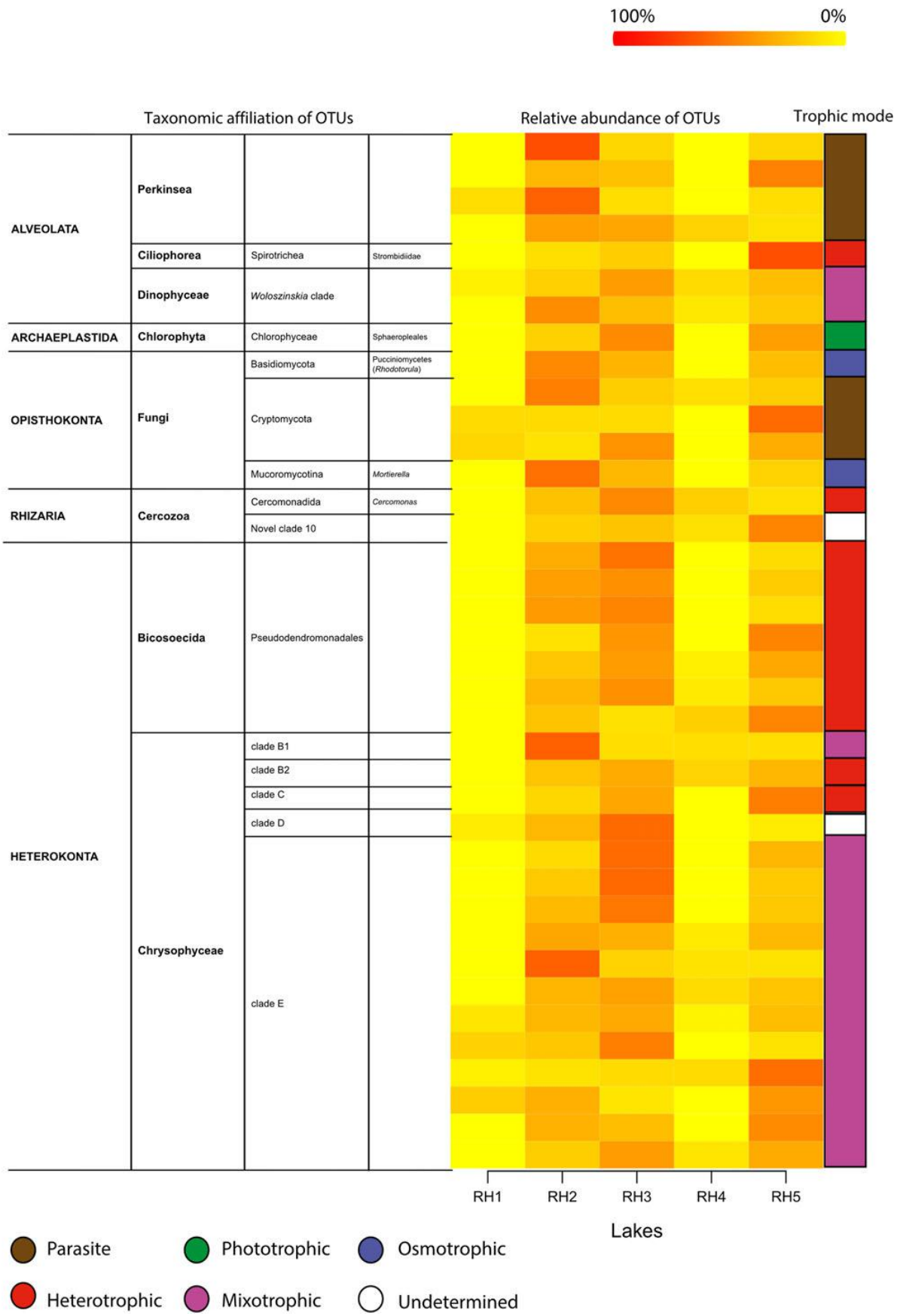

Figure A.4 – Taxonomical affiliation of the OTUs which have been found to be the 5% best indicators for ombrotrophy. The heat map indicates relative proportions of each read, from rare to frequent (see reference bar above). Inferred trophic types are shown on the right, next to the corresponding OTU.

Discussion

Taxonomic and functional composition of planktonic communities

A high abundance and diversity of sequences related to Chrysophyceae in oligotrophic freshwater systems has been observed in many environmental DNA surveys (Charvet et al., 2012; Richards et al., 2005), and less so in more eutrophic systems where plankton had been filtered following a protocol similar to the one described here (Lepere et al., 2006; Slapeta et al., 2005). Notably, they were very abundant in a pristine oligotrophic peat bog pool in Switzerland, an environment which resembles strongly the water bodies described in this study (Lara et al., 2011). Their abundance in the Rancho Hambre pools corroborates the importance of chrysophyceae in oligotrophic and acidic freshwater systems. Perkinsea are also a common group in freshwaters (Breiman, 2001; Mangot et al., 2011; Sime-Ngando and Niquil, 2011). Their dominance in sample RH2 may be due to a local peak of abundance limited in time, as it has been shown to occur in a peat bog (Lara et al., 2011). Ciliates also reach high numbers in one single lake, i.e. RH5, but occur in limited amounts in the other water bodies, possibly for the same reasons as for Perkinsea. A study that follows microeukaryotic populations through time could show if these high sequence abundances in certain lakes are persistent in time. In contrast, Clade 10 cercozoans are consistently found in high numbers in all water bodies. This clade has, to date, only been detected in the picoplanktonic fraction of oligotrophic (Richards et al., 2005) and mesoeutrophic lakes (Lefevre et al., 2008; Lefranc et al., 2005). These organisms appear also in high numbers in these studies, which suggests an important role in planktonic communities. However, they were not detected in surveys of planktonic communities based on morphological identification, probably because of their small size. Therefore, they have never been isolated, and nothing is known about their morphology, trophic strategy or life cycle.

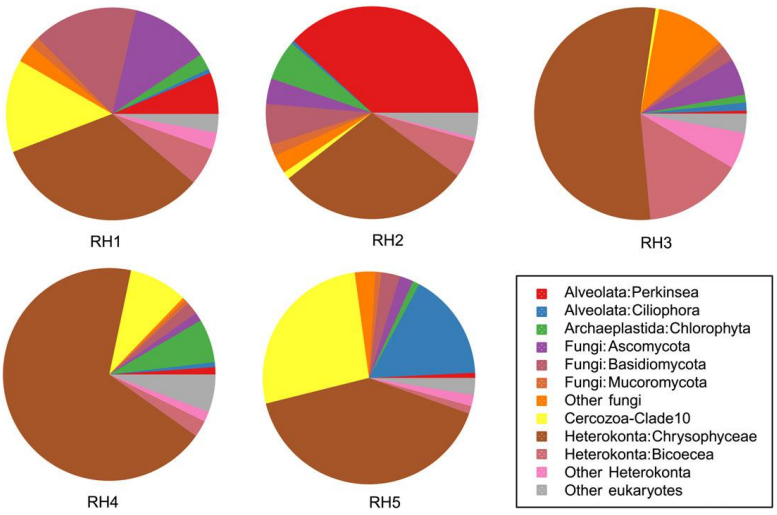

Figure A.5 – Pie charts representing the high-level taxonomic composition of the communities in the five water bodies RH1-5.

The difficulty of relating DNA sequences and lifestyle traits is a pervasive problem for the interpretation of environmental DNA sequences. Chrysophytes, for instance, have lost their photosynthetic ability several times in their evolutionary history, switching from a mixotrophic to a heterotrophic state. Consequently, small bacterivorous nanoflagellates that were previously grouped mostly within genera *Spumella* and *Monas* are now scattered over the chrysophycean tree (Boenigk et al., 2005a). Consequently, not all chrysophyte OTUs could be assigned to trophic modes in this study, either because mixotrophic and heterotrophic branches are intermingled in certain subgroups, or because no sequence has been related to a given morphotype (Richards et al., 2005). Likewise, the correspondence between parasite sequences and their hosts would be desirable to understand population fluctuations in planktonic communities. Although the correspondence between Cryptomycota SSU rRNA gene sequences and the ability to parasitize oomycetes and chytrids (Held, 1981), diatoms (Jones et al., 2011) and Amoebozoa (Corsaro et al., 2014b) has been demonstrated, the host range of the majority of these organisms still remains unknown. Likewise, Perkinsea, which have been only recently shown to be widespread in freshwaters (Brate et al., 2010; Mangot et al., 2011; Sime-Ngando and Niquil, 2011) have still not been characterized in that sense, except for frog tadpole pathogen (Davis et al., 2007). These gaps in knowledge call for detailed studies of the organisms that build the planktonic communities that include isolation, cultivation and also barcoding. A better knowledge of the organisms will open the way for an improved understanding of the population dynamics of the plankton ecosystems.

Correlation between community composition and environmental parameters

Hydrological condition of the water bodies influenced environmental parameters, which in turn shaped planktonic communities. When a parameter in GUniFrac was set between 0 and 0.4 (i.e. ranging from a presence – absence analysis to an intermediate weight given to abundant OTUs in each sample), the dendrogram based on community composition corresponded perfectly to the one built on abiotic parameters. In contrast, when

more weight was given to abundant OTUs by setting α above 0.4, the topology of the dendrogram changed totally. This suggests that the most common OTUs are not influenced by differing physicochemical parameters of the water bodies, but subordinate OTUs are. They are therefore the best indicators for minero-ombrotrophic conditions. The differential response of rare and common species to environmental communities has been studied in many different models, and results differ widely between cases. For instance, in plants from grassland communities, richness patterns of rare species were less predictable than those of common species (Lennon et al., 2011). Conversely, in benthic stream macroinvertebrates, both rare and common species react in the same way (Siqueira et al., 2012). A recent study on chironomid larvae in subtropical reservoirs showed that rare species underwent stronger niche selection and reacted more strongly to environmental fluctuation than common ones (Petsch et al., 2015). Accordingly, we suggest that abundant species may have wider ecological niches and can grow in minerotrophic and ombrotrophic pools alike. Less common OTUs, in turn, are probably more specialized and thus indicate better the trophic level of the ponds. Alternatively, decreasing the influence of frequent species by decreasing α reduces the effect of local peaks of abundance, such as the ones observed for Perkinsea and Ciliates (fig. A.5).

Characteristic organisms for minerotrophic and ombrotrophic pools

The 5% most characteristic OTUs for (i) minerotrophic and (ii) ombrotrophic pools comprised organisms that used all trophic strategies: autotrophy, heterotrophy, mixotrophy, osmotrophy and parasitism. However, proportions of each varied considerably. Characteristic mixotrophs were clearly more frequent in ombrotrophic environments, where they represented more than one-third of all OTUs selected for specificity versus only three in the minerotrophic pools. In contrast, only one strictly autotrophic indicator organism was present in ombrotrophic environments (a member of the Sphaeropleales), against five in the minerotrophic water bodies (fig. A.4). A higher diversity of mixotrophs in ombrotrophic environments can be paralleled to higher abundances, as it has been already widely documented in freshwater plankton through microscopic observation, including Argentinean Patagonia and Tierra del Fuego (Saad et al., 2013). Altogether, these results suggest that mixotrophy is a winning strategy in oligotrophic systems.

There was roughly the same number of heterotrophic environment-specific organisms in pools of both trophic status (respectively, 13 in minerotrophic ombrotrophic habitats), but their composition clearly differed. In the case of bacterivorous nanoflagellates, Bicosoecids (i.e. Dendromonadales), which here characterize ombrotrophic environments, have been found also in another diversity study of a peat bog based on environmental DNA (Lara et al., 2011). In contrast, minerotrophic water bodies were characterized by the chrysophyte genus *Paraphysomonas* (Scoble and Cavalier-Smith, 2014), a group of nanoflagellates with characteristic self-secreted silica scales which has been found abundantly in the highly mineralized Lake Alchichica in Mexico (Couradeau et al., 2011). They are also abundant in marine systems (Mazei and Tikhonenkov, 2006). As Paraphysomonas leave remains (silica scales of characteristic shape) that are preserved in the long run, their presence in core sediments can be used to infer ancient status of water bodies and past hydrographic conditions.

Only a few osmotrophs were selected as most characteristic organisms. Only one OTU corresponding to a yeast (*Rhodotorula* sp.) and one filamentous fungus (related to *Mortierella* sp.) appeared, in spite of the fact that total reads related to Ascomycetes and Basidiomycetes are very abundant. These two fungal genera are very widespread and are common in freshwater systems, and can therefore be considered as forming part of the indigenous communities (Cray et al., 2013). A possible explanation for the absence of more filamentous fungi is that organisms develop outside the pools and their spores are carried by the wind and fall randomly within one or another pool.

Characteristic parasitoid organisms were represented by Perkinsea and Cryptomycota; sequences from these two groups have been often recorded in freshwater environmental DNA surveys (Lara et al., 2009a; Lefevre et al., 2008; van Hannen et al., 1999). Their presence as best indicator organisms probably reflects that of a host which is, in turn, specific to a certain environment. Their abundance may reveal an important role in regulating host populations, as well as a re-mobilization of nutrients (Lefevre et al., 2008). Understanding their exact role in planktonic community dynamics requires a thorough study of their life cycles and host range, a task that requires isolation and cultivation of potential hosts as well as detection of the parasite. Environmental DNA surveys provide an unprecedentedly deep insight into the taxonomic diversity in planktonic microeukaryotic diversity. In that respect, they are indispensable. However, their interpretation still requires "traditional protistology skills," which have perhaps never been as useful as nowadays.

Acknowledgements

Thanks are due to the Dirección General de Recursos Hídricos de la Provincia de Tierra del Fuego, the Centro Austral de Investigaciones Científicas (CADIC) and Dr Daniel Fernández for most valuable logistic support. The help of Engineer Sergio Camargo in the fieldwork was much appreciated.

Funding

This work has been funded by the Swiss National Science Foundation (project FN 31003A-143960/1) given to E.L. Field work was financed by the CONICET project PIP 11220090100050.

Appendix B

Microbial eukaryote communities exhibit robust biogeographical patterns along a gradient of Patagonian and Antarctic lakes

M. Romina Schiaffino[1], Enrique Lara[2], Leonardo D. Fernández[2,3], Vanessa Balagué[4], David Singer[2], Christophe C.W. Seppey[2], Ramon Massana[4] , Irina Izaguirre[5]

[1]Centro de Investigaciones y transferencia del Noroeste de la Provincia de Buenos Aires, UNNOBA, 6000, Junín, Argentina.
[2]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland.
[3]Laboratorio de Ecología Evolutiva y Filoinformática, Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Barrio Universitario s/n, Concepción, Chile.
[4]Institut de Ciències del Mar, CSIC, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.
[5]Departamento de Ecología, Genética y Evolución, IEGEBA (CONICET-UBA), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Ciudad Universitaria, C1428EHA, Buenos Aires, Argentina.

Environmental Microbiology, Online published, october 2016
DOI: 10.1111/1462-2920.13566

Summary

Microbial eukaryotes play important roles in aquatic ecosystem functioning. Unravelling their distribution patterns and biogeography provides important base-line information to infer the underlying mechanisms that regulate the biodiversity and complexity of eco-systems. We studied the distribution patterns and factors driving diversity gradients in microeukaryote communities (total, abundant, uncommon and rare community composition) along a latitudinal gradient of lakes distributed from Argentinean Patagonia to Maritime Antarctica using both denaturing gradient gel electrophoresis (DGGE) and high-throughput sequencing (Illumina HiSeq). DGGE and abundant Illumina operational taxonomic units (OTUs) showed both decreasing richness with latitude and signifi-cant differences between Patagonian and Antarctic lakes communities. In contrast, total richness did not change significantly across the latitudinal gradient, although evenness and diversity indices were signifi-cantly higher in Patagonian lakes. Beta-diversity was characterized by a high species turnover, influenced by both environmental and geographical descriptors, although this pattern faded in the rare community. Our results suggest the co-existence of a 'core bio-sphere' containing reduced number of abundant/ dominant OTUs on which classical ecological rules apply, together with a much larger seedbank of rare OTUs driven by stochastic and reduced dispersal processes. These findings shed new light on the bio-geographical patterns and forces structuring inland microeukaryote composition across broad spatial scales.

Keywords: biogeographical patterns; core biosphere; DGGE; high-throughput sequencing; lakes; microbial eukaryote diversity; rare biosphere

Introduction

Phototrophic and heterotrophic small eukaryotes of picoplanktonic, nanoplanktonic and microplanktonic size (i.e. microbial eukaryotes with a size ranging between 0.2 and <20 µm) are abundant in the upper photic zone of all lakes and oceans (Caron et al., 1999; Foissner, 1999a; Li et al., 1983) and play important roles in the biogeochemical cycles (e.g. (Azam et al., 1983; Massana, 2011; Pomeroy, 1974; Stockner and Antia, 1986). The introduction of molecular techniques into microbial ecology has greatly increased our knowledge on microbial eukaryotes diversity (e.g. (Epstein and Lopez-Garcia, 2008). Fingerprinting techniques, such as denaturing gradient gel electrophoresis (DGGE), allow treating a relatively higher number of samples (van Hannen et al., 1998), but provide limited information on the taxonomical composition of microbial communities and retrieve only the abundant or dominant taxa, the so called 'core species' (Magurran and Henderson, 2003). In the last years, the introduction of novel high-throughput sequencing technologies permitted reaching saturation of microbial diversity, including the rarest taxa (the so-called 'rare biosphere', e.g. (Dawson and Hagen, 2009; Lepere et al., 2013; Logares et al., 2014; Monchy et al., 2011; Sogin et al., 2006; Taib et al., 2013)).

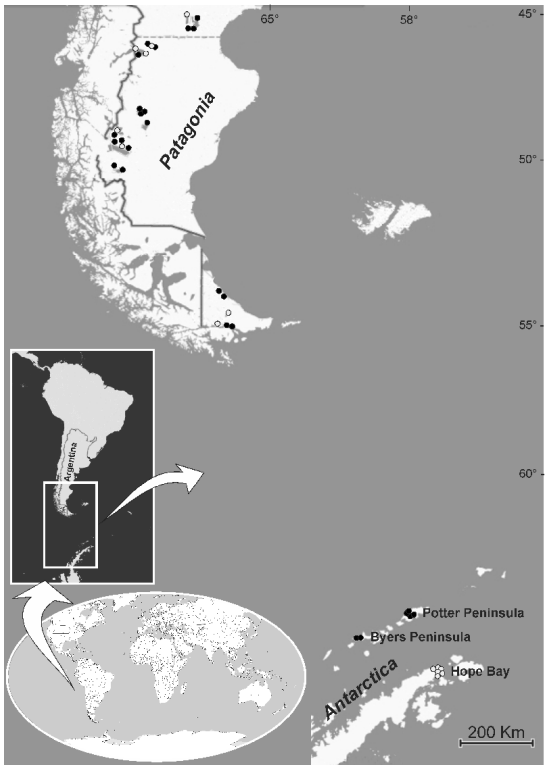

Figure B.1 – Map showing the 40 locations of the studied water bodies. All sites were studied with DGGE, while those that also were studied with Illumina HiSeq (14 lakes) are shown with white circles.

Diversity is considered a key property of ecosystems, necessary to understand the structure, functioning and dynamics of communities (e.g. (MacArthur and Wilson, 1967; Magurran, 2004)). It comprises two basic components, taxa richness (number of taxa in a community) and evenness (a measure of how similar taxa are in their abundances) (Magurran, 2004). Weak relationships between these two components in many aquatic datasets reflect independent components of diversity, suggesting that richness alone may be an incomplete surrogate for biodiversity (Soininen et al., 2011). Thus, richness and evenness should be treated separately (e.g. (Hurlbert, 1971; Legendre and Legendre, 1998; Magurran, 1988)) because both provide meaningful insights into community function (McNaughton, 1977) and may respond independently to different ecological processes (Ma, 2005). Beta-diversity (Whittaker, 1960) is usually applied in a broad sense to any measure of variation in taxa composition between sites (Anderson et al., 2011). As suggested by (Baselga, 2010; Baselga and Orme, 2012), it can be partitioned

into two components: turnover (the replacement of some taxa by others from site to site) and nestedness (the pattern characterized by the poorest site being a strict subset of the richest site, without replacement of taxa).

Ecological communities are also characterized by a few species exceptionally abundant, while the majority often remains remarkably rare (Fisher et al., 1943; Magurran, 2004; Preston, 1948). Therefore, communities can be separated into two components: 'core species', which are persistent and abundant, and 'rare or occasional species', which occur infrequently in the record and are typically low in abundance (Magurran and Henderson, 2003). This pattern is repeated across taxonomic groups (Magurran, 2004), including planktonic microorganisms (e.g. (Mangot et al., 2013; Pedros-Alio, 2006)). These two components are not independent from each other, and rare taxa can be promoted from the 'rare biosphere' to the 'core biosphere' if conditions alter sufficiently and become adequate for the growth of a particular taxon. Likewise, a member of the 'core biosphere' can be grazed down or reduced by viral lysis below the threshold level to become a member of the 'rare biosphere' without disappearing from the ecosystem (Pedros-Alio, 2006).

Recent studies have aimed to characterize the patterns of this 'rare biosphere' in ecosystems (Lepere et al., 2013; Logares et al., 2014). According to the 'seed bank theory' and as mentioned before, the 'rare biosphere' is supposed to play the role of a large inactive seed bank, dormant most of the time (e.g. as cysts or spores), and awaiting adequate conditions to develop (Gibbons et al., 2013; Lennon et al., 2011). Previous studies in freshwater seasonal variation (Lara et al., 2011; Simon et al., 2015) suggested that members of the eukaryotic community may enter some way of dormancy and constitute a 'seed bank' that participates to plankton community resilience over time. Furthermore, (Simon et al., 2016) recently observed that the capacity to enter dormancy via the generation of resting forms (e.g., cysts, spores or other resistant metabolic stages) is the major explanation for the resilience of microbial communities observed in shallow freshwater ecosystems undergoing frequent droughts. Since a dormant microbial cell can survive for a long time in sub-optimal environments and can be transported passively over large distances in the meantime, some researchers argued that microbial species would have a cosmopolitan distribution (Finlay, 2002). The implications of these assumptions are twofold. First, it would mean that the large spatial diversity patterns observed in the 'rare biosphere' are independent of the forces that drive current diversity patterns in other life-forms (e.g. plants and animals), such as environmental factors and isolation by distance. Second, it would mean that beta-diversity patterns (i.e. variation in OTU composition among sites) will tend to be null. This globally more or less homogeneous rare and inactive biosphere would then bring only redundant information between different high-throughput sequencing surveys, and early molecular techniques such as DGGE would suffice to describe the biologically relevant part of diversity. Thus, if it were true that all microbes are easily dispersed globally, the long tail of the distribution of species abundances in any ecosystem would include all of the microorganisms on Earth, and the biodiversity of any ecosystem would be identical to the biodiversity on the planet (Pedros-Alio, 2006). However, recent studies do not support these assumptions. It has been shown in freshwater (Debroas et al., 2015) and marine (Logares et al., 2014) systems that at least some of the rare microorganisms are active, and that microbial eukaryotes from high-mountain lakes are not globally distributed (e.g. (Filker et al., 2016)). In addition, studies focusing on marine prokaryotes showed that rare OTUs followed similar distribution patterns to those of the most abundant members of the community and of the entire community (Galand et al., 2009; Logares et al., 2013). Similar patterns have been also observed for marine and freshwater microbial eukaryotes (Lepere et al., 2013; Logares et al., 2014). The profound influence of microorganisms on human life and global biogeochemical cycles underlines the value of studying the biogeography of microorganisms (Lynch and Neufeld, 2015). The most common pattern in biogeographical studies is the decrease of richness with increasing latitude, although this pattern may depend on the spatial scale and the level of taxonomic resolution (e.g. (Hillebrand, 2004; Willig et al., 2003)). However, there are no clear geographical patterns in relation to evenness and contrasting patterns have been reported (e.g. (Willig et al., 2003)). South American and Antarctic lakes are propitious systems to study the diversity distribution patterns in microorganisms. Indeed, these freshwater systems can be found along a significant latitudinal gradient that exhibits relatively benign (low latitudes) to extremely harsh (higher latitudes) environmental conditions, as well as important barriers to dispersal (e.g. the Magellan Strait and the Drake Passage), that allow not only assessing their diversity patterns but also the importance of environmental and geographical factors in determining these patterns. Previous studies on microbial diversity patterns across Patagonian and Antarctic lakes have found a decreasing latitudinal diversity gradient for diatoms (Maidana et al., 2005), Chlorococcales (Tell et al., 2011), phytoplankton (Izaguirre et al., 2016), and bacterioplankton (Schiaffino et al., 2011).

Here, we have used both traditional fingerprinting (DGGE) and high-throughput sequencing technology (Illumina HiSeq) methods to study the patterns and drivers of latitudinal diversity gradients in total, abundant, uncommon and rare microbial eukaryote community along a gradient of freshwater environments (fig. B.1) stretching from Argentinean Patagonia (45°S) to Maritime Antarctica (63°S). We used DGGE, which focused on the dominant taxa, on the complete dataset (40 lakes), while the Illumina HiSeq approach that allowed reaching complete taxonomical reports was applied on a subset of samples (14 lakes). The use of both approaches provided

Table B.1 – Comparisons between cell abundances and diversity properties of microbial eukaryotes from Patagonian and Antarctic water bodies. Average per lake, n = 40 (28 Patagonian and 12 Antarctic water bodies) and n = 14 (8 Patagonian and 6 Antarctic water bodies).

			Patagonian lakes	**Antarctic lakes**
Epifluorescence	Autotrophic abundance	n = 40	7.7 x 104***	5.4 x 103***
	Heterotrophic abundance		1.1 x 104**	2.2 x 103**
DGGE	Dominant OTU richness	n = 40	21*	18*
	Dominant OTU richness	n = 14	20	18
Illumina HiSeq	Total OTU richness		347	313
	Chao 1		539	493
	Abundant OTUs ≥1% richness		19**	10**
	Uncommon OTUs >0.1 – <1% richness		52	38
	Rare OTUs ≤0.1% richness	n = 14	276	265
	Simpson Index (D)		13.4*	6.7*
	Shannon Index (H´)		3.4*	2.6*
	Evenness (E_D)		0.041	0.022
	Evenness ($E_{H´}$)		0.584*	0.445*

Student´s t-test *P <0.05, **P <0.01, ***P <0.001

Table B.2 – pearman Rho correlations between microbial eukaryote properties (abundance, richness and diversity) and some environmental variables from all the studied water bodies. n = 40 and n = 14 for DGGE and epifluorescence microscopy counts, and n = 14 for Illumina HiSeq reads.

		Lake area	**Latitude**	**Temperature**	**Chl A**
Epifluorescence (n = 40)	Autotrophic abundance	0.16	**-0.61****	**0.67****	**0.42***
	Heterotrophic abundance	-0.31	**-0.46***	**0.56****	**0.47***
DGGE	Dominant OTU richness (n = 40)	0.11	**-0.39**	0.14	**0.34**
	Dominant OTU richness (n = 14)	0.08	-0.32	0.15	**0.59**
Illumina HiSeq (n = 14)	Total OTU richness	-0.04	-0.01	0.1	-0.1
	Chao 1	-0.02	0.01	0.1	-0.29
	Abundant OTUs ≥1% richness	**0.74***	**-0.74***	**0.58**	-0.22
	Uncommon OTUs >0.1 – <1% richness	0.08	-0,18	0.22	0.19
	Rare OTUs ≤0.1% richness	-0.15	0.13	0.07	-0.2
	Simpson Index	0.48	-0.39	0.21	0.01
	Shannon Index	0.41	-0.39	0.3	0.05
	Evenness (E_D)	0.47	-0.29	0.1	-0.06
	Evenness ($E_{H´}$)	**0.54**	-0.53	0.4	0.02

Bold values represent P <0.05, *P <0.01, **P <0.0001.

a unique opportunity to compare their performance and conclusions. For both datasets, we hypothesised that OTU richness and diversity follow a decreasing pattern with increasing latitude, environmental harshness and geographical distance, just as it has been repeatedly shown for several other taxa (Brown and Lomolino, 1998; Gaston and Blackburn, 2000). In addition, given that environments from lower latitudes are less extreme and more productive than those from higher latitudes (Chase and Ryberg, 2004; Currie et al., 2004; Hawkins et al., 2003), we proposed that there are a higher proportion of dominant OTUs and evenness in lower latitudes than in higher latitudes. Furthermore, important changes in community compositions are assumed between Patagonian and Antarctic lakes due to environmental and geographical factors, which will be reflected by beta-diversity. Alternatively, we postulated that beta-diversity patterns of the rare biosphere are independent of the environmental and/or geographical distances, giving support to the 'seed bank theory'.

Results

Total abundances

In most water bodies, the abundance of autotrophic eukaryotes was greater than the abundance of heterotrophic ones (both 5 µm in diameter, (table B.1), and both showed the same behaviour with respect to the trophic and latitudinal gradients (table B.2). Patagonian water bodies presented significantly higher autotrophic and heterotrophic eukaryotes abundances than Antarctic ones (table B.1). In addition, the abundance of autotrophic and heterotrophic eukaryotes decreased significantly towards higher latitudes and accordingly increased significantly with temperature and higher chl α (table B.2).

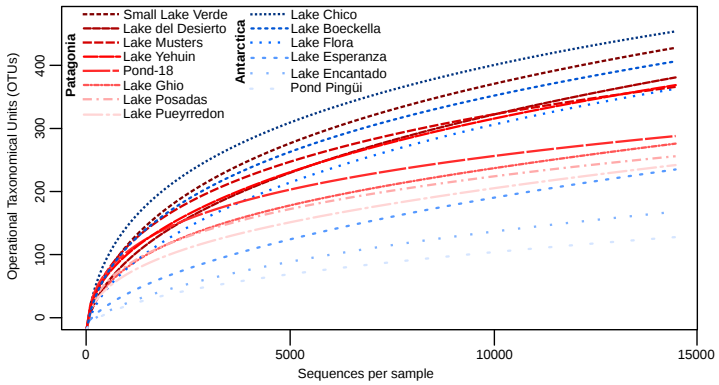

Figure B.2 – Rarefaction curves for each water body using Illumina HiSeq dataset.

DGGE analysis

The DGGE gels performed for 40 lakes yielded a total of 819 bands located in 90 different positions (total dominant richness): 24 positions (27%) were exclusive in Patagonian lakes and 3 (3%) in Antarctic lakes, whereas 63 positions (70%) were shared between Patagonian and Antarctic water bodies. We successfully sequenced 26 different OTUs (representing 40% of the total intensity of the gels). Bands with the same position were sequenced from different lakes to confirm that they corresponded to the same sequence: e.g. bands 3 (a, b), 7 (a, b) and 17 (a, b) (fig. B.3). We found members of five taxonomic groups: 10 (38%) Stramenopiles (8 Chrysophyceae-Synurophyceae, 1 Bicosoecida, 1 Dictyochophyceae), 10 (38%) Alveolata (1 Alveolata, 8 Ciliophora and 1 Apicomplexa), 3 (12%) Archaeplastida (Chlorophyta), 2 (8%) Fungi and 1 (4%) Telonemida (Supporting Information Table S1). Some sequences showed their closest match and phylogenetic affiliation (between 98 and 100% of similarity) with well-described species of small phytoplankton (generally smaller than 5 µm) *Vitreochlamys incisa* (band 15), *Pyramimonas australis* str. (band 16), *Ostreococcus tauri* (band 26). Other sequences matched with OTUs belonging to taxonomic groups represented by larger cell sizes (e.g. Alveolata), possibly due to the inefficient sequential filtration (Supporting Information Table S1).

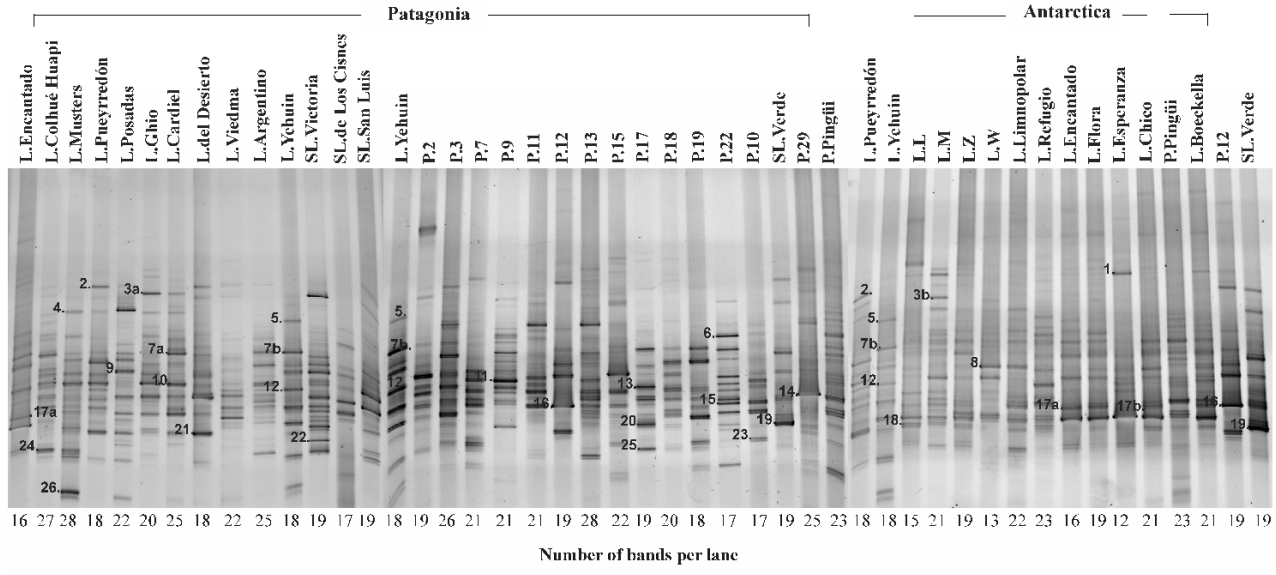

Figure B.3 – Image of the DGGE gels run for 40 water bodies. Numbers in the gels indicate excised and equenced bands. L.: lake, SL.: shallow lake, P.: pond.

Illumina HiSeq analysis

From a total of 5 552 925 paired-end sequences, 3 670 418 passed the quality check. The total richness across all studied lakes was 1 505 different OTUs and ranged from 147 to 475 OTUs in each individual water body (332 699 OTUs on average). Rarefaction curves performed for every water body showed that OTUs richness tended to reach a plateau, indicating that the sampling depth and sequencing coverage were good (fig. B.2). In general, Stramenopiles and Alveolata dominated diversity in the entire study (fig. B.4). Taxonomic composition showed clear differences between Patagonian and Antarctic water bodies: Cryptophyta were abundant in Patagonia but absent from Antarctic lakes, whereas Chrysophyta dominated in Antarctic lakes (fig. B.4). Ciliophora showed higher relative abundances and were more frequent in Patagonian than in Antarctic lakes, except in the Antarctic Pingüi Pond (fig. B.4). This system has a large contribution of organic matter from penguins and represents an outlier in our dataset.

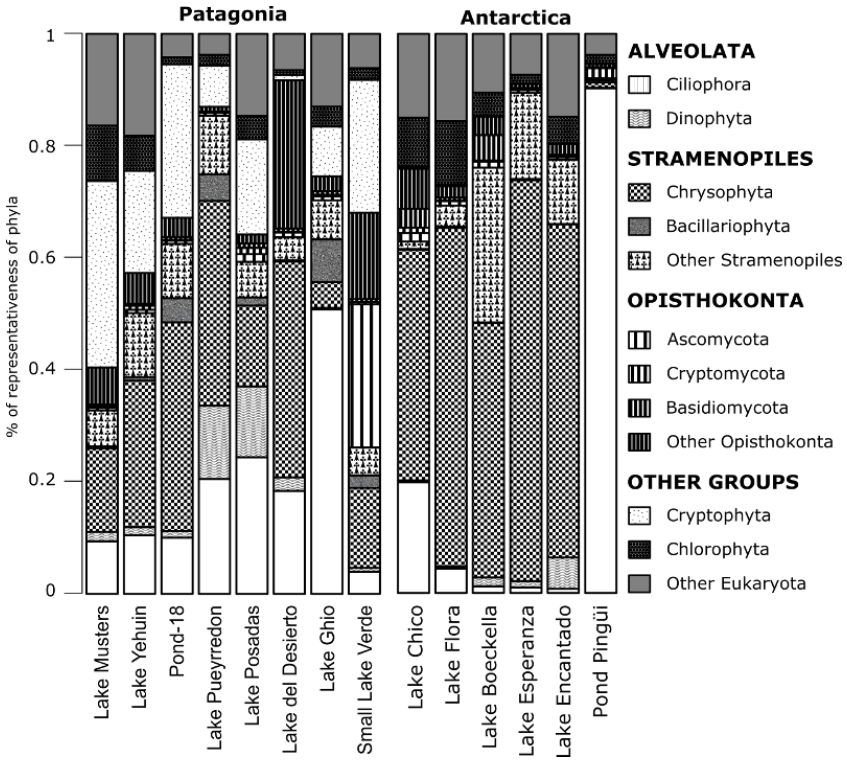

Figure B.4 – Relative composition of phyla from the 14 water bodies studied with Illumina HiSeq.

Globally, the 'uncommon and rare biosphere' presented a high percentage (92%) of exclusively uncommon or rare OTUs, whereas the 'core biosphere' showed a low percentage (4%) of exclusively dominant/abundant OTUs. Therefore, from the total 125

Table B.3 – Results of Mantel (simple and partial) and variation partitioning analyses to study the effect of geographical (PCNM) and environmental variables on microbial eukaryote community composition (based on DGGE band pattern and Illumina HiSeq reads). DOC: dissolved organic carbon. EF = Each Factor (%), BF = Both Factors (%). U = Unexplained (%)

Matrix type		**Mantel**		**Partial Mantel †**		**Variation partitioning**			***Forward selection of variables***
		r	***P***	***r***	***P***	**EF**	**BF**	**U**	
DGGE	Geo.	**-0.3**	0.002	**-0.3**	0.002	**6**	5	82	PCNM1, PCNM3
(n = 40 lakes)	Env.	**0.15**	0.02	**0.14**	0.02	**7**			temperature, conductivity
DGGE	Geo.	**-0.6**	0.002	**-0.4**	0.002	**8**	8	78	PCNM1, PCNM3
(n = 14 lakes)	Env.	**0.47**	0.002	**0.26**	0.02	**6**			temperature, lake area
Illumina total	Geo.	**-0.7**	0.003	**-0.4**	0.003	**9**	17	72	PCNM1, PCNM3
OTUs	Env.	**0.76**	0.002	**0.6**	0.003	2			temperature
Illumina abundant	Ge.	**-0.6**	0.002	**-0.6**	0.002	**6**	20	66	PCNM1, PCNM3
OTUs ≥1%	Env.	**0.57**	0.002	**0.54**	0.002	**8**			DOC,temperature, conductivity
Illumina uncommon	Geo.	**-0.7**	0.002	**-0.6**	0.002	**5**	17	73	PCNM1
OTUs >0.1 – <1%	Env.	**0.6**	0.002	**0.43**	0.002	**5**			conductivity, temperature
Illumina rare	Geog.	**-0.7**	0.002	**-0.6**	0.002	**9**	17	72	PCNM1, PCNM3
OTUs ≤0.1%	Env.	**0.43**	0.004	0.16	0.106	2			temperature

†The Partial Mantel test holds geographical or environmental matrix constant. Bold values are significant data ($P < 0.05$).

Table B.4 – Total, exclusive and shared number of sequence reads (Illumina HiSeq) for all Patagonian and Antarctic water bodies. Total OTU richness from 14 water bodies: 8 Patagonian and 6 Antarctic ones.

	Total OTUs	Shared OTUs	Exclusive OTUs	OTUs present in all water bodies	OTUs present in one water body
			Total matrix		
Patagonian lakes	1505	486	663	24	445
Antarctic lakes			356		
			Abundant OTUs ≥1%		
Patagonian lakes	125	2	91	0	84
Antarctic lakes			32		
			Uncommon OTUs >0.1% – <1%		
Patagonian lakes	422	24	257	1	295
Antarctic lakes			141		
			Rare OTUs ≤0.1%		
Patagonian lakes	1438	461	618	3	506
Antarctic lakes			359		

'core OTUs' found in 14 lakes, 120 (96%) were dominant in some lakes but also rare in other lakes, whereas only 5 (4%) were not found as rare in other lakes. In addition, 67% of these 125 OTUs were dominant just in one lake and 33% were dominant in two or more lakes (up to 6) lakes, and none of these OTUs was dominant in all studied lakes. On the other hand, from the total 1 500 OTUs found in the 'uncommon and rare biosphere' (OTUs ≤1%), 1 380 OTUS (92%) were always uncommon and/or rare (exclusive) while 120 of these OTUs (8%) were also found in the 'core biosphere' in other lakes. In all samples, we found a small number (around 8%) of abundant OTUs (≥1%), while most OTUs (around 92%) were rare (OTUs ≤0.1%). Patagonian lakes showed significantly higher numbers of dominant OTUs (≥1%) than Antarctic ones, whereas Antarctic lakes showed significantly higher numbers of rare OTUs (≤0.1%) than Patagonian lakes (fig. B.5). Comparisons among each abundance category (total Illumina matrix, OTUs ≥1%, OTUs >0.1% to <1% and OTUs ≤0.1%) for all Patagonian and Antarctic water bodies are shown table B.4. Patagonian lakes showed between 43 and 73% of exclusive OTUs, while Antarctic lakes showed between 24 and 33%. Patagonian and Antarctic lakes shared between 2 and 32% OTUs. For instance, the abundant OTUs community composition (≥ 1%) showed a total of 125 OTUs and only 2 OTUs belonging to Chrysophyceae were equally dominant in both Patagonian and Antarctic water bodies. All other abundant OTUs (123 OTUs) could dominate either in Antarctic (32) or in

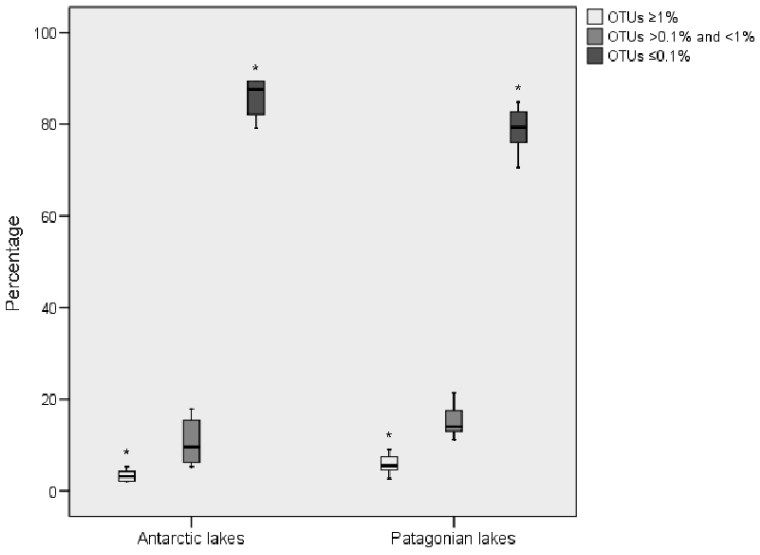

Figure B.5 – Average percentage per lake of the abundant, uncommon and rare OTUs in Patagonian and Antarctic lakes. Illumina community matrix was studied considering the abundant OTUs ≥1% of the total number of sequences in a given sample and the less frequent OTUs with two different criteria: uncommon OTUs >0.1% to <1% and rare OTUs ≤0.1% of the total number of sequences in a given sample. *Student's t-test P < 0.05 for comparing each abundance category (OTUs ≥1%, OTUs >0.1% to <1% and OTUs ≤0.1%) between Patagonian and Antarctic lakes.

Patagonian (91) lakes, but never in both. Contrarily, the rare community composition (≤0.1%) showed a total of 1438 OTUs. From these OTUs, 461 were present in both Patagonian and Antarctic lakes, whereas 977 OTUs could only be present in Antarctic (359) or in Patagonian (618) lakes. This abundance category showed that only 3 OTUs (related to Fungi) were present in all water bodies.

To compare the responses of the individual diversity components (richness and evenness) we studied the relationship between them. No correlations between richness and evenness were found (r=-0.07 for E_D and r=0.39 for E_H, both P>0.05 and n=14). Similar lack of relationship was found when studying the Patagonian lakes (r=-0.41 for E_D and r=-0.27 for E_H, both P>0.05 and n=8), or the Antarctic lakes separately (r=0.26 for E_D and r=0.77 for E_H, both P>0.05 n=6).

General Patterns and comparisons

The abundant part of Illumina reads and the number of DGGE bands were significantly higher in Patagonian water bodies than in Antarctic ones (Table 1), and also decreased significantly with higher latitudes (table B.2). On the other hand, the total richness (as shown by Illumina data) did not differ between Patagonian and Antarctic lakes (table B.1) and neither correlated with latitude (table B.2). Similar results were obtained with the richness estimator Chao 1 index (table B.1 table B.2). Both OTU diversity indices and evenness ($E_{H'}$) were significantly higher in Patagonian water bodies than in Antarctic ones, whereas neither diversity indices nor evenness changed with latitude, except Pielou evenness index ($E_{H'}$) that decreased with higher latitudes (table B.1 table B.2).

Microbial eukaryote communities exhibited high values of beta-diversity either based on DGGE (β_{SOR}=0.94, P<0.001) or on Illumina (β_{SOR}=0.84, P<0.001) datasets. In both cases, variation in community composition was mainly due to OTUs turnover (DGGE β_{SIM}=0.93 and Illumina β_{SIM}=0.86, both P<0.05) and barely due to nestedness (DGGE β_{NES}=0.01 and Illumina β_{NES}=0.02, both P<0.05) (fig. B.6).

Redundancy analyses (RDA) based on community composition versus environmental and geographical variables showed that in general Patagonian lakes ordinated apart from the Antarctic lakes (with some exceptions for the hypertrophic Antarctic Pingüi Pond and the Patagonian shallow lake Verde) when considering microbial communities from DGGE matrices (n=40 and n=14), total Illumina matrix and different Illumina abundance categories (abundant ≥1%, uncommon >0.1% to <1% and rare ≤0.1%) (fig. B.7). To study the main underlying driver of beta-diversity along the latitudinal gradient we applied variation partitioning and Mantel test (simple and partial) analyses (table B.3). RDA performed on each studied microbial eukaryote community matrix showed that forward selection retained the following variables as relevant: PCNM 1 and PCNM 3 (geographical variables), temperature, conductivity, DOC and lake area (environmental variables) (table B.3). Variation partitioning analyses showed that the amount of variation explained by each type of factor (geographical or environmental alone) varied with the molecular approach (DGGE or Illumina) and the community abundance category studied (Illumina data: abundant, uncommon or rare). Purely geographical and environmental factors explained similar and significant amount of variation on both DGGE (n=14) and abundant Illumina microbial eukaryote communities. However, when moving to less abundant communities (i.e. uncommon or rare) the respective amount of explained variance and the significance faded; only purely geographical descriptors influenced significantly the rare communities (table B.3). In agreement with the results obtained with variance partitioning, Mantel tests showed that the similarity in total and abundant OTU community composition (both from Illumina

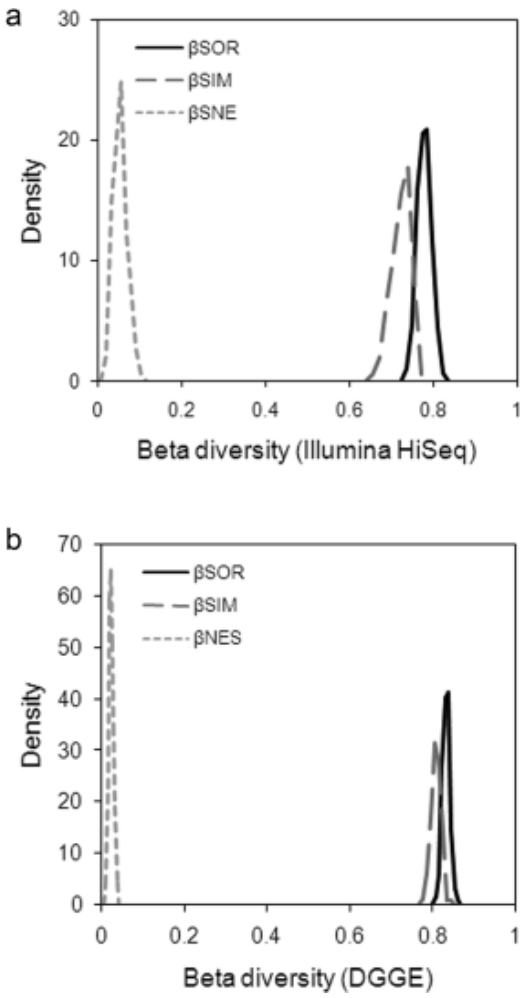

Figure B.6 – Mechanisms driving variation in OTU composition in Patagonian-Antarctic lakes. The plots show the distribution of dissimilarity values due to overall beta-diversity (β_{SOR}), turnover (β_{SIM}) and nestedness (β_{NES}) based on (a) Illumina HiSeq and (b) DGGE data sets (a) Illumina HiSeq and (b) DGGE data sets, both n 5 14 water bodies. Kernel density curves were constructed by resampling four lakes from each area 1000 times and computing the average β_{SOR} , β_{SIM} and β_{NES} respectively.

reads) and in dominant OTU community composition (DGGE band pattern) decreased significantly when the distance among water bodies increased and grew up significantly when the water bodies showed more similar local environmental factors. However, the rare OTU community composition (OTUs ≤0.1%) only decreased significantly when the distance among water bodies increased and showed an independent effect of environmental parameters (table B.3). On the other hand, both tests did not show the same results with the total communities from Illumina. The Mantel approach showed a significant influence of geographical and environmental descriptors, whereas variation partitioning analysis showed only a higher and significant effect of geographical factors (table B.3).

Discussion

Our results show that microbial eukaryote total richness did not change with latitude (table B.2) and did not differ significantly between Patagonian and Antarctic lakes (table B.1). This contrasts with the prevailing general view of decreasing richness with latitude (Gaston, 2000; Hillebrand, 2004; Pianka, 1966) also observed in soil protists (Lara et al., 2016), although there are a number of exceptions for aquatic (e.g. (Crow, 1993; Passy, 2010)) and soil (Fernández et al., 2016) eukaryotic microorganisms. Therefore, the decreasing latitudinal gradient in taxa richness often reported for larger organisms (Rosenzweig, 1995) is not straightforward for microbial freshwater eukaryotes taken as a whole. Our data are consistent with the report of (Hillebrand and Azovsky, 2001), which showed that the effect of latitudinal diversity gradient decreases with decreasing body size of the taxa under study and almost disappears for protists (though this paper was based on an incomplete taxonomic sampling for protists). However, when only the abundant fraction of OTUs was considered (from Illumina and DGGE), we observed both a decrease in richness with latitude (table B.2) and significant differences between Patagonian and Antarctic lakes (table B.1). Accordingly, cell abundances (both auto- and heterotrophic) decreased also with latitude and were significantly higher in Patagonia than in Antarctica, probably corresponding to the abundant part of diversity (table B.1, table B.2). Similar decreases with latitude of cell abundances and richness of the dominant taxa were previously observed for bacterioplankton (Schiaffino et al., 2011) and phytoplankton (Izaguirre et al., 2016) in the studied lakes, and for autotrophic picoplankton in the Atlantic Southern Ocean (Doolittle et al., 2008).

The diversity and evenness indices of the whole community did not correlate with latitude (except evenness Pielou index $E_{H'}$ that correlated negatively with latitude, table B.2), whereas both were significantly higher in Patagonian lakes than in Antarctic ones (except E_D, table B.1). As mentioned before, diversity has two basic components, richness (number of OTUs in a community) and evenness (how similar are OTUs in their abundances) (Magurran, 2004), which are combined in diversity indices (e.g. Shannon or Simpson indices). We found that total richness did not change in different latitudes, but as abundance was influenced by latitude (table B.1, table B.2), diversity changed significantly between Patagonian and Antarctic lakes. This pattern of diversity would be due to an increase in harsh climatic conditions with latitude, limiting the development of taxa having lower tolerance to climatic adversity. Protists in high-latitude lakes are constrained by cold temperatures, low inorganic nutrient supply and low light availability for much of the year due to ice cover and polar darkness (Charvet et al., 2014; Rengefors et al., 2012). Basically, many of the OTUs found in Antarctic systems could not reach the threshold to be abundant, likely because these ecophysiological constrains. These could also explain why Patagonian lakes showed more dominant OTUs, whereas Antarctic lakes showed more rare OTUs.

Many authors suggest that richness and evenness should be treated separately (e.g. (Hurlbert, 1971; Legendre and Legendre, 1998; Magurran, 1988)) because both provide meaningful insights into community function (McNaughton, 1977) and may respond independently to different ecological processes (Ma, 2005). (Wilsey and Stirling, 2007) proposed that the richness component is more sensitive to migration rates, while evenness is more sensitive to biotic and environmental interactions. We found that evenness and richness were not correlated along the latitudinal gradient of lakes, suggesting that microbial eukaryote communities responded independently to different ecological factors. As no changes in total OTU richness were found in our dataset, but changes in evenness Pielou index ($E_{H'}$) along the latitudinal gradient of lakes emerged (**??**, table B.2), we can suppose that microbial eukaryotes would be more regulated by biotic interactions and environmental stressors, than dispersal and migration. Accordingly, (Soininen et al., 2011) stated that species richness and evenness may respond to different environmental factors or in a different way to a given factor, thus reflecting independent components of biodiversity.

We found a small number (around 8%) of highly abundant OTUs (≥1%), representing the 'core biosphere', while most OTUs (around 92%) were less frequent (fig. B.5), the so called 'rare biosphere'. In line with our results, (Nolte et al., 2010) showed that a small number of highly abundant protistan OTUs are present throughout the entire year, while most OTUs are rare and highly restricted. Similarly, (Debroas et al., 2015) found that freshwater protists were represented by a high number (77%) of rare OTUs.

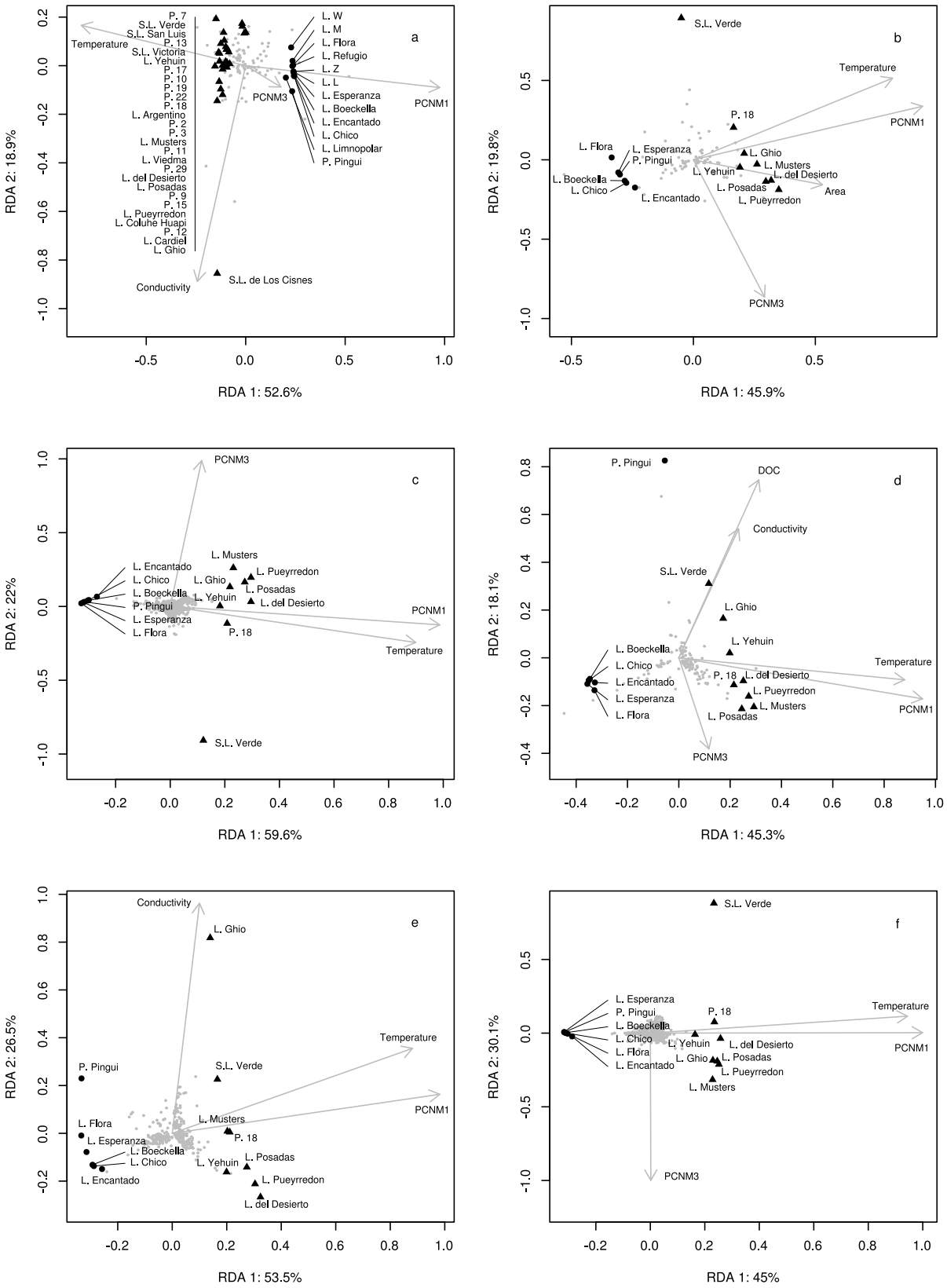

Figure B.7 – Redundancy analyses using community composition versus environmental and geographical (PCNM) variables when considering microbial communities from DGGE matrices n 5 40 (a), n 514 (b), total Illumina matrix (c), abundant ≥1% OTUs (d), uncommon >0.1% to <1% OTUs (e) and rare ≤0.1% OTUs (f). Antarctic lakes are represented by circles and Patagonian lakes by triangles. DOC: dissolved organic carbon.

Interestingly, we observed differences in taxonomic composition between Patagonian and Antarctic water bodies, being Cryptophyta frequent in Patagonia and absent from Antarctic lakes, whereas Chrysophyta domi-

nated in Antarctic lakes (fig. B.4). This result was also supported by microscopic observations (Izaguirre et al., 2016). Accordingly, (Charvet et al., 2012, 2014) found that Arctic lakes often contain many OTUs related to Chrysophyta, suggesting they are well adapted to cope with the low nutrient supply and strong seasonality that characterize Polar Regions. Indeed, it has been shown that at least some heterotrophic strains are specifically adapted to the Antarctica (Boenigk et al., 2007).

Multiple-sites dissimilarity measures of variation in OTU composition conducted on Illumina and DGGE dataset revealed that there is a high and significant dissimilarity in the OTU composition between any pair of lakes taken at random from our study site. These analyses also revealed for both datasets that variation in OTU composition is characterized by turnover and to a much lower extent to nestedness (fig. B.6). Most OTUs are replaced among lakes while very few OTUs co-occur among lakes. These results give further support to those shown by the Mantel test and variation partitioning analyses and suggest that each lake may have a unique and little explored microbial eukaryote community. It is worth mentioning that, although a percentage of OTUs is shared between Patagonian and Antarctic lakes (i.e. around 32% of the total Illumina matrix and 70% from DGGE data), turnover remains as an important phenomenon in driving variation in OTUs composition between any pair of lakes (i.e. OTUs are replaced between lakes). Therefore, these values of shared OTUs between Patagonian and Antarctic lakes should not be confused with the amount of dissimilarity that exists between any pair of lakes in terms of their OTU composition. Indeed, an OTU only needs to be present in one Patagonian lake and in one Antarctic lake to be considered a shared OTU. Moreover, the number of shared OTUs is a descriptive approach that does not say anything about the phenomenon characterising beta-diversity patterns (turnover or nestedness) along the studied gradient.

Mantel test and variation partitioning analyses showed that beta-diversity changed significantly along geographical and environmental gradients (fig. B.4, table B.3). The distance-based approach Mantel test (that provides an assessment of the significant interactions) and canonical variation partitioning (that allows quantification of the major sources of variation in the dependent variables using transformed raw data) are complementary approaches (Tuomisto and Ruokolainen, 2006). Particularly, these two approaches showed that environmental and geographical factors were important in shaping abundant/dominant microbial eukaryote communities, whereas only the geographical factors were important driving rare microeukaryote communities ($\leq$0.1%). These results point towards the existence of an abundant fraction of diversity ('core biosphere') ruled by standard macroecological rules, such as the latitudinal gradient of diversity and environmental filters. It coexists with a 'rare biosphere' that seems to be only weakly influenced by geographical position and not at all by environmental filters. Besides, variation partitioning showed that the percentage of variance explained in the communities decreased in the rarest fractions of diversity, suggesting higher randomness. The combined effect of environmental and geographical factors on the microbial community structure was also found by (Cottenie, 2005) and (Lovejoy and Potvin, 2011). (Soininen et al., 2011) found that freshwater planktonic communities may be controlled by both dispersal-driven assembly and local ecological determinism, with a scale-dependant balance between these two forces. In addition, (Lepere et al., 2013) and (Logares et al., 2014) also found a highly significant effect of geographical distance on both rare and abundant/dominant freshwater and marine microbial eukaryotes respectively.

Direct observation of microbial eukaryotes confirmed that autotrophic and heterotrophic eukaryote abundances ($\leq$5 µm in diameter) not only showed a latitudinal gradient (**??**, table B.2), but also their abundances increased with increasing chl α (table B.2). Similarly, (Rochera et al., 2013) found that the abundance of autotrophic picoplankton from 12 lakes of Byers Peninsula increased along a trophic gradient.

It is well known that in all aquatic ecosystems, the sequential filtration process allows the passage of cells larger than their nominal pore sizes and can lead to the retention of smaller cells if the filters are clogged (Diez et al., 2001). (Sorensen et al., 2013) found that the abundance of large protist (>10 µm as ciliate and dinoflagellate), often found among picoeukaryotes studies, correlated negatively with the pore size chosen for the end filter in the sequential filtration, suggesting that extracellular DNA adhering to small particles may be the source of larger protist OTUs in picoplanktonic size fractions. We observed also large ciliates by microscopy (such as *Stentor* sp.), which may have contracted and passed through the filters (Gabriela Küppers, personal communication). On the other hand, temporal samplings are important for adequate diversity and species richness estimates (Nolte et al., 2010). It is likely that a broader seasonal sampling would have increased the taxonomic diversity even further and possibly defined better the 'core and rare biosphere'. However, given the considerable amount of sampled lakes and the relatively broad scale studied (2100 km) in our work, we decided to take an integrated water sample in each lake during the same season to make them comparable. (Lepere et al., 2013) found that temporal variations in the composition of the small protist community do not affect the importance of geographical distance in explaining community composition, and they suggested that the use of a single sample per lake should be enough to analyse the spatial distribution of lacustrine protists.

Although the high-throughput sequencing approach brought obviously more information than DGGE on community composition, comparing both approaches has certain limitations. Indeed, the first PCR step used

different primers sets amplifying, respectively, the V9 and V1-3 regions of the SSU rRNA. It has been shown that primers amplifying these regions in eukaryotes are not entirely universal, and therefore, biases may be observed in the respective community compositions (Hadziavdic et al., 2014). Moreover, the taxonomic resolution of the different variable regions of the gene is not equivalent and varies from taxon to taxon (Hu et al., 2015). Nevertheless, in spite of these important methodological differences, a similar amount of variance could be explained with both Illumina abundant OTUs (≥1%) and DGGE (table B.3), even though patterns observed in the RDA analyses showed some differences (fig. B.7). In that sense, DGGE gave a picture of diversity comparable with the abundant fraction of the communities, and therefore does not lose its relevance as a fingerprinting technique (Pedros-Alio, 2006). Our data is consistent with that of (Pommier et al., 2010), who found that fingerprinting approaches such as DGGE clustered samples in a similar way than a high-throughput sequencing approach (in that case, pyrosequencing).

In summary, we found more abundant OTUs in Patagonian lakes (higher E_{H} evenness indices and higher dominant/abundant richness) than in Antarctic lakes, while total OTU richness did not change. Accordingly, the 'core biosphere' was significantly larger in Patagonian lakes than in Antarctic ones, while the 'rare biosphere' was larger in Antarctic lakes. In addition, the microbial eukaryote community composition (beta-diversity) changed significantly between Patagonian and Antarctic lakes and was mainly characterized by OTUs turnover. Both geographical position of the water bodies and the local environmental descriptors influenced beta-diversity (mainly abundant/dominant microbial eukaryote community composition), whereas only the geographical factors were important in shaping the rare OTU community composition. Thus, our results suggest that the latitudinal gradient becomes evident depending on the diversity component under evaluation (i.e. total richness, dominant richness, diversity indices and composition). These findings definitely shed new light on the biogeographical patterns and forces that structure inland microbial eukaryote composition across broad spatial scales. Growing information about microbial eukaryote diversity and distribution patterns makes protist the subject of active arguments over their ecology (Caron et al., 2009). However, microbial eukaryotes comprise an immense array of different sizes and lifestyles, and, even though some general conclusions can be drawn, the study of individual taxa may lead to diverging conclusions. Altogether, patterns in the spatial distribution of organisms provide important baseline information about mechanisms that regulate the diversity of life and the complexity of ecosystems (Green et al., 2004; Levin, 1992).

Experimental procedures

Studied lakes and samplings

Samples were collected once in 40 freshwater bodies (ranging from oligotrophic to eutrophic) from Chubut Province, Argentinean Patagonia, to Hope Bay, Antarctic Peninsula (45°22 to 63°24S of latitude) (fig. B.1). In Antarctic lakes, samples were taken during the austral summer 2004, whereas in Patagonian lakes, samples were collected in late spring 2007 and 2008. Integrated samples were collected within the euphotic zone from the surface down to 5 m in deep lakes and from about 30 cm below the surface in shallow lakes. Temperature, pH, conductivity and dissolved oxygen (DO) were measured in situ with portable meters (Horiba D-54 meter, Kyoto, Japan and Hanna HI 9146, Villafranca, Italy). Nutrients (Ammonium N-NH_4, nitrate N-NO_3, nitrite N-NO_2 and phosphate P-PO_4), phytoplanktonic chlorophyll α(chl α), dissolved organic carbon (DOC) and light diffuse attenuation coefficient (Kd) were analysed following the methods described in (Schiaffino et al., 2011).

A complete description of the studied area and the main characteristics of the water bodies were previously reported in different papers: (Quirós and Cuch, 1985; Quirós and Drago, 1999) for Patagonia, (Vinocur and Unrein, 2000) for Potter Peninsula, (Toro et al., 2007) for Byers Peninsula and (Izaguirre et al., 1998) for the Antarctic Peninsula (Hope Bay).The geographical positions of the studied water bodies, as well as their main limnological characteristics were detailed in (Schiaffino et al., 2011). The classification of the trophic status of the water bodies was based mainly on chl α taking into account the ranges presented in the (Wetzel, 2001) compilation.

Quantification of total abundances

Abundances of autotrophic and heterotrophic eukaryotes of picoplanktonic and nanoplanktonic size (0.2–5 µm) from each water body were obtained by epifluorescence, using an Olympus microscope (Olympus BX40F4, Tokyo, Japan) at 1000X magnification. Samples for epifluorescence microscopy were preserved with filtered cold glutaraldehyde 10% (1% final concentration) and filtered through a 0.2 and 0.6 µm pore-size polycarbonate filters. Autotrophic eukaryotes (≤3 mm in diameter) counts were performed using both blue and green-wavelength excitation and DAPI stained heterotrophic eukaryotes (flagellates of ≤5 µm in diameter) counts were performed using blue and UV-wavelength excitation.

Molecular methods

DNA extraction. Around 1000 ml of water samples were prefiltered in situ through a 50 µm net to remove zooplankton, then filtered with a vacuum pump first through a 20 µm pore-size polycarbonate filter and then through a 3 µm and 0.2 µm pore-size polycarbonate filters (diameter 47 mm; Millipore, Cork, Ireland). The filters were placed in cryovials with 1.8 ml of lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) and stored at -80°C until DNA extraction. Fractionation with 3 µm filters was chosen to be consistent with previous work (e.g. (Massana et al., 2004a; Vaulot et al., 2008)) focusing on microbial eukaryotes. The 0.2–3 µm size fraction was used for this study. The procedures followed for DNA extraction (phenol/chloroform extraction) and touchdown polymerase chain reaction (PCR) amplifications were previously described in detail (Unrein et al., 2005).

DGGE analysis. This approach was performed for a total of 40 Patagonian and Antarctic lakes. Results from this finger-printing method were presented showing the analysis of the entire dataset (n=40) and the analysis of the 14 (n=14) lakes. To perform DGGE amplifications we used the primers designed by (Diez et al., 2001) targeting the V1-V3 variable regions of 18S rRNA gene: Euk1F (50-AAC CTG GTT GAT CCT GCC AGT-30) and Euk516r-GC (50-ACC AGA CTT GCC CTC C-30) with a 40 bp GC-clamp. DGGE analyses were carried out with a DGGE-2000 system (CBS Scientific Company, Del Mar, CA). The 0.75 mm-thick gels of polyacrylamide (acrylamide:bisacrylamide ratio of 37:5:1) were run at 100 V and 60°C for 16 h in a linear 35 to 55% denaturant agent gradient (100% denaturant agent was defined as 7 M urea and 40% deionised formamide). The gels were stained for 45 min in 1X TAE buffer with Sybr-Gold nucleic acid stain (Invitrogen, Grand Island, NY) and visualized with UV radiation using a Chemidoc system and the Quantity One software (Bio-Rad, Hercules, CA). Samples were run in 3 DGGE gels and about 800 ng of PCR product from environmental samples were applied to individual lanes in the gels. Between 2 and 4 samples were run in all the DGGEs to allow comparison among the gels; thus, respective positions of individual bands in different gels could be determined. Digitized DGGE images were analysed using the Gel-Pro 4.0 software (National Institutes of Health, Bethesda, MD). Changes in band intensities are likely due to relative changes in the abundances of the corresponding populations. With DGGE approach we obtained the dominant OTUs.

DGGE bands were removed from the gel and resuspended in 20 µl of Milli-Q water. The PCR reamplifications were performed with the original primer set and the PCR products were purified with the QIAquick PCR Purification kit (QIAGEN, Hilden, Germany). The reamplified PCR products were used for a sequencing reaction with the corresponding forward primer in Macrogen Sequencing Service (Macrogen, Republic of Korea). The sequences obtained (around 500 bp) were screened for chimeras with KeyDNATools and then compared with the curated ribosomal eukaryotic database PR2 (Guillou et al., 2013). All 18S rRNA gene sequences obtained in this study were deposited in GenBank under accession numbers KC923040-KC923068.

Illumina HiSeq. To further explore the diversity distribution patterns, a subset of 14 water bodies (8 Patagonian and 6 Antarctic lakes) from the entire dataset (40 lakes) were analysed with this technology. We amplified extracted DNA using primers specific to the V9 variable region of the 18S rRNA gene using the protocol as in (Amaral-Zettler et al., 2009), and adapted after (Lara et al., 2015). Sequencing was performed by the company Fasteris (Geneva, Switzerland) using Illumina HiSeq 2500 technology; paired end reads were around 200 bp in length.

Quality check (Phred score filtering, trimming of the primers and chimera removal) of the sequences was performed following the pipeline developed by de (De Vargas et al., 2015); Briefly, we kept sequences that passed the filtering based on quality values from the paired fastq file, by evaluating the expected error in a 50 bp sliding window and discarding sequences with more than 1% of error in the worst quality window. Sequences were clustered using SWARM (Mahé et al., 2014). The obtained sequences were then taxonomically classified with the curated ribosomal eukaryotic database PR2 (Guillou et al., 2013) using the software GGSearch (McWilliam et al., 2013). Sequences affiliated with Metazoa and Embryophyta were removed from the analysis. Sequences were also aligned against the SILVA bacterial and archaeal database to remove possible prokaryotic OTUs.

Data analyses

The Illumina community matrix was studied considering the total OTUs, the abundant OTUs 1% of the total number of sequences in a given sample (Fuhrman, 2009; Pedros-Alio, 2006) and the less frequent OTUs with two different criteria: uncommon $>0.1\%$ to $<1\%$ and rare $\leq 0.1\%$ (Fuhrman, 2009; Vergin et al., 2013). The justification for the use of these thresholds is arbitrary, and lies in studies that indicate that organisms that make $\geq 1\%$ of the total cell number can be detected with PCR-dependent techniques, while those that comprise $\leq 0.1\%$ are difficult to retrieve (Casamayor et al., 2000; Muyzer et al., 1993). In order to compare the different samples, we randomly selected the same number of Illumina sequences from each lake (matching the sample with the lowest number of reads, i.e. 14 437 sequences after primers trimming and quality filtering).

To study the diversity patterns, we calculated two different diversity indices for each sample using the total OTUs from the Illumina data: Shannon's index (H'=$\sum pi\ \ln(pi)$) and Simpson's reciprocal index (D=$1/\sum/pi^2$), were pi is the relative abundance of each OTU (n/N). In addition, two different evenness indices were also calculated based on the Simpsons reciprocal index (E_{D}=D/S) and the Shannons index ($E_{\mathrm{H'}}$=$H'/\ln S$, Pielou index), where S is the OTU richness. To estimate total richness, we computed the Chao 1 index (Chao 1 = S_{obs} + ($f1^2/2f2$), where S_{obs} is the observed number of OTUs, $f1$ is the number of singleton OTUs (represented by a single Illumina read), and $f2$ is the number of doubleton OTUs (Chao, 1984). Furthermore, to determine whether sampling was deep enough to get a reasonable estimate of OTU richness, we performed rarefaction curves with the rarefaction function R package vegan v. 2-3, 2. The latitudinal diversity pattern was assessed by correlating (Spearman Rho correlations) the diversity components (richness and evenness) and indices recorded in each lake with latitude. In addition, to compare the responses of the individual diversity components, we performed Spearman Rho correlations between them. Student's t-tests were used to compare the differences between Patagonian and Antarctic datasets (abundances, richness, evenness and diversity indices). Some variables were LN-transformed to meet normality and homoscedasticity assumptions.

We computed multiple-sites dissimilarity measures of variation in OTUs composition (Baselga, 2012) based on Illumina and DGGE community matrices to explicitly assess the spatial patterns of beta-diversity and their underlying phenomena (pure turnover, pure nestedness, or a combination of both with one of them prevailing over the other) in the model groups under study. In this approach, beta-diversity (β_{SOR}) is first calculated and then partitioned into its spatial turnover (β_{SIM}) and nestedness-resultant dissimilarity (β_{NES}) components. β_{SOR} and β_{SIM} are represented by the dissimilarity values retrieved through the calculation of Sørensen and Simpson dissimilarity indices respectively; while β_{NES} is computed as the difference between these metrics (*op. cit.*). These metrics range from 0 to 1, where higher values are consistent with higher values of dissimilarity among sites. Differences in the number of lakes among the areas studied were controlled for by re-sampling four lakes from each area 1000 times and computing the average β_{SOR}, β_{SIM} and β_{NES} respectively. Multiple-site dissimilarity measures of beta-diversity were computed using the R package betapart (Baselga and Orme, 2012).

We finally determined if environmental parameters and/or geographical distance influenced microbial eukaryote community composition (beta-diversity), and if so, which one had the strongest influence. We also wanted to know if the three abundance categories (abundant $\geq$1%, uncommon $>$0.1% to $<$1% and rare $\leq$0.1%) were influenced by the same parameters and if these parameters explained the same amount of variance. We also treated DGGE data in order to evaluate its performance with Illumina data, comparing the patterns using DGGE and the abundant part of the Illumina reads. We first transformed OTU abundance data using a Hellinger transformation prior to applying redundancy analysis (RDA) (Legendre and Gallagher, 2001; Ramette, 2007). Geographical coordinates were transformed into Principle Coordinates of Neighbourhood Matrix (PCNM; (Borcard and Legendre, 2002). Raw environmental variables were transformed to their standard normal deviate equivalents [(x—mean) divided by the SD] to accommodate the different units of the different variables (Legendre and Birks, 2012). Relevant variables were selected for each abundance category and DGGE data by forward selection based on RDA (ordistep function, R package vegan v. 2-3, 2). In order to assess the relative contribution of each factor (environmental vs. geographical descriptors), we performed a variation partitioning based on partial RDA (varpart R package vegan v. 2-3, 2) on each community matrix (DGGE data and each abundance category from Illumina), also based on the variables that were kept after forward selection. We also performed standard and partial Mantel tests to examine the respective effect of geographical distance versus environmental factors on community composition; this distance-based approach is efficient in analyzing spatial patterns in community similarity and this is an alternative way to assess beta-diversity (Legendre et al., 2005; Soininen et al., 2007, 2011; Tuomisto and Ruokolainen, 2006). We calculated Bray Curtis distance matrices for community matrices (each abundance category from Illumina and DGGE datasets) and Euclidean distance matrices for the standardised environmental parameters (Martiny et al., 2006) and the PCNM variables that were selected previously.

Acknowledgements

This work was supported by a grant from the Argentinean Funds for Technical and Scientific Investigation (FONCYT, PICT 32732 and PICT 2013-0794), the Spanish Project MIXANTAR (REN 2002-11396-E/ANT), the Swiss NSF project 310003A_143960 and the Program 'Luis Santalo' of the National Research Council of Spain and the National Council of Scientific and Technical Research of Argentina (CSIC-CONICET, PROBA 2007AR0018). The authors also thank the members of the Antarctic Esperanza Station for logistic support and Dr A. Quesada and his team for providing the samples from the Byers Peninsula. L.D.F. is supported by CONICYT (doctoral scholarships N8 21110037 and 78130011) and the Wildlife Conservation Society (Karukinka grant N8 2012).

Appendix C

Response of forest soil euglyphid testate amoebae (Rhizaria: Cercozoa) to pig cadavers assessed by high-throughput sequencing

Christophe V.W. Seppey[1], Bertrand Fournier[1,2], Ildikò Szelecz[1,3], David Singer[1], Edward A.D. Mitchell[1,4], Enrique Lara[1]

[1]Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland
[2]Evolutionary Community Ecology Group, CNRS, University of Montpellier 2, 34095 Montpellier Cedex 05, France
[3]nstitute of Forensic Medicine, Goethe University, Kennedyallee 104, 60596 Frankfurt, Germany
[4]Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, 2000 Neuchâtel, Switzerland

International Journal of Legal Medicine, Vol. 130(2), p.645-655, March 2016
DOI 10.1007/s00414-015-1149-7

Abstract

Decomposing cadavers modify the soil environment, but the effect on soil organisms and especially on soil protists is still poorly documented. We conducted a 35-month experiment in a deciduous forest where soil samples were taken under pig cadavers, control plots and fake pigs (bags of similar volume as the pigs). We extracted total soil DNA, amplified the SSU ribosomal RNA (rRNA) gene V9 region and sequenced it by Illumina technology and analysed the data for euglyphid testate amoebae (Rhizaria: Euglyphida), a common group of protozoa known to respond to micro-environmental changes. We found 51 euglyphid operational taxonomic units (OTUs), 45 of which did not match any known sequence. Most OTUs decreased in abundance underneath cadavers between days 0 and 309, but some responded positively after a time lag. We sequenced the full-length SSU rRNA gene of two common OTUs that responded positively to cadavers; a phylogenetic analysis showed that they did not belong to any known euglyphid family. This study confirmed the existence of an unknown diversity of euglyphids and that they react to cadavers. Results suggest that metabarcoding of soil euglyphids could be used as a forensic tool to estimate the post-mortem interval (PMI) particularly for long-term (>2 months) PMI, for which no reliable tool exists.

Keywords: Environmental DNA; Euglyphid testate amoebae; Illumina high-throughput sequencing; Metabarcoding; SSU rRNA gene V9 region; Forensic ecology

Introduction

The estimation of time since death and more generally the ability to detect the presence of cadavers even when the remains are no longer present are the two major objectives in forensic research (Prangnell and McGowan, 2009; Rodriguez and Bass, 1985).

Calculation of the post-mortem interval (PMI), an essential element of legal medicine and criminal investigation to establish the timing of events that led to the death of a person, becomes less precise with the advance of the decomposition process. Until now, two main approaches are used to estimate the PMI. The medical method provides information ranging from a few hours to several days after death (Amendt et al., 2004; Henssge et al., 2000; Wyss and Cherix, 2006). The second method, forensic entomology, is based on the observation of larval stages of necrophagous flies and beetles and can be used to establish a PMI up to some weeks (Amendt et al., 2004; Wyss and Cherix, 2006). Although well established, the accuracy of entomological methods has been questioned after the full development of the first generation of necrophagous insects (Wyss and Cherix, 2006). Therefore, the development of additional indicators for PMI estimates beyond 1 month would constitute a welcome addition to the toolkit of forensic criminal investigators.

According to (Payne, 1965), the decomposition of cadavers can be separated into six stages: fresh, bloated, active decay, advanced decay, dry and remain stages. During the 'bloated' and 'active decay' stages (Carter et al., 2007; Payne, 1965), the release of cadaver liquids into the soil changes the chemical parameters drastically (Vass et al., 1992). This perturbation of the soil environment has been referred to as 'ephemeral resource patches' (Barton et al., 2013) leading to the development of 'cadaver decomposition islands' (CDI) (Carter et al., 2007). Although most of the decomposition takes place in the first few weeks under optimal conditions, cadaver effects on the soil environment can be long lasting. For example, (Towne, 2000) showed that nitrogen and phosphorus concentration and pH were still significantly enhanced in soil samples taken under cadavers 2 years after laying ungulate cadavers on a prairie, while (Melis et al., 2007) reported enhanced soil calcium content and pH as late as 7 years post mortem in a CDI. Such environmental changes were shown to have an effect on the soil fauna (Bornemissza, 1957), bacteria (Horswell et al., 2002; Howard et al., 2010; Moreno et al., 2011) and fungi (Carter and Tibbett, 2003; Hawksworth and Wiltshire, 2011). However, knowledge about cadaver effects on soil communities remains very limited, and almost nothing is known about the response of soil protists (Szelecz et al., 2014).

In this study, we focused on euglyphid testate amoebae (Rhizaria: Cercozoa), a highly diverse and abundant group of protists that reacts rapidly to environmental changes by shifts in community structure and abundance (Foissner, 1999b). Euglyphids include about one quarter of the ca. 300 testate amoeba morphospecies known to occur in soils (Foissner, 1999b). These amoeboid unicellular protists range mostly between 20 and 150 µm in length, and their densities typically range between ca. 10^6 and 10^8 individuals per square metre (Foissner, 1987). They build a shell (test) reinforced with ornamented self-secreted siliceous plates, and these shells allow species identification even after the death of the organism (Meisterfeld, 2000a,b). Most euglyphids are heterotrophs and feed mainly on bacteria and fungi (Daniel et al., 2000). The distribution patterns of soil testate amoebae along environmental gradients and their response to environmental changes have been well studied, including soil humidity (Bobrov et al., 1999; Booth, 2008; Swindles and Ruffell, 2009); temperature (Beyens et al., 2009; Tsyganov et al., 2011); pH (Booth, 2001; Charman, 2001; Mitchell et al., 1999, 2008); and pesticide (Petz and Foissner, 1989), nitrogen, phosphorus and sulphate concentration (Gilbert et al., 1998b; Mitchell, 2004;

Payne et al., 2010). They can thus be expected to also respond to the presence of decomposing cadavers. The generation time of euglyphids, which ranges from ca. 2 days to 1 week under natural conditions (Heal, 1964), is considerably longer than that of bacteria or smaller protozoa such as nanoflagellates, and this represents an advantage regarding their use as bioindicators. It is indeed short enough to allow them to (re)colonise rapidly suitable habitats (Wanner and Elmer, 2009; Wanner et al., 2008) and respond to environmental change over a period of weeks. However, as euglyphids are highly sensitive to environmental conditions, the effects on communities can be expected to be long lasting under continuous environmental stress. So especially for estimating longer PMIs, euglyphids might be a group to consider for forensic applications.

However, a current limitation to the development of euglyphid analysis (or that of other soil protists) as a standard tool for PMI estimates is taxonomy. Sound taxonomy is indeed a prerequisite for the use of a group of organisms as bioindicators. Up to now, all ecological studies on testate amoebae were based on morphology-based species identifications. The morphological identification of testate amoebae requires taxonomic expertise and is time-consuming. Furthermore, recent molecular taxonomy studies on euglyphids have revealed the existence of a substantial higher diversity than estimated based only on morphology (Chatelain et al., 2013; Heger et al., 2011b), and this hidden/unknown diversity may prove to have bioindication value. The molecular approach presented in this study overcomes the current limitation of morphology based taxonomy and is also faster (i.e. weeks instead of months for the number of samples analysed here).

Analysis of environmental samples (e.g. soil, water, faeces) targeting a specific DNA barcode gene and aiming at characterising the entire community is referred to as metabarcoding (Taberlet et al., 2012). The V9 region of the 18S rRNA gene has sufficient variability for obtaining reasonably high taxonomic resolution (Adl et al., 2014; Amaral-Zettler et al., 2009) and two main advantages for biodiversity surveys as follows: (1) it is short and thus likely to be well preserved in environmental DNA samples and (2) it contains highly conserved sites allowing to designing primers for virtually all eukaryotes (Valentini et al., 2009). The advent of high-throughput sequencing (HTS) now allows using the metabarcoding approach in ecological studies with high sample numbers (e.g. spatial and/or temporal sampling).

In this study, we used a DNA metabarcoding approach applied to the V9 region of the 18S rRNA gene to assess the temporal response of soil euglyphid testate amoebae to decomposing cadavers over a period of 35 months (1051 days). Given the sensitivity of the technique and the well-documented response of testate amoebae to ecological gradients, changes and disturbances, we expected to find (1) a higher diversity of soil Euglyphids than generally inferred from microscopic analyses and (2) a strong generally negative response of communities to decomposing cadavers with i) rapid disappearance of the majority of taxa following the massive release of cadaver fluids in the soil and ii) slow recovery after the end of the active decay phase, ca. 1–2 months after the peak of cadaver fluid input in the soil. If such patterns were indeed found, this may lead to the development of new PMI indicators in the future.

Materials and methods

Sampling site

The experimental site is situated in a beech- (*Fagus sylvatica*) and oak- (*Quercus robur, Q. petraea*) dominated forest near Neuchâtel (Switzerland 47°00' N; 06°56' E, elevation 478 m). The overall average temperature measured over the course of the experiment was 10.4 °C (SD 6.09 °C). The mean annual precipitation of the nearest meteorological station (Neuchâtel) for 1993 to 2013 was 974 mm per year (MétéoSuisse).

Table C.1 – Sampling dates and corresponding decomposition stages of the pig cadavers in a *Fagus-Quercus* forest above Neuchâtel, Switzerland

Sample code	Sampling date	Decomposition stage
D0	05.08.2009	Fresh
D8	13.08.2009	Active decay
D15	20.08.2009	Dry stage
D22	27.08.2009	Dry stage
D33	07.09.2009	Dry stage
D64	08.10.2009	Dry stage
D309	10.06.2010	Dry stage
D1051	21.06.2012	Dry stage

Experimental setup

The experiment included three treatments: control (plots of forest soil left under natural conditions), fake pig (plastic bags filled with a volume of soil similar to that of the pigs placed in a cotton cloth) and pig (*Sus scrofa*). The fake pig treatment was used to differentiate the chemical effect of pig cadaveric liquids from the physical effects (i.e. humidity, soil compaction) due to the presence of a carcass on the soil. The bag volume was kept approximately similar to that of the pigs by removing soil from the bag to mirror the volume loss of the pig cadavers over time. The pigs (20 kg±1 kg) were killed on the farm with captive bolt stunning and the cadavers immediately brought to the experimental site. The cadavers were placed in

strong metal wire cages (90×100×50 cm) to protect them from scavengers. The cages also allowed moving the cadavers for sampling. Pigs are commonly used in comparable forensic studies due to the similarities with humans, comparable thoracic cage size and almost naked skin (Stokes et al., 2013). Each treatment was run in triplicate. The sampling plots were organised into three randomised blocks (15–34 m apart). Within each block, the plots were at least 4 m apart.

Sampling and chemical analyses

Eight sets of samples were collected from the onset of the study (August 5th, 2009=D0) until June 21st, 2012 (table C.1). At the onset of the experiment (D0, before the pigs and fake pigs were placed), initial control samples were collected from all sampling plots and pooled for each block (i.e. three pooled samples in total). Sampling days were scheduled according to decomposition stages (table C.1) (Payne, 1965). On each sampling day, ca. 25 g of soil was taken to a depth of 10 cm in each plot and stored at -80 °C.

Soil subsamples (3 g) were dehydrated (40 °C, 12 h), ground to powder and analysed for total organic carbon (Soil_C) and total nitrogen (Soil_N) using combustion infrared spectroscopy (CHNEA1108-Elemental analyser, Carlo Erba Instrument) after decarbonation with HCl (Harris et al., 2001).

Molecular analyses

DNA was extracted from soil samples using the MoBio PowerSoil DNA Isolation Kit (Carlsbad, CA, USA) following the manufacturer instructions. The SSU rRNA V9 region was amplified by PCR using the specific eukaryotic primers 1380f/1510r (CCCTGCCHTTTGTACACAC/CCTT CYGCAGGTTCACCTAC) (Amaral-Zettler et al., 2009). Forward primers were tagged on the 5' end with a 10 nucleotides strand, specific to each sample. PCR reactions were run in triplicate with a PTC-200 Peltier Thermo Cycler (BioConcept, Allswill, Switzerland) with 1 ng of environmental DNA, 6 µL of 10x PCR buffer, 0.6 µl of each primer, 0.6 µl of each dNTP 400 µM (Promega) and 0.2 µl of 0.05 U/µl Go Taq (Promega). The volume was adjusted to 30 µL with ultra-pure water. Amplification was conducted with the following conditions: denaturation at 94°C for 3 min, 30 cycles at 94°C for 30 s, 57°C for 60 s and 72°C for 90 s and final extension at 72°C for 10 min (Amaral-Zettler et al., 2009). PCR products were purified through QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and pooled together with a 4 ng DNA of each sample. A DNA library was prepared using the New England Biolabs's kit NEBNext DNA Sample Prep Master Mix Set 1 (http://www.neb.com/nebecomm/ManualFiles/manualE6040.pdf) except for the size selection step. Sequencing was done by the Genomics Core Facility at Brown University (Providence, USA) with an Illumina ® HiSeq 2000 sequencer to obtain paired-end reads covering the full length of the V9 region.

Sequence treatment

A database was constructed by selecting 44 complete euglyphid V9 sequences from the GenBank database, using sequences derived both from identified organisms and from related environmental sequences retrieved from GenBank. Each environmental V9 read was compared to the database using the BLASTn algorithm (Altschul et al., 1997) in order to select euglyphid sequences. Before the BLASTn, each nucleotide with a Phred score below 28 was changed to an unknown nucleotide 'N' in order to avoid unreliable nucleotides. The BLASTn algorithm was setup with a match/mismatch ratio of 1-1, gap open and extend penalty, respectively, of 0 and 2 and a word size of 32 nucleotides.

Table C.2 – Summary of the sequence filtering of euglyphid testate amoeba from the control, fake pig and pig cadaver plots in a *Fagus-Quercus* forest above Neuchâtel, Switzerland

Analysis steps	Total reads	Euglyphid sequences	Euglyphid dereplicated reads	Euglyphid OTUs
Raw fastq	247366905	–	–	–
Blast selection	187566	187566	–	–
Reads>=5 times	57533	57533	2621	–
OTU building	57724	57724	2621	198
OTU selection	52860	52860	–	51
Triplication D0	57640	57640	–	51

We used an empirically determined e-value threshold as the criterion for classifying a read as belonging to the euglyphids. To determine the appropriate e-value threshold, a subset of eukaryotic V9 sequences (sample D309, block 2, pig treatment) was compared by BLASTn to the previously established euglyphid V9 sequence database with a permissive e-value (i.e. 10). The hit results were sorted by increasing e-value and compared

Table C.3 – Summary of total euglyphid testate amoeba OTU abundance in the control, fake pig and pig cadaver plots in a *Fagus-Quercus* forest above Neuchâtel, Switzerland

Euglyphid OTUs	Total abundance	Euglyphid OTUs	Total abundance
eugly_59	4234	eugly_60	417
eugly_13	4205	eugly_290	411
eugly_2	4149	eugly_991	402
eugly_12	3873	eugly_98	382
eugly_666	3161	eugly_320	367
Euglypha rotunda AJ418783.1	3056	eugly_992	349
Euglypha filifera AJ418786.1	2583	eugly_16	314
eugly_66	2530	eugly_473	273
Uncultured eukaryote EF025028.1	2279	eugly_862	253
eugly_81	2048	eugly_82	238
eugly_151	1949	eugly_76	211
eugly_156	1933	eugly_80	203
eugly_5	1916	eugly_973	195
eugly_136	1630	eugly_233	182
eugly_183	1451	eugly_1245	177
eugly_322	1400	eugly_854	176
eugly_38	1227	*Tracheleuglypha dentata*	176
eugly_307	1199	eugly_282	172
eugly_54	1186	eugly_885	170
Assulina muscorum AJ418791.1	1162	eugly_371	165
eugly_113	1086	eugly_250	141
eugly_234	946	eugly_120	134
Cercomonadida env sample EF024983.1	858	eugly_1777	105
eugly_79	675	eugly_1716	96
eugly_33	594	eugly_1890	87
eugly_41	514		

to the GenBank database by BLASTn, using the previous setup, until sequences corresponding to taxa other than euglyphids were found. Once the e-value threshold was found (i.e. $8e^{-29}$), each environmental sequence was compared against the V9 euglyphid database using BLASTn.

Only sequences over 130 nucleotides long and occurring at least five times in the 66 samples were retained, in order to remove possible false-positive sequences. As our database showed that some closely related but nevertheless morphologically and genetically (e.g. COI gene or full SSU) distinct euglyphid morphospecies shared exactly the same V9 sequence (e.g. *Euglypha penardi* (EF456753) and *Euglypha* cf. *ciliata* (EF456754) (Lara et al., 2007c)), we considered each unambiguous difference in the nucleotide sequence as sufficient for discriminating two OTUs. Conversely, when two sequences differed only in ambiguous nucleotides, they were considered as belonging to the same OTU. The resulting OTU sequences were then counted in each sample.

Numerical analyses

We assessed the response of the 51 OTUs found in the 66 samples to the different treatments using partial redundancy analysis (RDA) on Hellinger-transformed data (Legendre and Gallagher, 2001) with the blocks used as conditional variable. Rare OTUs (present less than three times in a minimum of ten samples) were removed to reduce noise in the model and optimise the adjusted R^2 (Borcard et al., 2011). These thresholds were selected after testing several options (presence threshold 1, 3, 5, 7; minimum number of presence 7, 10, 12, 14). The significance of variables (Soil_C, Soil_N, treatment) and ordination axes (first, second and third) were assessed using Monte Carlo tests (999 permutations, p value threshold=0.05).

We assessed the effect of the treatments, relative to control, on the OTU responses over time using a principal response curve (PRC) (Van den Brink and Ter Braak, 1999). The model was also tested using a Monte Carlo procedure (999 permutation, p value threshold=0.05).

All statistical analyses were performed with R-2.13.1 (R Core Team, 2013) using package 'vegan' (Oksanen et al., 2012) for the Hellinger transformation, RDA and PRC analyses.

Retrieval of full-length SSU rRNA gene sequences of selected taxa and phylogenetic analysis

Because sequences of the V9 variable region of the SSU rRNA gene are short (i.e. generally less than 200 bp), they are not suited for inferring the position of OTUs in phylogenetic trees. This is especially problematic if the considered sequences are suspected not to cover a large part of the diversity of the group of interest (Dunthorn et al., 2014). In order to place the OTUs of interest (i.e. showing a strong response to cadavers) in a phylogenetic tree, we used the sequence information included in the V9 region to design specific reverse primers and amplified the rest of the SSU rRNA gene. We designed specific primers to amplify specifically two phylotypes that responded positively to the pig treatment: eugly_13R (CACGAACTGAAGGCAAGCCCA) and eugly_666R (TTCACTTCCAATCACAG GAG). The newly designed primers were used in combination with the euglyphid specific forward primer Eugly1SSUF (GCGTACAGCTCATTATATCAGCA (Chatelain et al., 2013)) located at the beginning of the SSU rRNA gene. DNA extractions, where the OTU was most abundant, were selected for specific amplification of the SSU rRNA gene of interest. Cycling profile was the same as described above (with 40 cycles). PCR products were cloned into pCR2.1 Topo TA cloning vector (Invitrogen) and used to transform E. coli TOP10' OneShot cells (Invitrogen) according to the manufacturer's instructions. Up to five clones per PCR product were chosen for sequencing. Sequencing was performed with an ABI-3130xl DNA Sequencer (Applied Biosystems). The new sequences obtained were placed into an alignment that comprised all euglyphid sequences retrieved from GenBank, which included both environmental clones and sequences derived from identified organisms. The alignment was performed using MUSCLE (Edgar, 2004). A maximum likelihood tree was built using the RAxML v7.2.8 algorithm (Stamatakis, 2006) as proposed on the portal (http://phylobench.vitalit.ch/raxml-bb/) using a general time-reversible model. Rate heterogeneity was estimated using a CAT model.

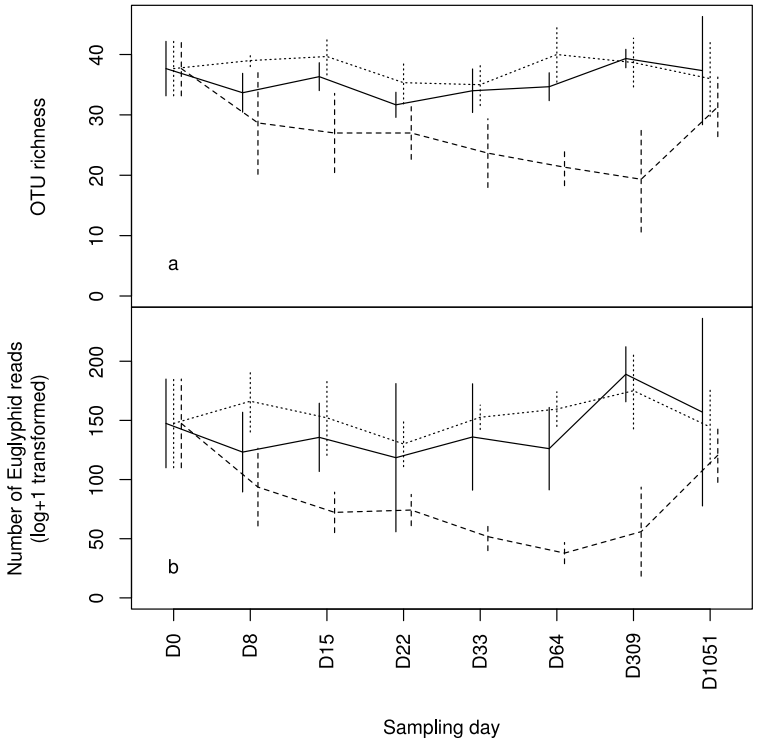

Figure C.1 – Temporal patterns of euglyphid OTU richness (a) and number of reads log +1 transformed (b) in soil samples from control, fake pig and pig cadaver plots in a *Fagus-Quercus* forest above Neuchâtel, Switzerland. Treatments are represented by line type (plain: control, dashed: pig, dotted: fake pig). The vertical lines show the standard deviation of the richness and number of reads for each treatment and sampling date. The lines for the three treatments are slightly offset to improve readability

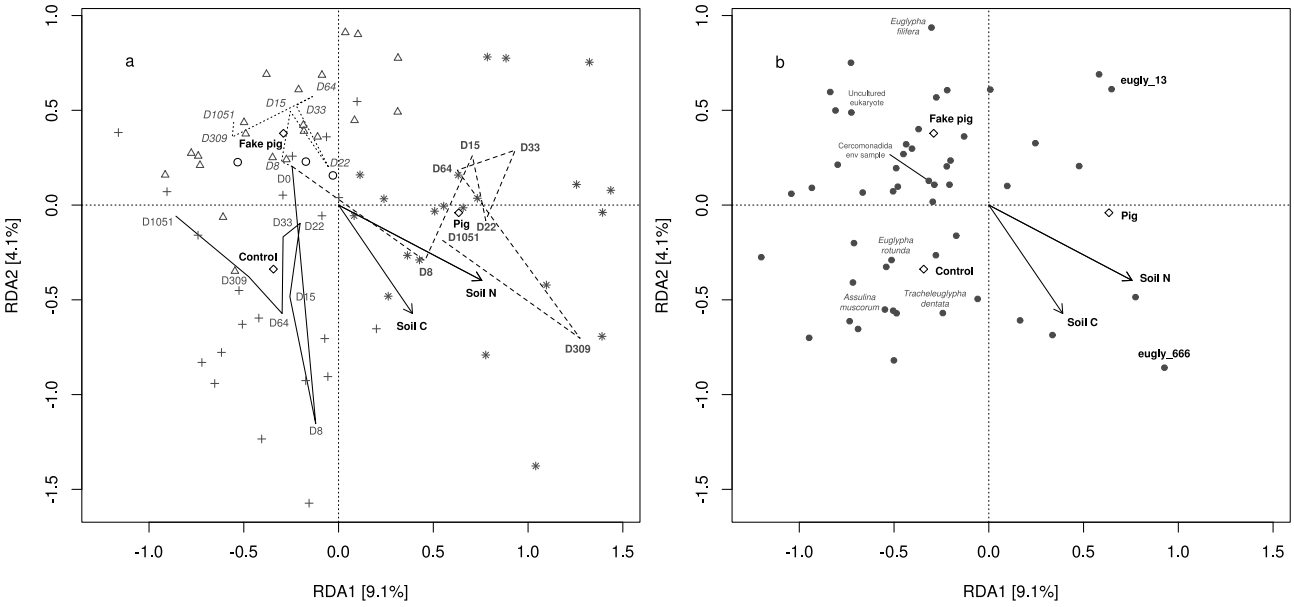

Figure C.2 – Partial redundancy analysis (RDA) ordination diagram showing the temporal patterns of soil euglyphid testate amoeba communities (OTUs) in control, fake pig and pig cadaver plots in a *Fagus-Quercus* forest above Neuchâtel, Switzerland. Diamonds represent treatment centroids, and arrows represent weight percentage of total organic carbon (soil C) and total nitrogen (soil N). Percentages of variance explained by axes are shown in brackets. In a, successive sampling dates for each treatment (days 0, 8, 15, 22, 33, 64, 309, 1051) are connected by lines. The line corners correspond to the centroid (average for RDA1 and RDA2 coordinates) of three samples of the same treatment and same sampling day. Treatments are indicated by line type (plain: control, dashed: pig, dotted: fake pig), fonts of the sample labels (plain: control, bold: pig, italic: fake pig) and symbols ('+': control, 'triangle': fake pig, 'stars': pig). Empty circles represent day 0 for each block. In b, OTU responses are represented by dots. The two OTUs for which full SSU sequences were obtained are represented in bold. The OTUs with a perfect match with a database sequence are represented with their names

Results

Diversity and structure of euglyphid OTU assemblages

Of the 247,366,905 raw Eukaryote reads, 187,566 were identified as euglyphids and 57,533 of these were found at least five times overall (table C.2). These 57,533 reads were divided into 198 OTUs. Of these, 51 OTUs respected the thresholds and were thus retained for further analyses. Six of these OTUs matched exactly with sequences from our database. Total OTU abundance data are summarised in table C.3, and OTU richness and number of euglyphid reads along time for the three treatments are shown in fig. C.1.

The partial redundancy analysis (fig. C.2) with the blocks used as conditional variable revealed a significant correlation between euglyphid communities and Soil_N and Soil_C (Monte Carlo test, 999 permutations, both p=0.01). Axes 1 and 2 were significant (p=0.005 for both). The RDA ordination showed that the pig treatment samples diverged from the control and fake pigs along the soil nitrogen content gradient until day 309 after which they converged again with the samples of the other two treatments. The RDA also showed that most OTUs responded negatively to the pig treatment. However, some OTUs responded positively to the pig treatment (e.g. eugly_13, eugly_666).

The principal response curve (fig. C.3) summarises the treatment effects on OTUs over time and shows the average responses of individual OTUs. The first PRC axis explained significantly ($p<0.03$) 42% of the model variance, while time and treatments explained, respectively, 10 and 27% of the variance. Qualitatively, the PRC diagram showed an overall negative effect of the pig treatment (D8 to D1051) on the majority of euglyphid OTUs and the positive response of a few OTUs, especially eugly_666 and eugly_13, which were therefore further studied.

Retrieval of full-length SSU rRNA gene sequences of eugly_13 and eugly_666

All clone sequences obtained were identical (n=5 and 8, respectively). Phylogenetic analyses confirmed the position of the two phylotypes within euglyphid testate amoebae (supported with 100% bootstrap value) and showed that they did not belong to any barcoded family (fig. C.4) (Heger et al., 2010). They were basal to all

known euglyphid families. Eugly_13 branched robustly (80% bootstrap) with an environmental sequence from fresh-water sediments (freshwater 13_2.2 AY620297). By contrast, eugly_ 666 did not branch robustly with any sequence—be it from environmental samples, cultures, or isolated cells.

Discussion

Euglyphid community responses to decomposing pigs

This study showed that the presence of decomposing pig cadavers significantly affected the community structure of euglyphid testate amoebae, showing a drastic decrease in sequence abundance and in OTU richness (fig. C.1). This result is in agreement with our general working hypothesis. The negative effect of a cadaver on euglyphid communities was correlated to the large input of nitrogen and organic carbon in the soil. This result was consistent with previous studies, which show that inputs of nitrogen strongly and negatively influenced testate amoeba communities (Gilbert et al., 1998a,b; Mitchell, 2004). It is probable that most euglyphids died because of anoxic conditions, but a direct or indirect effect of high nitrogen content is also possible. However, two well-represented OTUs, namely eugly_ 13 and eugly_666 (eugly_991 also responded positively but was less abundant), responded positively to the presence of cadavers, but only in the late decomposition stage (i.e. after 1 month to 1 year). These OTUs were present but rare at the beginning of the decomposition process as well as in the control and fake pig treatment, but their abundance peaked, respectively, at D33 and D309 in the pig treatments only and in the three replicates simultaneously (fig. C.5). This suggests that they did not benefit from the initial perturbation brought by the release of cadaveric fluids but rather found optimal conditions (i.e. abiotic, e.g. soil water chemistry, and biotic, e.g. prey and/or low level of competition or predation) for their growth in later stages. These organisms probably benefited from changes in the bacterial communities, as these are supposed to change deeply and progressively underneath a cadaver (Allison and Martiny, 2008; Fierer et al., 2012). Indeed, previous studies have shown that decomposing carcasses cause an increase in soil bacterial biomass (Barton et al., 2013) but also drastically change bacterial community structure (Howard et al., 2010; Moreno et al., 2011). As bacteria constitute a large part of euglyphid food regime (Daniel et al., 2000), any change in the abundance or community structure of bacteria is likely to also influence the abundance and community composition of euglyphids. It may also be that these taxa represent nutrient-tolerant organisms that benefit well from high abundance of prey organisms, but with low competitive ability in the normally more oligotrophic conditions. The precise mechanism for this response however remains to be elucidated.

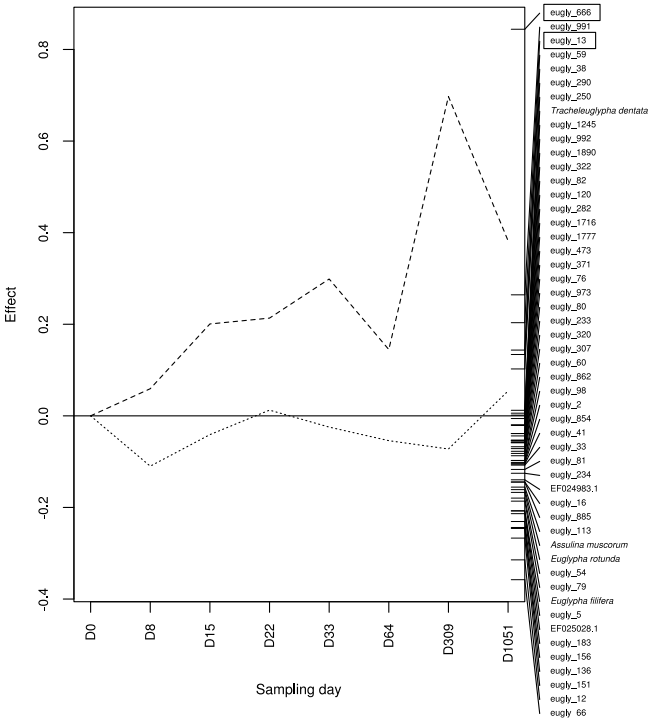

Figure C.3 – Principal response curve (PRC) diagram showing the effects of pig (dashed line) and fake pig (dotted line) treatments relative to control treatment over time on soil euglyphid testate amoeba communities in a *Fagus-Quercus* forest above Neuchâtel, Switzerland. The average response of individual OTUs is shown on the right axis. The two OTUs for which full SSU sequences were obtained are framed. GenBank accession numbers represent sequences that matched perfectly with the database

A possible bias could have been due to the import of euglyphids with the cadavers, either from the farm or during transport. However, at D8, cadaver samples were less different from the control than samples from fake cadavers (which could not have been contaminated by the new plastic bags and cotton cloth), and we therefore conclude that such contamination was negligible.

The effect of cadavers on euglyphids peaked at D309 (fig. C.3). This time interval seemed quite long in comparison with the results obtained by (Szelecz et al., 2014) from the same field experiment (i.e. complete die-out of testate amoebae 22 days post mortem). However, Szelecz et al. (2014) studied litter and not the underlying mineral soil horizon as done here, and they used a direct observation (microscopy) approach, which most likely underestimated diversity. Indeed, as OTUs eugly_13 and eugly_666 did not belong to any known

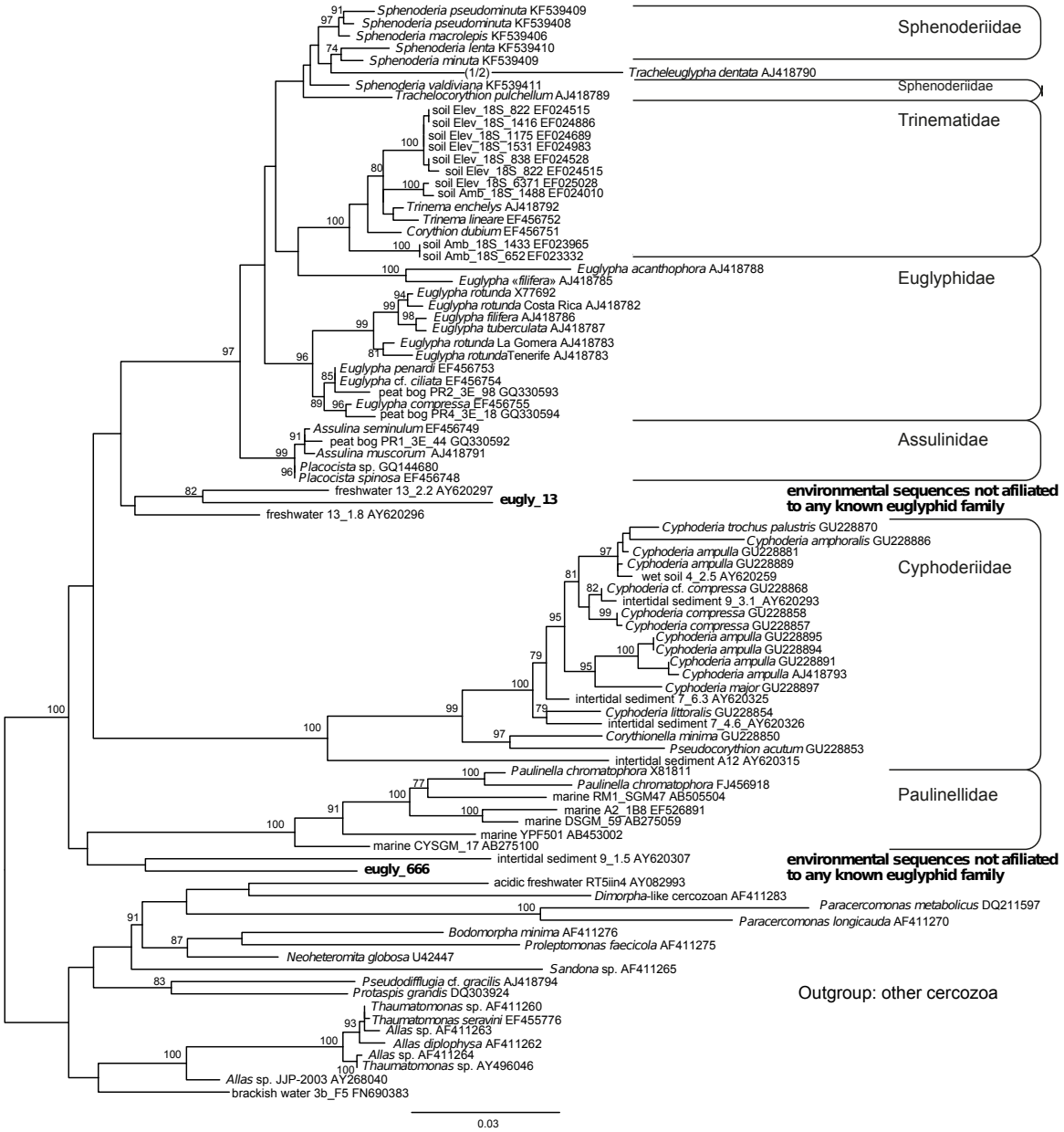

Figure C.4 – Maximum likelihood tree built on full SSU rRNA gene sequences of Euglyphida showing the phylogenetic position of eugly_13 and eugly_666 full-length sequences (in bold) obtained from pig cadaver plots in a *Fagus-Quercus* forest above Neuchâtel, Switzerland. The tree was built using RaxML on 1440 positions and rooted with several cercozoan sequences retrieved from GenBank. Robustness of nodes was evaluated by bootstrapping (1000 replicates)

euglyphid family (fig. C.4), their morphology may differ significantly from known forms, they may be very small and thus overlooked or lost in classical sample preparation protocols using 10-20 µm filters and/or they may represent naked forms (i.e. without test), as documented in foraminiferans (Pawlowski et al., 1999).

The RDA and PRC both showed that euglyphid communities had not completely recovered from the influence of cadavers by the end of the experiment (i.e. D1051). This pattern is in line with the observations of Szelecz et al. (2014), who did not observe a full recovery by D309 (end of their experiment). This long resilience time suggests that euglyphid communities (and probably testate amoebae in general) could be used as indicators of cadaver presence over very long periods. The fact that euglyphid communities still indicated an effect of cadavers either shows a lag in return to pre-disturbance community structure or that they still responded to other differences (e.g. soil chemistry, abundance and composition of prey).

In addition—and this is in itself an unexpected result—the positive response of certain euglyphid OTUs at certain time points (eugly_13 at D33, eugly_666 at D309) (fig. C.5) suggests that individual taxa may respond specifically and positively to some decomposition stages. Such patterns suggest the possibility to use soil testate amoebae as bioindicators for estimating the time elapsed since death (post-mortal interval, PMI), a parameter of considerable importance in forensic sciences.

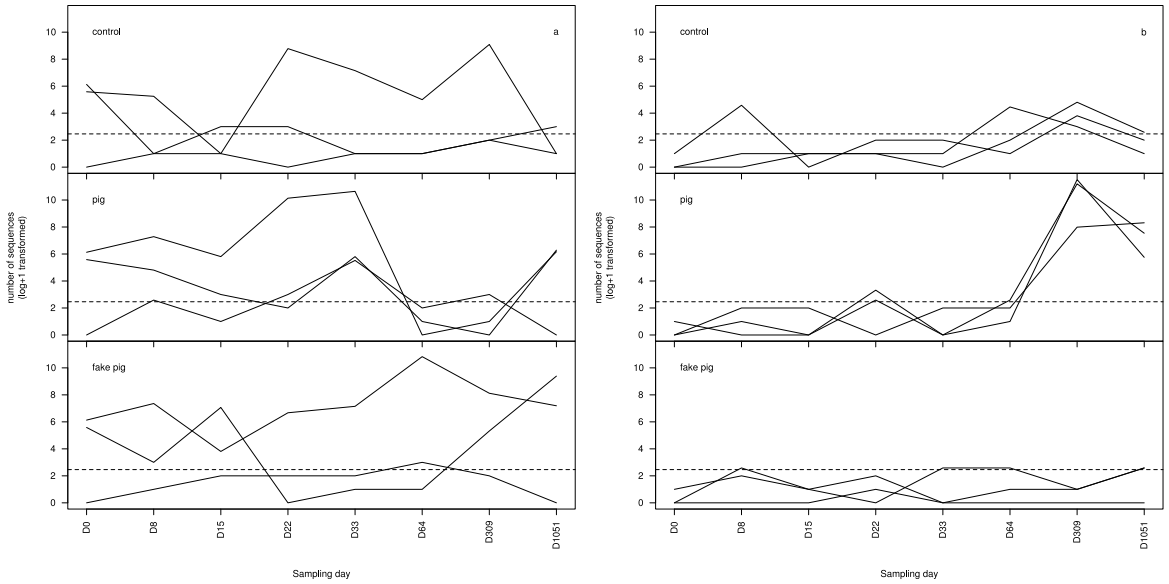

Figure C.5 – Temporal pattern of number of sequences (log +1 transformed) over time in control, fake pig and pig cadaver plots in a *Fagus-Quercus* forest above Neuchâtel, Switzerland, for euglyphid testate amoeba OTUs eugly_13 (a) and eugly_666 (b). Full lines represent the number of sequences in each block. The dashed lines represent the average sequence number in a sample independently of the treatment, block or sampling

Unknown diversity of soil euglyphid testate amoebae

Even after removing rare OTUs, we still found 51 OTUs, 45 of which did not match any sequence in the database. The V9 region does not allow discrimination between close-related species, and it is unclear to how many morphologically and genetically different taxa these 51 OTUs correspond. Regardless of the short length of the barcode, these results reveal the existence of a very high overall diversity of euglyphids in forest soils. This technique yields large amounts of data from small sample volumes, requires much less taxonomic expertise than classical morphological analyses and does not depend on the existence of a reliable taxonomy (which is often lacking for protists).

Perspectives and potential future application

Focusing our study on a specific taxonomic group allowed us to define OTUs at high resolution, using a threshold adapted to already barcoded morphospecies. This approach allowed us to use metabarcoding at a taxonomic resolution close to morphological analysis—much more than what is generally achieved in studies using general eukaryotic marker. Indeed in most studies, more sequences are pooled into OTUs, each of which corresponds to broader taxonomic units than what we achieved in this study. The approach we used to study the response of euglyphid testate amoebae to the impact of decomposing cadavers can also be used to study the responses of any other group of soil eukaryotes. It is indeed very likely that many other taxonomic groups will also show comparable responses to those documented here for euglyphids. Our study shows that some of this unknown diversity could be of potential use for applied purposes such as forensic science. If such patterns can be explored in details, we believe that it will be possible to develop accurate and reliable new molecular bioindicator tools for PMI estimations and other applications.

Acknowledgments, Ethical approval and Conflict of interest

This study was funded by the Stiftung Forensisches Forum Frankfurt/Main, Germany; Vereinigung von Freunden und Förderern der Goethe-Universität, Frankfurt/Main, Germany; the University of Neuchâtel, Switzerland; and the Swiss National Science Foundation (project. No 31003A_141188 to E.M.). We thank Jan Boni (forest engineer of Neuchâtel) for providing the permission to use the site, Dr. Sonia Estelle Tarnawski and Luc Dolivo for the help with DNA extraction, Dr. Roxane Kohler-Milleret for infrared spectroscopy and Emanuela Samaritani for the assistance with the Illumina sequencing. We also thank Balazs Laurenczy for the constructive discussion on the informatics part. All applicable international, national and/or institutional guidelines for the care and use of animals were followed. The authors declare no conflict of interest.

Appendix D

Mycamoeba gemmipara nov. gen., nov. sp., the First Cultured Member of the Environmental Dermamoebidae Clade LKM74 and its Unusual Life Cycle

Quentin Blandenier[1], Christophe C.W. Seppey[1], David Singer[1], Michèle Vlimant[2], Anaële Simon[3], Clément Duckert[1] Enrique Lara[1]

[1]Laboratory of Soil Biodiversity, University of Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland
[2]Laboratory of Ecology and Evolution of Parasites, University of Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland
[3]Laboratory of Microbiology, University of Neuchâtel, Emile Argand 11, Neuchâtel 2000, Switzerland

Journal of Eukaryotic Microbiology, Online published, september 2016
DOI:10.1111/jeu.12357

Abstract

Since the first environmental DNA surveys, entire groups of sequences called "environmental clades" did not have any cultured representative. LKM74 is an amoebozoan clade affiliated to Dermamoebidae, whose presence is pervasively reported in soil and freshwater. We obtained an isolate from soil that we assigned to LKM74 by molecular phylogeny, close related to freshwater clones. We described *Mycamoeba gemmipara* based on observations made with light- and transmission electron microscopy. It is an extremely small amoeba with typical lingulate shape. Unlike other Dermamoebidae, it lacked ornamentation on its cell membrane, and condensed chromatin formed characteristic patterns in the nucleus. *M. gemmipara* displayed a unique life cycle: trophozoites formed walled coccoid stages which grew through successive buddings and developed into branched structures holding cysts. These structures, measuring hundreds of micrometres, are built as the exclusive product of osmotrophic feeding. To demonstrate that *M. gemmipara* is a genuine soil inhabitant, we screened its presence in an environmental soil DNA diversity survey performed on an experimental setup where pig cadavers were left to decompose in soils to follow changes in eukaryotic communities. *Mycamoeba gemmipara* was present in all samples, although related reads were uncommon underneath the cadaver.

Keywords: Budding; Discosea; eukaryotic diversity; fungus; high throughput sequencing; Longamoebia; ribosomal genes; serial dilution; yeast

Introduction

Our vision of protist diversity has been radically challenged since the introduction of observation-independent environmental DNA surveys. Large and deep-branching groups of eukaryotes have been discovered with classical cloning/sequencing strategies, thus overtaking our estimations on eukaryotic environmental diversity. Unsuspected alveolate clades appeared to be extremely diverse in marine systems (López-García et al., 2001), and were found later to be exclusively composed by parasitoids (Guillou et al., 2008). Likewise, many new stramenopile lineages were discovered in the early 2000s in marine systems (Massana et al., 2014). These organisms, which include nowadays 25 different lineages spread all across the tree of stramenopiles, were found to be the most diverse and numerous bacterivores in the sunlit part of oceans (Massana et al., 2004b), and thus vesting them with a prominent role in the oceanic microbial loop (Azam et al., 1983). In deeper waters, a particular group of excavates, the diplonemids revealed an immense diversity (Lara et al., 2009b; Lukes et al., 2015). New massive sequencing technologies also revealed a large diversity in opisthokonts (del Campo et al., 2015).

Other systems like soils, however, have been by far not as deeply studied as the ocean. The high prevalence of fungal, plant, and metazoan sequences has been for a long time a major hindrance for studies on soil protist diversity (Lesaulnier et al., 2008). Still, previously unsuspected deep branching clades have also been discovered in soils, like the Opisthosporidia (also known as Rozel-lomycota) (Karpov et al., 2013; Lara et al., 2009a). Recently, the development of high throughput sequencing has allowed obtaining high numbers of phylotypes, showing promising results in terms of overall microeukaryotic diversity (Geisen et al., 2015b). Relationships among organisms have been inferred using sound experimental designs and approaches such as co-occurrence networks (Lentendu et al., 2014). However, strong conclusions on the organisms morphology and function can only be provided by direct observation or, even better, culturing. Therefore, a current challenge in eukaryotic microbiology is to identify the organisms hiding behind these environmental clades and to infer their ecological function. For this purpose several approaches have been used. Fluorescence in situ hybridization (FISH) in combination with scanning electron microscopy has been recently applied to the characterization of *Paulinellida* (*Euglyphida* testate amoebae) living in forest litter (Tarnawski and Lara, 2015). Bulk soil protists remain, however, widely inaccessible to FISH probes because of the large amount of soil particles unless organ-isms are large enough to be isolated individually (e.g. ciliates, macroscopic mycetozoa, and testate amoebae). Typically, naked amoebozoans are numerous in soils (Geisen et al., 2015b), and harbour a wide array of lifestyles and morphologies (Shadwick et al., 2009).

The amoebozoan environmental clade LKM74, named after the first clone encountered in an environmental DNA survey (van Hannen et al., 1999) is quite abundant and well distributed in soils (Corsaro and Venditti, 2013), but also present in freshwaters (Di Filippo et al., 2015; Richards et al., 2005) and peat bogs (Lara et al., 2011). This clade has been repeatedly placed in the vicinity of *Dermamoeba algensis* in small subunit ribosomal gene trees (18S rRNA) (Corsaro and Venditti, 2013; Kudryavtsev and Pawlowski, 2015), although this relationship remained weakly supported. Despite its pervasive presence in many environments, the organisms have never been kept in culture and their morphology remains unknown. In this study, we describe a tiny naked amoeba isolated from the bulk soil of a coniferous forest by serial dilutions. We affiliated it to LKM74 based on 18S rRNA gene sequences, characterized its complex life cycle and feeding strategy, and documented its ultrastructure. Furthermore, we demonstrated that this species is a typical soil inhabitant by following related sequence reads in an environmental eukaryotic DNA survey of soils. This study was conducted in an

experimental setup where pig cadavers were left to decompose and samples were taken at regular intervals to follow modifications of the microbial eukaryotic communities in the underlying soil during the process of decay.

Material and Methods

Sample collection and identification

The original sample from which the species was described has been taken from a coniferous forest (dominated by *Picea abies*) near Neuchâtel, Switzerland (47°00'85.30"N; 6°55'95.90"E) in August 2011. Soil was suspended into phosphate buffer, and serially diluted into a low nutrient medium in 96 well plates as described in (Lara et al., 2007a). Active amoebae were transferred into culture flasks containing Page's Amoeba Saline medium amended with 1 g per 100 ml of Tryptone Soy Broth and *Escherichia coli* as food organism. Amoebae were subcultured regularly to obtain pure, monoprotistan strains and to lower the proportion of environmental bacteria. Cultures were kept at 12 °C.

Cultures were observed using Utermöhl's plankton chambers with an inverted microscope (Olympus IX81). Cells were measured at different life stages, which were morphologically documented and photographed with light microscopy. Different life stages were also documented using Methyl blue ($C_{37}H_{27}N_3Na_2O_9S_3$) to stain cell walls and condensed cytoplasm.

We also used a full flask containing active and coccoid life stages for transmission electron microscopy (TEM). In that purpose, we pelleted cells from a thriving culture. Fixation, staining, and mounting were achieved as described in (Lara et al., 2006). Observations were made on a Philips CM 100 transmission electron microscope.

Molecular analyses

In addition, a flask containing 10 ml of a thriving culture was used for DNA extraction. Cells were removed from the flask bottom with a cell scraper, and the resulting supernatant was placed into a Falcon tube and centrifuged at maximum speed during 20 min. The obtained pellet was placed into 200 µl of Guanidine thiocyanate buffer and nucleic acids were extracted following a protocol (Chomczynski and Sacchi, 1987) adapted after (Lara et al., 2007c). We amplified the 18S rRNA gene using primers EK 82F (GAAACTGCGAATGGCTC) and EK 1498R (CACCTACGGAAACCTTGTTA) in a total volume of 30 µl with an amplification profile consisting of 4 min at 95 °C followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C, and 1 min 30 s at 72 °C with a final elongation of 10 min at 72 °C. Sequencing was carried out using a BigDyeTM Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems, Geneve, Switzerland) and analysed with a ABI-3130XL DNA sequencer ABI PRISM 3700 DNA Analyzer (PE Biosystems). The sequence was deposited in GenBank with the following accession number: KX687875.

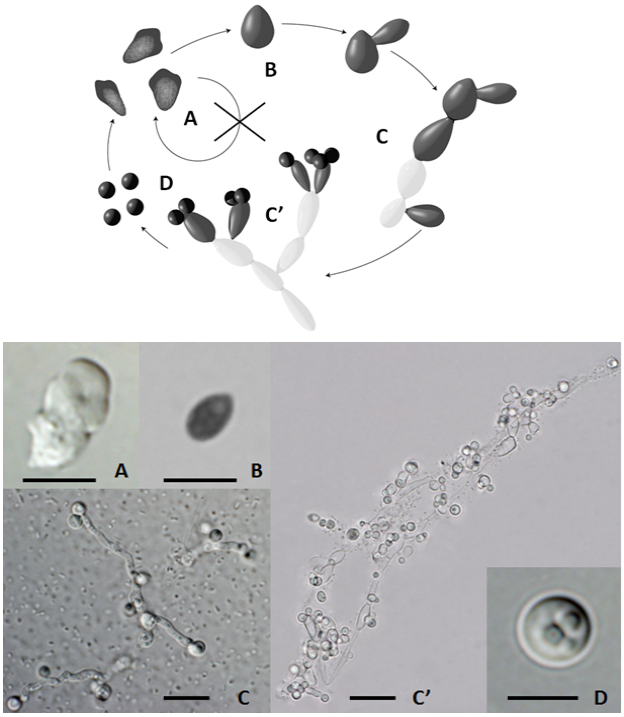

Figure D.1 – Life cycle of *Mycamoeba gemmipara* including its four developmental stages. (A) Trophozoites. (B) Coccoid stage, stained with methyl blue. (C–C') bud like ramification. (D) Cysts released from branches of the pseudomycelium ramification. Scale bar: A–B, D = 5 µm, C–C' = 10 µm.

Phylogenetic analysis

The obtained sequence was placed in an alignment containing various sequences from Discosea (with an emphasis on Longamoebia as defined in (Smirnov et al., 2011a)) derived either from isolated cells or cultures, or from environmental clone sequences. The root was placed on Vannellida. The alignment is available from the authors upon request. We build a maximum likelihood phylogenetic tree using the RAxML algorithm (Stamatakis et al., 2008) as implemented on the web server "http://embnet.vital-it.ch/raxml-bb", and evaluated the robustness of the nodes by bootstrapping.

Screening of high throughput sequencing data

The experimental setup from which the soil samples were retrieved has already been published in (Szelecz et al., 2014). Briefly, three pig (*Sus scrofa*) carcasses were left to decompose during 1,051 days in a forest and soil was sampled at different time steps (21 samples in total). Three control soils (without pig) situated at a few meters of distance were sampled at the same time (24 samples). This experiment was performed in the context of a forensic study aiming at improving the estimation of the post mortem interval based on the observation of soil eukaryotic communities. DNA extraction, PCR of environmental DNA and sequencing of the V9 region of the 18S rRNA gene were performed as in (Seppey et al., 2016). The obtained reads were pretreated as in (De Vargas et al., 2015). The environmental sequences were aligned against a database constructed from publicly available V9 sequences of clones related to *Mycamoeba gemmipara* as determined with BLAST (GenBank: AY919786.1, AY919722.1, GQ861575.1, GQ861565.1, GQ861560.1), plus the sequence derived from the *M. gemmipara* culture. We determined empirically a threshold based on e-values to assess if a given environmental read belongs to the *M. gemmipara* group. The selected reads were then clustered into phylotypes using the SWARM v: 1.2.5 clustering algorithm (Mahé et al., 2014). To determine if the abundance of reads related to *Mycamoeba* changed significantly between control and cadaver plots, we performed Wilcoxon test on the data. All statistical analysis was done with R (R Development Core Team 2011).

Results

Active trophozoites were extremely small (maximum 7 µm in length when moving towards a single direction, and 2 µm width), had a flattened shape and lobose pseudopodia. When moving in a single direction, trophozoites had the typical lingulate shape observed in other Dermamoebidae (fig. D.2A), and conical-shaped pseudopodia can be observed when the amoeba changes direction (fig. D.2B). The stationary shape (fig. D.2C) showed small pseudopods radiating in several directions. The hyaline zone seemed to be generally only restricted on the extensive pseudopodia and was not visible laterally. No uroid or surface ridges were visible.

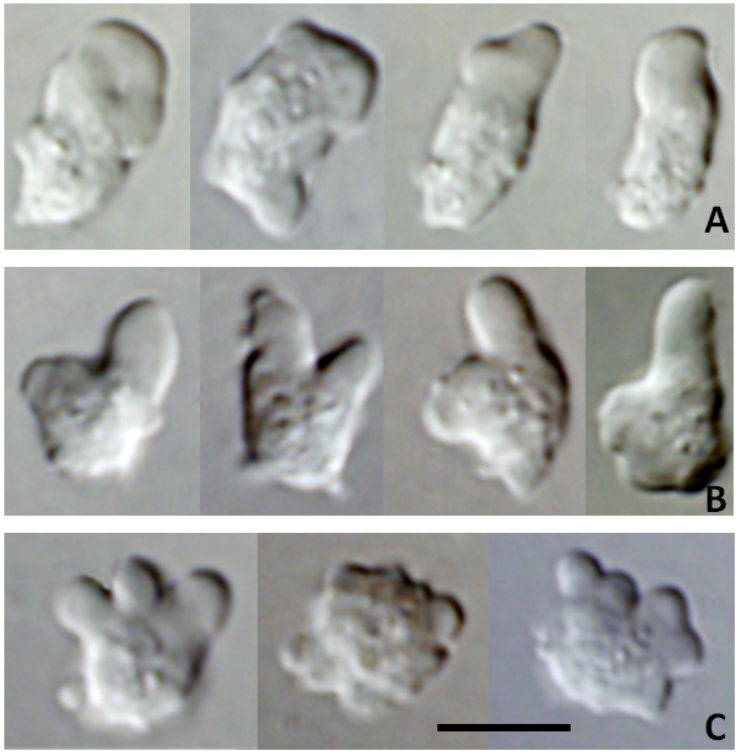

Figure D.2 – Different morphologies of active trophozoites. (A) Trophozoites showing a typical lingulate shape. (B) Trophozoites when changing direction. (C) Stationary form. Scale bar: A–C = 5 µm.

The life cycle of *M. gemmipara* comprised four stages (fig. D.1). (i) Active trophozoites were observed moving and feeding on *E. coli*. At this stage, no cell division could be observed. (ii) Cells then became rounded-elongated and smooth (coccoid phase), and stopped moving and feeding (fig. D.1B). Cytoplasm became highly condensed, and a cell wall was built. (iii) After about 24 h, the coccoid cells started budding several times successively, in a modular growth manner (fig. D.1C, fig. D.4A). The resulting structures grew, reaching up to 200 µm or more in length (fig. D.1C'). These structures in which cells cling together in chains correspond to a pseudomycelium as it has been described in several yeasts and bacteria. Old modules were devoid of cytoplasm, which probably migrated towards newly developed structures (fig. D.4B); still, cell walls without cytoplasm remained up to several months. Spherical dispersal cysts (fig. D.1B, fig. D.4B), were formed at the tip of the branched structures, the latter being finally degraded (iv). Dispersal cysts could then germinate into active amoebae immediately, but will do so most often when fresh medium is provided, thus closing the cycle. A movie (Movie S1) where the first steps of budding in coccoid cells can be observed is available on the journal website.

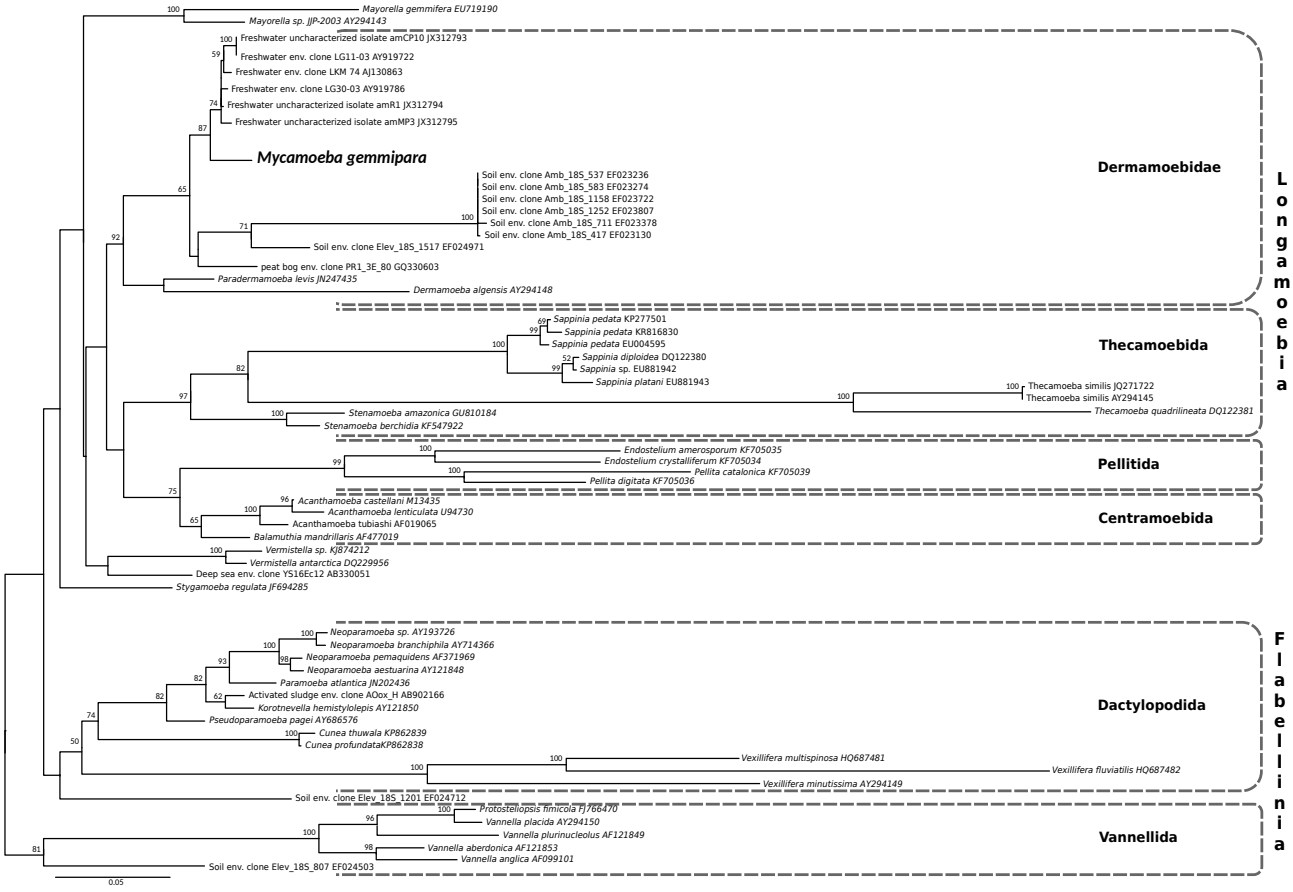

Figure D.3 – Maximum likelihood phylogenetic tree based on complete 18S rRNA sequences of several Longamoebida, showing the position of *Mycamoeba gemmipara* within Dermamoebidae. The tree was built using the RAxML algorithm (Stamatakis et al. 2008).

Light/electron microscopy observations

Transmission electron microscopy pictures did not show any ornamentation on the cellular membrane (fig. D.4). Chromatin in the nucleus appeared inhomogeneous and presented several conspicuous zones with higher density within the nucleus. These zones were situated against the nuclear membrane, as well as in the centre of the nucleus, and a large nucleolus could be observed. Mitochondrial cristae were tubular. Several phagocytosis vac-uoles could be seen containing bacteria at different stages of digestion. Small vesicles, probably containing digestive enzymes could be seen surrounding the bacterial cells.

Molecular phylogeny

Mycamoeba gemmipara branched within a group that comprised the original LKM74 environmental clone plus several other clones and unidentified isolates (fig. D.3). This clade, named genus *Mycamoeba* hereafter, received a moderate support (bootstrap value (BV) = 65). Furthermore, it branched robustly (BV = 87) at the base of a clade comprising only freshwater forms, the B1 clade sensu (Corsaro and Venditti, 2013). The monophyly of family Dermamoebidae was recovered with high statistical support (BV = 92), including genera *Dermamoeba, Paradermamoeba* and, now, *Mycamoeba*.

Presence of *Mycamoeba* reads in environmental DNA surveys

Our screen through the Illumina reads obtained from the soil from the forensic experiment allowed assigning 7,482 sequences to genus *Mycamoeba* (using an e-value alignment threshold below $1e^{-45}$) out of 25,579,257 environmental sequences, thus representing a total of 0.03% of all reads. These reads were present in all the 45 samples. The clustering of the environmental sequences resulted in a single Operational taxonomic unit (OTU) corresponding at 100% of identity with the original sequences of *M. gemmipara*. fig. D.5 shows the

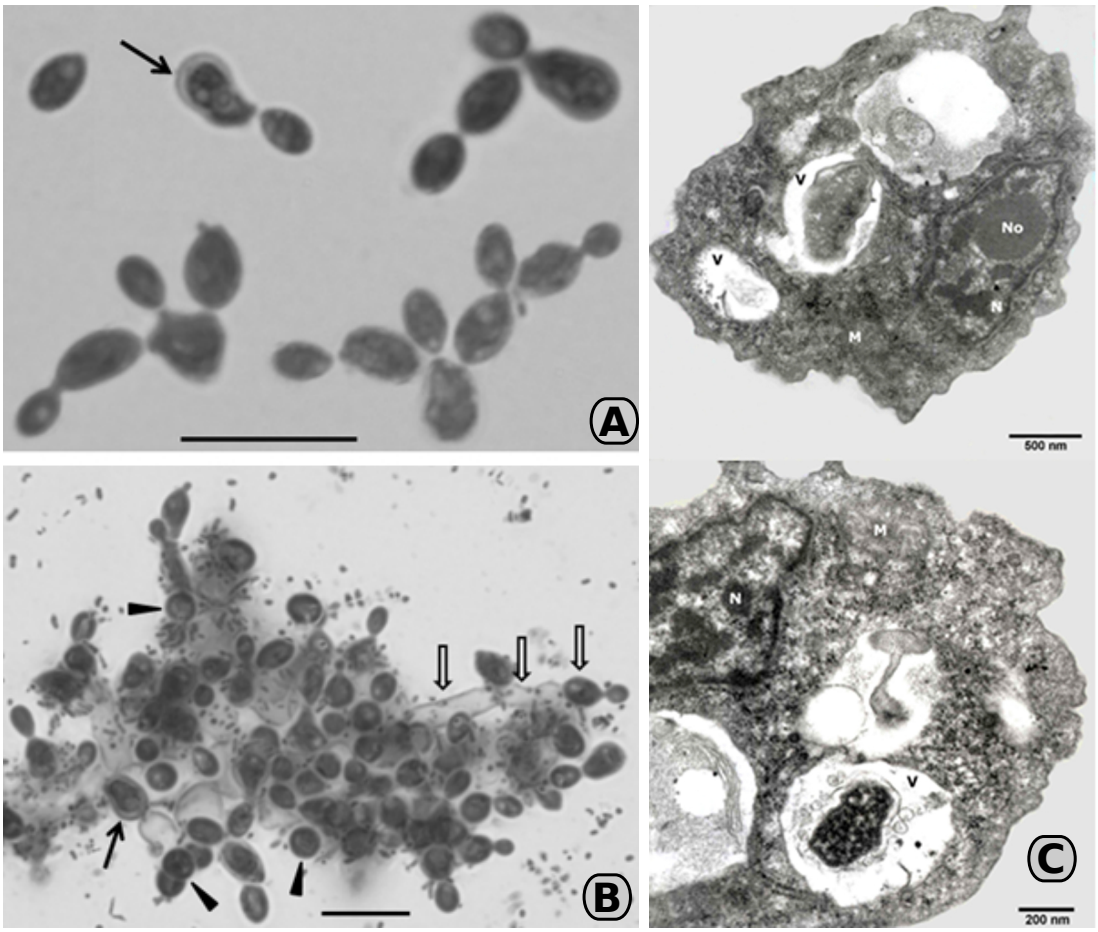

Figure D.4 – Development of the coccoid cells. (A) Budding of coccoid (ovoid shaped) cells. Black arrow indicated cell wall. (B) Mature fructifications. Black triangles indicate mature cysts (spherical shape); a young coccoid cell filled with cytoplasm is indicated with a black arrow; old (empty) coccoid cells are shown with white arrows. Scale bar: A–B = 10 µm. All structures have been stained with methyl blue coloration. (C) TEM sections of *Mycamoeba gemmipara* showing: "M" the tubulicristate mitochondria; "N" the nucleus with the particular arrangement of the chromatin and "No" the nucleolus. Phagocytosed bacterial cells are visible in the vacuoles "V", where lysosomes can be noticed in the process of releasing digestive enzymes.

sequences abundances distributions according to the control group and pig treatment. Number of reads is significantly lower in the pig treatment samples than in the control group (Wilcoxon test: p-value < 0.05). *Mycamoeba gemmipara* is the only described representative of environmental clade LKM74. Its size, clearly below 10 µm, places it among the smallest known amoebozoans. *Parvamoeba rugata* (Rogerson, 1993) *Paravanella minima* (Kudryavtsev et al., 2011) and *Sapocribrum chiconteaguense* (Lahr et al., 2015) may be shorter in average but the thinner and flattened (lingulated) shape of *M. gemmipara* gives it a smaller biovolume. Coccoid forms measure a little above 3 µm. The size of *M. gemmipara* and its inconspicuous aspect, plus the fact that the amoeboid stage lasts a relatively short amount of time is probably the reason why it had never been detected previously. This small size may be characteristic of the whole LKM74 clade, as the isolates observed by (Corsaro and Venditti, 2013) were also smaller than 10 µm. Recently, there has been an increasing number of descriptions of nano-sized Amoebozoa, summarized in (Kudryavtsev and Pawlowski, 2015). As

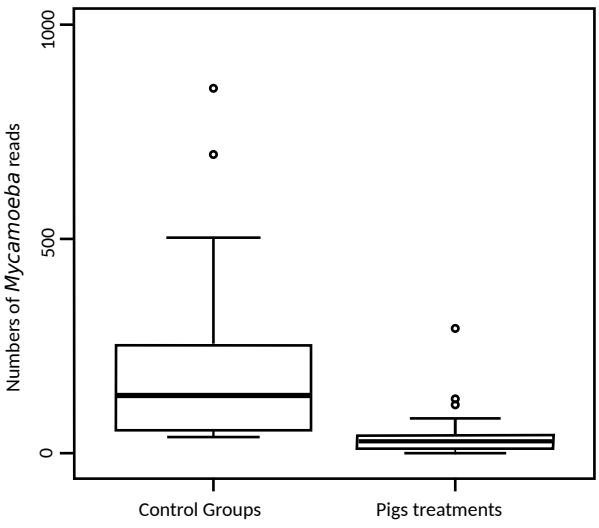

Figure D.5 – Boxplot showing the number of V9 18S rRNA reads related to *Mycamoeba gemmipara* in control treatments and under the decomposing pig, suggesting a negative effect of cadaveric fluids on the amoebae.

their 18S rRNA genes are often fast-evolving like in many other Amoebozoa, they remain undetected in DNA-based environmental diversity surveys, either because "general eukaryotic" primers fail in amplifying their 18S rRNA, or because their sequences cannot be assigned with confidence to a given group of eukaryotes. In that sense, application of systematic cultural approaches to soils and sediments are promising in revealing new lineages of Amoebozoa.

Discussion

Our DNA survey demonstrated that *M. gemmipara* was a genuine forest soil inhabitant, as it has been found in all samples taken in the control group and under the pig carcasses. The site where *M. gemmipara* was isolated was located a few kilometres away from the experimental setup (Szelecz et al., 2014), and we can therefore suppose that this amoeba is widespread and probably abundant in the forests surrounding Neuchâtel (Switzerland). Nevertheless, it branched at the base of a clade comprising exclusively species from aquatic environments (Corsaro and Venditti, 2013). Other fruiting amoebae, such as *mycetozoa* and "protosteliales" were also recovered from aquatic environments, where trophozoites are probably active (Lindley et al., 2007). Therefore, it cannot be excluded that *M. gemmipara* lives in both freshwater sediments and soils. Its ability to build bud like ramifications in a liquid medium differs from most fruiting amoebae, although instances of myxogastrids completing their whole life underwater have been reported (Gottsberger and Nannenga.N.E, 1971; Kappel and Anken, 1992). Moreover, this organism could be used as an indicator of certain environmental perturbation as it responded significantly to the effect of the cadaver decomposition (release of high concentration of nutrient, anoxic conditions; fig. D.5).

Order Dermamoebida has been first suggested by (Cavalier-Smith et al., 2004). It hosted a single family, Thecamoebidae, with genera *Thecamoeba* and *Dermamoeba*. Further analyses based on 18S rRNA gene sequences suggested a separated grouping for Thecamoeba (Kudryavtsev et al., 2005), which formed the order Thecamoebida including also genera *Sappinia* (Brown et al., 2007), *Stenamoeba* (Smirnov et al., 2007), and *Vermistella* (Tekle et al., 2016). A new, emended order Dermamoebida comprised genera *Dermamoeba* and *Paradermamoeba* (Dermamoebidae), and *Mayorella* (Mayorellidae). To our knowledge, the monophyly of order Dermamoebida has never been recovered using 18S rRNA genes, as the grouping of genus *Mayorella* with Dermamoebidae remained either weakly supported (Pawlowski and Burki, 2009) or not at all (Corsaro and Venditti, 2013). Its inclusion within order Dermamoebida has been suggested on the base of the presence of a cell coat without wrinkles (Smirnov et al., 2011a); however, a recent multigene tree analysis suggested that *Mayorella* was rather related to the mainly marine amoebozoan clade Dactylopodida (Tekle et al., 2016).

The general lingulate shape of the locomotive form and the conical pseudopodia of *M. gemmipara* are typically found in all other members of Dermamoebidae (Kudryavtsev et al., 2011). Chromatin patterns in the nucleus of *M. gemmipara* are very distinctive, makes condensed granules all through the nucleus and especially near the borders of the nuclear membrane. All other known Dermamoebidae possess homogeneous chromatin in their nucleus with the exception of a large nucleolus.

This includes *D. algensis* (Smirnov et al., 2011b) and *Paradermamoeba* (Smirnov and Goodkov, 2004). Genera placed within Dermamoebidae are characterized by a conspicuous cell coat. *Dermamoeba* and *Paradermamoeba* have a thick, highly structured cell coat (Smirnov et al., 2011a), either cuticle-like or consisting analogous glycostyle-like structures. In contrast, *M. gemmipara* does not possess any structure around its cell membrane, at least during its trophozoite life stage.

The peculiar life cycle of *M. gemmipara* has no known equivalents in Amoebozoa. While trophozoites actively ingest bacteria (as shown in TEM images; fig. D.4), the cell walls that are formed around coccoid cells preclude any phagocytosis. Nevertheless, the organisms undergo a considerable biovolume increase during pseudomycelium formation. Biomass incorporation can therefore only occur by osmotrophy. Its principle may remind the polyphyletic "protosteliales" (Shadwick et al., 2009), where a prespore stage (without cell wall) precede the formation of a stalk, which is used for spore dispersal (Olive, 1967), However, the simple shape of protostelids stalks differs considerably from the branched formations observed in *M. gemmipara* pseudomycelium (fig. D.1, fig. D.4). Moreover, in protostelids, stalks are formed from cysts when environmental conditions are degrading, whereas active amoebae (trophozoites) undergo cell division (Dykstra and Keller, 2000). In *M. gemmipara*, we did not observe any cell division at the trophozoite stage, which suggests that multiplication occurs only during coccoid budding process. Globally, the structures produced by *M. gemmipara* resemble those observed in Fungi. The reproductive mode by budding reminds strongly of yeasts such as *Saccharomyces cerevisiae*, and many fungi revert from yeasts to mycelial growth in a single organism (Rippon, 1982). As *M. gemmipara* original culture was regularly subcultured for about five years in the laboratory (thus generating dozens of replicates), we rule out the possibility that budding coccoid cells and pseudomycelia could be originated by fungal contaminants, which should have logically either disappeared or invaded our cultures. The main difference between the pseudomycelium observed in *M. gemmipara* and a classical mycelium is that its structures

are not perennial and degrade once the cysts are released. *M. gemmipara* combines thus two life strategies: phagotrophy in a first stage of its life cycle (probably accumulating enough biomass to enter its next life stage) and osmotrophy afterwards. This strategy appears successful in freshwater and soil environments, and suggests a similar evolutionary pathway as in the Nucletmycea. These latter also evolved from amoeboid organisms, such as *Nuclearia*, towards mycelial growing organisms such as the true Fungi (Brown et al., 2009), which may have occurred also in soils or freshwater.

Classification summary

Taxonomic summary: *Mycamoeba gemmipara* nov. gen. nov. spec

Amoebozoa Eudiscosea Longamoebia Dermamoebidae *Mycamoeba gemmipara*

Genus *Mycamoeba* gen. nov

Description: flattened amoebae with lobose or conical pseudopodia and a lingulate shape. Mitochondrial cristae are tubular. No glycocalyx or other ornaments on the cell surface. Possess a complex life cycle where active cells transform into coccoid stages, which undergo subsequent buddings, eventually turning into ramified structures (pseudomycelia) with spherical cysts in a terminal position on the ramifications. These structures disappear and cysts are released prior to germinating into active trophozoites. Etymology: From ancient Greek: myces; mould, fungus, in reference to the fungal-like structures appearing in the osmotrophic stage of its life-cycle, as well as its peculiar mode of reproduction, atypical for an amoeba, but reminding strongly yeasts. Sole species: *Mycamoeba gemmipara*.

Mycamoeba gemmipara sp. nov

Description: Small amoeboid cells up to 7 µm. Chromatin is not distributed in a homogeneous manner: there are conspicuous condensed regions pressed against the nuclear membrane and also in the centre of the nucleus. Ecology: has been detected to date only in forest soils. Seems to avoid important amounts of nitrogen (release of cadaveric fluids) or other perturbations generated by cadaver decomposition. Hapantotype: a culture has been deposited at the culture collection "Culture Collection of Algae and Protozoa". 18S rRNA gene sequence of *Mycamoeba gemmipara* (1,738 bp) was deposited in GenBank (KX687875). Etymology: *gemmipara*, as a reference to the reproduction mode (pario, giving birth in latin) through budding (gemma, a bud in latin).

Acknowledgements

We thank the Swiss NSF for funding (project no. 310003A_ 143960). We are grateful to Daniel J.G. Lahr (University of Sao Paulo, Brazil) for his insightful comments, and also to an anonymous reviewer whose constructive comments helped us in improving the manuscript. We thank also Idliko Szelecz for the soil samples. This work has been realized by Quentin Blandenier during his undergraduate studies.

Appendix E

Soil Protists Communities in Three Neotropical Rainforests are Hyperdiverse and Dominated by Parasites

Frédéric Mahé[1], Colomban de Vargas[2,3], David Bass[4,5], Lucas Czech[6], Alexandros Stamatakis[6,7], Enrique Lara[8], David Singer[8], Jordan Mayor[9], John Bunge[10], Sarah Sernaker[11], Tobias Siemensmeyer[1], Isabelle Trautmann[1], Sarah Romac[2,3], Cédric Berney[2,3], Alexey Kozlov[6], Edward A. D. Mitchell[8,12], Christophe V. W. Seppey[8], Elianne Egge[13], Guillaume Lentendu[1], Rainer Wirth[14], Gabriel Trueba[15], and Micah Dunthorn[1]

[1]Department of Ecology, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany,
[2]CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France,
[3]Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Station Biologique, Place Georges Teissier, 29680 Roscoff, France,
[4]Department of Life Sciences, The Natural History Museum London, Cromwell Road, London SW7 5BD, United Kingdom,
[5]Centre for Environment, Fisheries and Aquaculture Science, Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK,
[6]Heidelberg Institute for Theoretical Studies, Schloß-Wolfsbrunnenweg, 69118 Heidelberg, Germany,
[7]Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Am Fasanengarten, 76128 Karlsruhe, Germany,
[8]Laboratory of Soil Biodiversity, Université de Neuchâtel, Rue Emile-Argand, 2000 Neuchâtel, Switzerland,
[9]Department of Forest Ecology and Management, University of Agricultural Sciences, Skogsmarksgränd, 90183 Umeå, Sweden,
[10]Department of Computer Science, Cornell University, Gates Hall, Ithaca, New York 14853, USA,
[11]Department of Statistical Science, Cornell University, Malott Hall, Ithaca, New York 14853, USA,
[12]Jardin Botanique de Neuchâtel, Pertuis-du-Sault, 2000 Neuchâtel, Switzerland,
[13]Department of Biosciences, University of Oslo, Blindernveien, 0316 Oslo, Norway,
[14]Department of Plant Ecology and Systematics, University of Kaiserslautern, 67663 Kaiserslautern, Germany,
[15]Instituto de Microbiología, Universidad San Francisco de Quito, Diego de Robles, Quito, Ecuador.

Nature Ecology & Evolution, accepted

Abstract

High animal and plant richness in tropical rainforest communities has long intrigued naturalists. It is unknown if similar hyperdiversity patterns are reflected at the microbial scale with unicellular eukaryotes or protists. Here we show, using environmental metabarcoding of soil samples and a novel phylogeny-aware cleaning step, that protists communities in Neotropical rainforests are hyperdiverse and dominated by the parasitic Apicomplexa, which infect arthropods and other animals. These host-specific protist parasites potentially contribute to the high animal diversity in the forests by reducing population growth in a density-dependent manner. By contrast, too few Oomycota OTUs were found to broadly drive high tropical tree diversity in a host-specific manner under the Janzen-Connell model. Extremely high OTU diversity and high heterogeneity between samples within the same forests suggest that protists, not arthropods, are the most diverse eukaryotes in tropical rainforests. Our data show that protists play a large role in tropical terrestrial ecosystems long viewed as being dominated by macro-organisms.

Introduction

Since the works of early naturalists such as von Humboldt and Bonpland (Von Humboldt and Bonpland, 1853), we have known that animal and plant communities in tropical rainforests are exceedingly species rich. For example, one hectare can contain more than 400 tree species (Valencia et al., 1994) and one tree can harbour more than 40 ant species (Wilson, 1987). This hyperdiversity of trees has been partially explained by the Janzen-Connell model (Connell, 1970; Janzen, 1970), which hypothesizes that host-specific predators and parasites reduce plant population growth in a density-dependent manner (Bagchi et al., 2014; Terborgh, 2012). Sampling up in the tree canopies and below on the ground has further led to the view that arthropods are the most diverse eukaryotes in tropical rainforests (Basset et al., 2012; Erwin, 1982).

The focus on eukaryotic macro-organisms in these studies is primarily because they are familiar and readily observable to us. We do not know if the less familiar and less readily observable protists—microbial eukaryotes excluding animals, plants, and fungi (Pawlowski et al., 2012) —inhabiting these same ecosystems exhibit similar diversity patterns. To evaluate if macro-organismic diversity patterns are reflected at the microbial scale with protists, we conducted an environmental DNA metabarcoding study by sampling soils in 279 locations in a variety of lowland Neotropical forest types in: La Selva Biological Station, Costa Rica; Barro Colorado Island, Panama; and Tiputini Biodiversity Station, Ecuador. This metabarcoding approach has the power to uncover known and new taxa on a massive scale (De Vargas et al., 2015). By amplifying DNA extracted from the soils with broadly-targeted primers for the V4 region of 18S rRNA and sequencing it with Illumina MiSeq, we were able to detect most eukaryotic lineages, and assess the diversity and the relative dominance of free-living as well as parasitic lineages.

Methods

Code availability

All codes used here can be found in HTML format (Supplementary Files S1 and S2).

Neotropical soil samples

Two-hundred and seventy-nine soil samples were taken from: La Selva Biological Station, Costa Rica in October 2012 and June 2013; Barro Colorado Island (BCI), Panama October 2012 and June 2013; and Tiputini Biodiversity Station, Ecuador in October 2013 (see Supplementary File 1 for sample coordinates). The aim was to collect soils from around each field station (excluding areas where humans were not permitted and areas of secondary growth) and no specific sub-environments were targeted during sampling (e.g., samples were collected in a variety of conditions such as swamp mud, sandy soils, hill-top dry soils). Fieldwork and government-issued sampling permits were made possible by the Organization for Tropical Studies and the Ministerio del Ambiente y Energía and Sistema Nacional de Areas de Conservación in Costa Rica, the Smithsonian Tropical Research Institute in Panama, and David Romo Vallejo and Consuelo de Romo at the Universidad San Francisco de Quito in Ecuador.

At the site of collection, the loose-leaf layer and stones were removed. Five ml of the top- most layer of surface soil (with visible plant roots and visible animals removed) were immediately placed into eight ml of LifeGuard Soil Preservation Solution (MO BIO, Carlsbad, CA). DNA was isolated with the PowerSoil DNA Elution Accessory Kit (MO BIO). Like other large-scale sequencing studies, e.g., (De Vargas et al., 2015; Tedersoo et al., 2014), only DNA was used in this study, not RNA which can likewise remain even in ancient

sediments (Orsi et al., 2013; Pawlowski et al., 2014). See Supplementary Figure 13 to see that RNA can be found in cysts, and for a comparison of DNA and RNA abundances; additionally, the low similarity of OTUs among samples within each forest, even those that were collected near each other (Supplementary Figure 12), point to the soils not having the same taxa that would be expected if most of the DNA was from a large seed bank of dead or inactive species that collected overtime. The samples were taken mostly at least 100 meters apart. DNA from every two consecutive samples was combined in equal concentration to reduce costs; many of these samples failed to amplify and only those that worked are shown here. The combined samples were amplified with a high fidelity polymerase for the V4 region of 18S rRNA following (Mahé et al., 2015) with universal eukaryotic V4 primers (TAReuk454FWD1 and TAReukRev3) (De Vargas et al., 2015). This primer pair has not shown a preference for amplifying Apicomplexa, and the V4 primers were able to amplify all Oomycota in GenBank in an in silico analysis.

After PCR cleanup, a second PCR with primers containing sample-specific tags and Illumina MiSeq's sequencing adapters was performed following manufacturer's instructions. Amplicons were gel-size selected for appropriate length, and library quality was assessed using the Bioanalyzer 2100 (Agilent, Santa Clara, CA) and Qubit 2.0 (Life Technologies, Darmstadt, Germany). Illumina MiSeq sequencing was performed at GATC Biotech AG (Konstanz, Germany) with v3 chemistry and software MCS v2.3.0.3 and RTA v1.18.42.0. 20% PhiX DNA was spiked-in to the MiSeq flowcell to increase base diversity. Twenty amplified products (each from a combined two samples) were loaded onto one MiSeq flowcell.

Ten samples from La Selva, Costa Rica, and ten samples from Barro Colorado Island, Panama, were also amplified with primers designed to specifically target the closely related Ciliophora (Lara et al., 2007a). After PCR cleanup, a second, nested PCR was performed with the above V4 primers, and then sequenced with Roche/454 GS FLX+ with Titanium chemistry and software v2.6. The 20 amplified products were loaded onto 1/4 of a plate.

Marine samples

Reads from the open oceans came from the Tara-Oceans consortium (De Vargas et al., 2015). Starting with Tara-Oceans' raw reads (SRA accessions PRJEB6610 and PRJEB7315) allowed us to treat the sequencing data with the same bioinformatic pipeline. Our pipeline's goal was to improve the sensitivity of OTU detection by applying most quality- and frequency-based filtering steps after clustering instead of applying them before, and removal of low abundant OTUs was performed on the combined data and not at the sample level. We also replaced the paired-read merger FLASH (Magoč and Salzberg, 2011) with the newer PEAR (Zhang et al., 2014), which is advertised to yield higher quality results. After a careful investigation, it turned out that the large increase in OTU number we observed compared to (De Vargas et al., 2015) is mainly due to that later modification.

Samples from European near-shore marine sites (see Supplementary File 1 for coordinates for each sample) came from the BioMarKs consortium (Logares et al., 2014; Massana et al., 2015). We re-sequenced the samples here to obtain deeper sequencing. Total nucleic acids from sediment samples were extracted using the RNA Power Soil Total Isolation kit combined with DNA Elution Accessory kit (MO BIO) following manufacture's instructions. To remove contaminating DNA in the RNA extracts, the samples were treated twice with DNase for 25min at 37°C with 2U of Turbo DNA- free kit (Ambion) followings manufacturer's instructions. RNA was reverse-transcribed into first-strand cDNA using Invitrogen's Superscript III (Thermo Fisher Scientific, Waltham, MA) following manufacturer's instructions. The V4 region was amplified with the above primers, and the V9 region was amplified with general eukaryotic primers (1389F and 1510R) (Amaral-Zettler et al., 2009). Amplified products were sequenced on an Illumina Genome Analyser IIx sequencer at CEA Genoscope (Évry, France).

Read cleaning and clustering

Fastq files were assembled with PEAR v0.9.8 55 using default parameters and converted to fasta format. Following (Mahé et al., 2015), assembled paired-end reads were filtered with Cutadapt v1.9 (Martin, 2011) and retained if they had both primers and no Ns (or ambiguously called nucleotides). Reads were dereplicated into strictly-identical amplicons with VSEARCH v1.6.0 (Rognes et al., 2016), then clustered with Swarm v2.1.5 (Mahé et al., 2014) using d=1 with the fastidious option on. The most abundant amplicon in each OTU was searched for chimeric sequences with VSEARCH, and their OTUs were removed even if they occurred in multiple samples.

Taxonomic assignment

Taxonomic assignment used VSEARCH's global pairwise alignments with the Protist Ribosomal Reference (PR^2) database v203 (Guillou et al., 2013) based on GenBank release v203. Using Cutadapt, the PR^2 database was extracted for just the specific regions that were amplified and sequenced to allow for comparisons using a global pairwise alignment. Amplicons were assigned to their best hit, or co-best hits, in the reference database as reported by VSEARCH. The different steps of that taxonomic assignment strategy are grouped in a pipeline called Stampa (`https://github.com/frederic-mahe/stampa`). Low abundant OTUs were removed from the combined dataset only if they included $\leq$2 reads, and were found in only one sample, and were $<$99% similar to an accession in PR2. All reads were deposited at GenBank's SRA bioproject SUB582348.

Stampa plots

To assess globally the results of Stampa, our taxonomic assignment pipeline, we produced "Stampa plots" (`https://github.com/frederic-mahe/stampa`). Stampa plots are simple distribution plots showing the number of reads per similarity value, where the similarity value is the best match between environmental and reference sequences. Stampa plots are a visual assessment of the taxonomic coverage of our reference database sequences: if most environmental reads or OTUs have high similarity values with references, then the coverage is good. On the contrary, low similarity values indicate a lack of coverage, which can be the sign of amplification-sequencing artifacts (chimeras for instance), the presence of known taxa without reference sequences, the presence of new taxa, or if the coverage deficit is massive, a previously unexplored environment (such as Neotropical rainforest soils).

Phylogenetic placement pipeline

To put the 50,118,536 soil protist reads into a phylogenetic context, they were dereplicated into 10,567,804 strictly identical amplicons and placed onto a comprehensive eukaryotic reference tree. The corresponding multiple sequence alignment (MSA) used to build this tree contained 512 full length sequences from all major eukaryotic clades (see Supplementary File 2 for GenBank accession numbers) based on the taxon sampling as summarized in a range of pan-eukaryotic phylogenomic studies reviewed in (Burki, 2014; del Campo et al., 2014), with a bias towards lineages known to occur in soils from environmental sequencing studies; e.g., (Bates et al., 2013; Dupont et al., 2016; Geisen et al., 2015b). Only high quality, full- or near-full length 18S rRNA reads were selected, and reads that have previously been observed to form long branches omitted, to reduce phylogenetic artifacts. It is acknowledged that single gene trees are unable to resolve the backbone of the eukaryote phylogeny; however, this was not the intended purpose of this taxon selection or the phylogenetic analyses. The intent was to produce a comprehensive eukaryote sample tree on which to place the V4 reads generated by this study. Phylogenetic placements of query OTU representatives that were taxonomically assigned to the Apicomplexa were conducted using a full-length reference MSA and corresponding phylogeny comprising 190 taxa from all major Alveolate clades (see Supplementary File 2 for GenBank accession numbers). The taxon sampling was collated from a range of publications relating to eukaryotic diversity in general and Alveolata diversity and phylogeny in particular; e.g. (Adl et al., 2012; Barta et al., 2012; Bass and Cavalier-Smith, 2004; Gómez et al., 2009; Rueckert and Leander, 2009; Rueckert et al., 2010; Slapeta and Linares, 2013; Wakeman and Leander, 2012). It is also acknowledged that single gene trees are unable to resolve the backbone of the Alveolate phylogeny.

For a full explanation of the phylogenetic placement see Supplementary File 2. In brief, we used RAxML v8.1.15 (Stamatakis, 2014) to infer reference trees from the reference alignments, then used PaPaRa v2.4 (Berger and Stamatakis, 2011) to align the query sequences to those reference alignments, and finally used the Evolutionary Placement Algorithm (Berger et al., 2011) as implemented in RAxML to place the queries onto the trees. Placement results were visualized as "Heattrees" using Genesis v0.2.0 (`https://github.com/lczech/genesis`). The trees were inferred with and without taxonomic constraints, although the results were similar. We also assessed the quality of the phylogenetic placement positions to determine how confident the algorithm is when placing a query sequence on the branches of the reference tree by analyzing the distribution of the likelihood weights for the placements, and by analyzing the locality of placement distributions for each amplicon over the tree (Supplementary Fig. 14).

Phylogenetic placement of Haptophyta

For the 28 Haptophyta OTUs, we fist downloaded 106 GenBank accession following the taxon sampling of (Edvardsen et al., 2016; Egge et al., 2015; Shalchian-Tabrizi et al., 2011), and aligned in MAFFT v7 (Katoh et al., 2002). OTU representatives were then added to this alignment using the MAFFT "add-in" function, and

the tree was inferred with RAxML, with substitution model GTR-G-I, algorithm New rapid hill- climbing and 100 bootstrap runs.

Oomycete analyses

To tentatively assign functional roles of the Oomycota OTUs, we first compared each individual OTU with related sequences in Genbank using BLAST (Altschul et al., 1997) for a taxonomic affiliation. In cases of identical BLAST hits, we proceeded to a quick phylogenetic neighbor joining analysis to determine as closely as possible its affiliation. Functional affiliation of the OTUs was then determined mostly based on (Lara and Belbahri, 2011).

Statistical analyses

We analyzed frequency count data derived from OTU clustering to estimate the total (i.e., observed + unobserved) OTU richness in the population from which the observed samples were drawn. These data consisted of the number of OTUs with one representative in the sample (the "singletons"), the number with two representatives, three, and so on. We used two sets of statistical procedures. The first was implemented in the software package Breakaway v2.0 (Willis and Bunge, 2015). This method estimates a family of statistical models, known as "Kemp"-type distributions, from the frequency count data, by fitting ratios of successive frequency counts via nonlinear regression; an optimal selected model is returned, along with error and goodness-of-fit assessments. In the case of the combined (global) OTU data the lowest-order model was selected, which is equivalent to fitting a weighted linear regression to the ratios of successive frequency counts. The second was implemented in the software package CatchAll (Bunge et al., 2012). This method fits a family of mixed-Poisson models to the frequency data, as well as calculating a suite of nonparametric procedures. It returns an optimal selected model with errors and goodness-of-fit assessments, and comparisons of parametric and nonparametric analyses. CatchAll selected a mixture of three exponential-mixed Poisson models. In all cases the results were virtually identical, showing almost no diversity extant in the population beyond that observed in the sample. The R package Vegan (Oksanen et al., 2015) was used to analyze frequency count data derived from OTU clustering. Different functions of Vegan were called to compute diversity indexes (Shannon, Simpson, and Rényi diversities), to randomly subsample our samples (rrarefy function) and to estimate and compare species compositions, using Bray-Curtis distance and NMDS ordination (monoMDS function). Figures were made with R (R Core Team, 2013) and ggplot2 (Wickham, 2009).

Detecting RNA in cysts

To assess if viable and sequenceable RNA can be detected in protist cysts, we designed fluorescent in situ hybridization (FISH) probes specific to the kinetoplastids with 18S rRNA probes coupled to Cy3 fluorochrome and designed to hybridize to ribosomes: KIN516, a wide-spectrum kinetoplastic probe (ACCAGACTTGTCCTCC); and Bodo1757, a probe designed on the V8 region and specific to the clade *Bodo* (CGAGCAAGTGAAACACTCGCC). Probes were checked for specificity using pure cultures of *Neobodo designis* strain AND31 (AY965872) and *Bodo saltans* strain SCCAP BS364. Cultures were fixed with three volumes of paraformaldehyde 4% (PFA) for two hours. Fixed cell suspensions were collected on a 0.2 µm nitrocellulose ISOPORE filter (Millipore, Billerica, MA, USA) and rinsed with demineralized water. Hybridizations were performed at 45°C for 90 min in a buffer containing 0.9 M NaCl, 20mM Tris HCl, and 0,01% SDS, pH 7.2. Probes were used at a concentration of 50ng/µl. The optimal amount of formamide in the hybridization buffer to reduce non-specific binding of the probe was set by performing hybridizations with increasing formamide concentrations for the different strains investigated. In general, hybridization with 35% formamide did not result in any unspecific binding of the probes to non-target organisms. For unspecific staining of all cells, DAPI (4', 6-diamidino-2-phenylindole) was added at a concentration of 1 µg/ml to the hybridization buffer. After hybridization, filters were washed for 15 min at 48°C in a buffer containing 20 mM Tris-HCl (pH 7.2), 10 mM EDTA, 0.01% SDS and 80 mM NaCl, subsequently rinsed with distilled water and air-dried. Prior to microscopic observation, filters were mounted with Citifluor solution. Probes were applied to a soil taken from a garden compost in a garden in Switzerland. Soil was mixed and sieved through a 2 mm grid to remove large debris. The soil was then spread as a 1cm layer on a 16 x 23 cm plastic tray, and wetted with autoclaved demineralized water (without flooding). Two layers of lens-cleaning tissues were spread on the surface, and four microscope cover slips measuring 24 per 50 mm each (Menzel Gläser, Braunschweig, Germany) were laid upon the wet tissues. The tray was left in the dark at 18°C for one week. Protists attached to the coverslips were removed by rinsing with Neff's Amoeba Saline buffer and collected into a 10 ml Falcon tube. Cells were directly fixed with PFA on a nitrocellulose filter (0.2µm) and hybridized using the specific designed probes as well as with DAPI as described above.

Measuring DNA and RNA

Ciliates were measured for the DNA and RNA abundances of the 18S rRNA locus: *Colpoda magna*, from the American Type Culture Collection 50128; *Dileptus* sp., from Carolina Biological Supply Company; *Halteria grandinella*, collected on October 2013 at a pond at the University of Kaiserslautern; *Paramecium tetraurelia* strain D4-2, provided by Martin Simon, Saarland University; and *Tetrahymena thermophila* wildtype B, provided by Josef Loidl, University of Vienna. Clonal populations were cultured with wheat grass powder medium inoculated with the bacterium *Klebsiella minuta*. Standard curve were made with cells from each species were amplified for the 3' end of 18S rRNA with primers 1391-F (Goodfellow and Stackebrandt, 1991) and Euk-B (Medlin et al., 1988). PCR products were cleaned with MinElute PCR Purification Kit from Qiagen and ligated to pGEM®-T vector (Promega, Madison, Wisconsin, U.S.A.) and cloned. Plasmids were isolated with the FastPlasmid Mini Kit (5 Prime, Hilden, Germany), and checked with an amplification using vector primers and Sanger sequencing. Plasmids were diluted to a six tenfold standard for the qPCR ranging from 10^{-1} -10^{-6} ng/µl. For quantitative real-time PCR (qPCR) of RNA, total RNA was extracted 12 separate times from single cells first washed three times in Volvic water, then lysed with the Ambion Single Cell Lysis Kit (Thermo Fisher Scientific) with an integrated DNase step. cDNA was generated using the QuantiTect Reverse Transcription Kit (Qiagen, Venlo, Netherlands) with another integrated DNase step. For quantitative real-time PCR (qPCR) of DNA, measurements were taken directly from 12 separate single cells. Amplifications were performed in a final volume of 20 µl, containing 10 µl iQTM SYBR Green Supermix from Bio-Rad (Hercules, California, U.S.A.), 1.0 µl template DNA (cDNA or a single cell), 10 pg of each primer, and 7.0 µl RNase free water. Cycling conditions were: 95°C for 2 min; and 40 cycles of 95°C for 15s, 57°C for 25s, and 72°C for 40s. Melt curve data were collected after cycle 2 between 57°C and 95°C with a temperature increase at a rate of 0.5°C/10s. qPCR reactions were performed with iQ Single Color Real-Time PCR Detection System (Bio-Rad), with negative control and the six tenfold dilution in triplicates of standard curves plotting against Ct values. The amplification efficiency E was estimated with the equation $E=(10^{-1/slope})-1$ with the calculated slope of the standard curves. We obtained efficiencies E between 90.1% and 105.5% and R^2 between 0.987 and 1.000. Copy numbers were calculated using the following equation: number of copies = (amount × 6.022×10^{23}) / (length × 1×109 × 660) where amount is the concentration of DNA or cDNA (in ng), and length is the length of the linear fragment plus plasmid, 660 is the average molecular weight of one base pair for double strand DNA and 6.022×10^{23} is the molar constant (Zhu et al., 2005).

Results

Sequencing of the Neotropical rainforest soil samples resulted in 132.3 million cleaned V4 reads (Supplementary Table 1). Of the 50.1 million reads assigned to the protists, 75.3% had a maximum similarity of <80% to references in the Protist Ribosomal Reference (PR^2) database (Guillou et al., 2013) (fig. E.1, Supplementary Fig. 1). By contrast, our re-analysis of 367.8 million reads from the hyper-variable V9 region of the 18S rRNA from the open oceans 11 produced only 3.1% reads <80% similar to the PR^2 database, and most of our new V4 and V9 reads from European near-shore marine environments were likewise highly similar to the PR^2 database (fig. E.1, Supplementary Fig. 1). Reads <80% similar to known references are often considered spurious and removed in environmental protist sequencing studies (Stoeck et al., 2010). Three quarters of our rainforest soil protist data would be discarded if we applied this conservative cleaning step. Furthermore, PR^2 and similar databases are biased towards marine and temperate references. To solve these problems, reads were dereplicated into 10.6 million amplicons (= strictly identical reads to which an abundance value can be attached), and subsequently placed with the Evolutionary Placement Algorithm (EPA) (Berger and Stamatakis, 2011) as implemented in RAxML (Stamatakis, 2014) onto a phylogenetic tree inferred from 512 full- length references from all major eukaryotic clades (fig. E.3, Supplementary Fig. 2). We conservatively retained operational taxonomic units (OTUs) constructed with Swarm (Mahé et al., 2014) (Supplementary Fig. 3) whose most abundant amplicon fell only within known clades with a high likelihood-weight score. This novel phylogeny-aware cleaning step effectively discarded highly divergent amplicons (Dunthorn et al., 2014), resulting in the removal of only 6.8% of the cleaned reads and 7.7% of the OTUs. The retained protist reads from the rainforest soils clustered into 26,860 OTUs. As in a recent sampling of the sunlit surface layer of the world's open oceans (De Vargas et al., 2015), more protist OTUs were detected than animals (4,374; of which 39% were assigned to the Arthropoda), plants (3,089), and fungi (17,849) combined (Supplementary Table 1). The OTUs found in the samples may not all correspond to soil-dwelling species: some of these hyperdiverse protists and other eukaryotes could be a shadow of the tree-canopy communities from cells that have rained down from above. Taxonomic assignment of the protists showed that 84.4% of the reads and 50.6% of the OTUs were affiliated to the Apicomplexa (fig. E.2, Supplementary Fig. 4). Apicomplexa are widespread parasites of animals (Desportes and Schrével, 2013; Lee et al., 2000). As an independent line of evidence for their dominance in the Neotropical rainforest soils, ten samples from both Costa Rica and Panama were amplified with primers designed to target specifically the

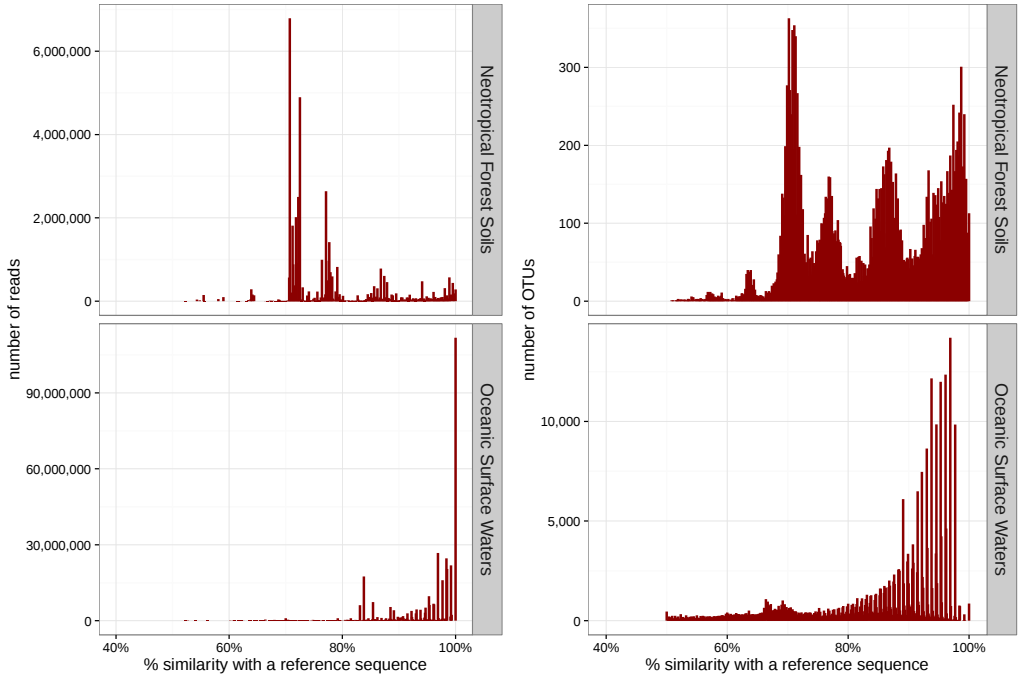

Figure E.1 – Similarity of protists to the taxonomic reference database PR² (Guillou et al., 2013). In contrast to marine data, most of the reads and OTUs from the Neotropical rainforest soils were <80% similar to references in the PR² database. Only 8.1% soil reads had a similarity ≥95% with references, while 68.1% of the marine reads from the Tara-Oceans's study of the world's open oceans had a similarity ≥95% with reference sequences.

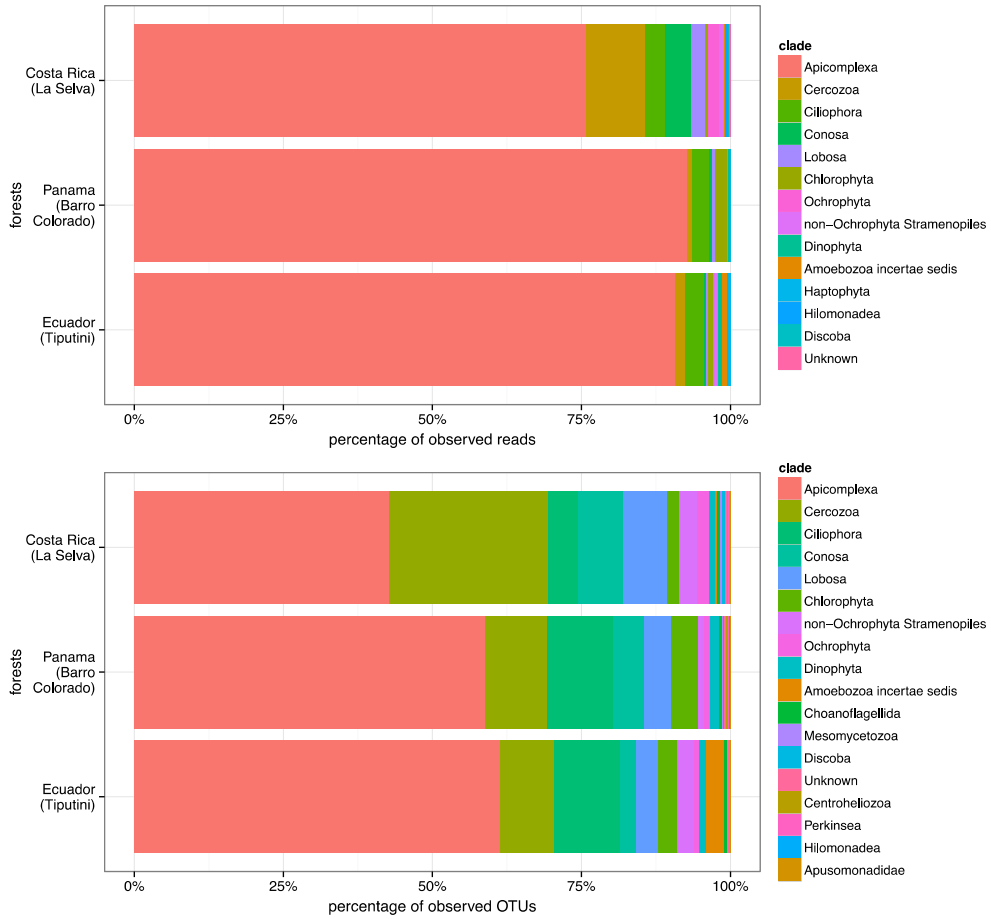

Figure E.2 – Taxonomic identity and relative abundances of soil protist reads and OTUs in three Neotropical rainforests. Each taxa shown represent at least >0.1% of the total data. Between 76.6% and 93.0% of the reads, and between 43.1% and 61.5% of the OTUs, per forest were assigned to the Apicomplexa (in red).

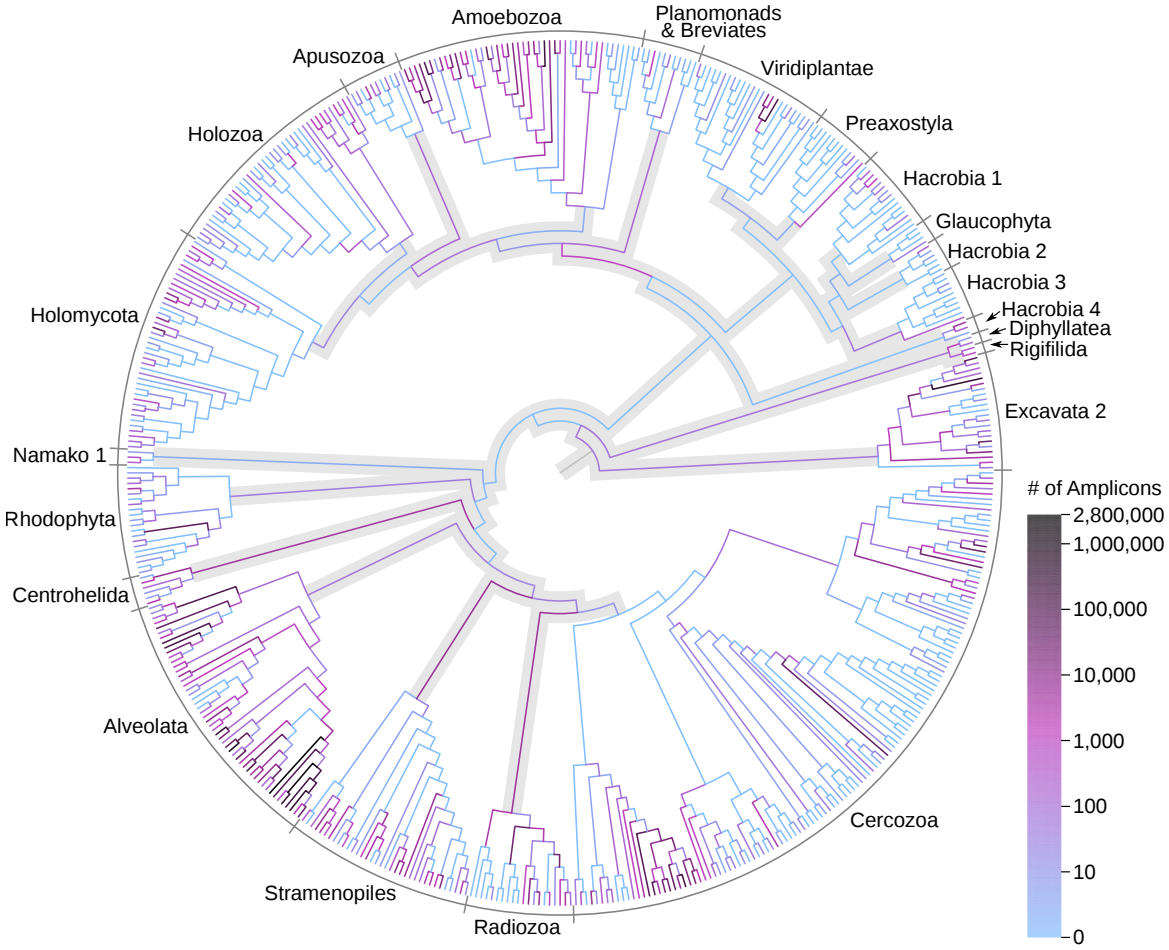

Figure E.3 – Phylogenetic placement of Neotropical soil protist reads on a taxonomically unconstrained global eukaryotic tree. Reads were dereplicated into strictly identical amplicons. Inferred relationships between these major taxa may differ from those obtained with phylogenomic data. Alveolate includes Apicomplexa and Ciliophora; Holozoa includes animals; Holomycota includes fungi. Branches and nodes outside of known clades are shaded in gray. In our conservative approach only OTUs that placed within known clades with high likelihood scores were retained.

closely related Ciliophora (both the Ciliophora and the Apicomplexa are in the Alveolata (Adl et al., 2012)) and sequenced with Roche/454. From these ciliate- specific amplifications, 47.8% of the 297,892 reads and 28.1% of the 1,082 OTUs were assigned to the Apicomplexa (Supplementary Fig. 5). In contrast to the results presented here, read- and OTU-abundances of the Apicomplexa were observed to be substantially lower in marine and other terrestrial environments (Supplementary Fig. 4).

We placed the Apicomplexa OTUs with EPA into a more focused phylogeny inferred from 190 full-length references from all major Alveolata clades. While the OTUs were generally distributed across the whole tree (Supplementary Fig. 6), 80.2% of them grouped with the gregarines. Gregarines predominantly infect arthropods and other invertebrates (Desportes and Schrével, 2013). About 23.8% of these gregarine OTUs placed within the lineage formed by the millipede parasite *Stenophora*, and the insect parasites *Amoebogregarina, Gregarina, Leidyana*, and *Protomagalhaensia*. An additional 13.5% of the OTUs placed with two environmental gregarine sequences collected from brackish sediment (Dawson and Pace, 2002). Many other OTUs grouped within gregarine lineages thought to be primarily parasites of marine annelids and polychaetes. Non-gregarine Apicomplexa OTUs largely grouped with the blood parasites *Plasmodium* (some of which cause malaria) and close relatives, including those that can cycle through arthropods and vertebrates such as birds.

The second through fifth most diverse protist taxa were the predominantly predatory Cercozoa, Ciliophora, Conosa and Lobosa (fig. E.2), accounting for a combined total of 5,572,490 reads and 10,338 OTUs. Fewer photo- or mixo-trophic Chlorophyta, Dinophyta, Haptophyta, and Rhodophyta were also found, accounting for a combined total of 699,187 reads and 1,096 OTUs. Haptophyta are mostly marine (Edvardsen et al., 2016), for example, but a phylogenetic inference of these OTUs showed most had close relationships to freshwater species (Supplementary Fig. 7). Although algae are thought to affect the net uptake of carbon in some terrestrial ecosystems (Elbert et al., 2012), the contribution of these soil-inhabiting algae to the carbon cycle in Neotropical rainforests is unknown. Only OTUs were assigned to the Oomycota. Oomycota are efficient

parasites with flagellated stages that disperse through soils. Using best BLAST hits to GenBank references, we inferred close relationships of many of these Oomycota OTUs to taxa that infect either animals or plants (Supplementary Fig. 8).

Non-metric multidimensional scaling (Supplementary Fig. 9) and Bray-Curtis dendrograms (Supplementary Fig. 10) showed that the protist community composition was slightly more similar among the samples from Costa Rica and Panama than those from Ecuador, reflecting similar diversity patterns seen in animals and plants in Central and South American forests (Gentry, 1993). Within each forest we estimated less than 1,732 unobserved OTUs exclusively from the samples taken (Supplementary Table 2), suggesting that our sequencing depth detected a high fraction of the total diversity within each of the samples. However, within-forest OTU rarefaction curves based on the sample accumulation estimated logarithmic increases without plateaus (Supplementary Fig. 11), and the Jaccard similarity index estimated high OTU heterogeneity between samples even within the same forest (Supplementary Fig. 12). This high heterogeneity between samples indicates that our sampling effort unveiled only a fraction of the protist hyperdiversity in the three Neotropical rainforests.

Discussion

The broad sampling and deep sequencing of soils from three Neotropical rainforests revealed numerous protist taxa. Much of this unravelled diversity was detected by our phylogeny-aware cleaning step that retained OTUs with low similarity to existing 18S rRNA references (fig. E.1, fig. E.3). We show that using EPA phylogenetic placements, rather than relying solely on pairwise sequence similarities, represents a powerful approach in metabarcoding studies aimed at discovering novel taxa in environments where we know very little and very few references are available—such as tropical soils. Even if almost all the tips are missing in the tree (i.e., there are no references closely related to the environmental data), a tree that comprises at least the major eukaryotic lineages allows phylogenetic placements to retain most of the novel diversity that would have otherwise been discarded. While the placements can be evaluated at the amplicon or OTU levels, placing only the OTU representatives produced by Swarm drastically reduces computation time without significant information loss.

Even with the phylogeny-aware cleaning step, our molecular approach likely underestimated diversity across all microbial eukaryotic and macro-organismic taxa. For example, while the broadly-targeted V4 primers used here amplify a wide variety of eukaryotic lineages (Filker et al., 2015; Forster et al., 2016a; Hu et al., 2016; Logares et al., 2014; Richards et al., 2015; Stoeck et al., 2010), other primers could have amplified additional taxa (Adl et al., 2014; Hu et al., 2015; Lentendu et al., 2014) (Supplementary Table 1). Additionally, the resolution power of metabarcoding has its limits. The Swarm clustering method, which relies on iterative local clustering thresholds, often results in fine-grained OTUs that can reveal additional diversity in comparison to traditional clustering methods that rely on global clustering thresholds (Forster et al., 2016b; Mahé et al., 2014) (Supplementary Fig. 3). Species with identical barcode sequences (i.e., no variation between taxa) can nevertheless be indistinguishable by any clustering method, which may mask potentially different ecological roles and functions.

The parasitic Apicomplexa dominated the Neotropical soil protist communities by accounting for a majority of both reads and OTUs (fig. E.2). This pattern of dominating parasites contrasts with the prevailing view that soil protist communities are dominated by predators of bacteria (Adl and Gupta, 2006), although a considerable presence of protist predators of fungi and animals as well as protist parasites have been observed elsewhere (Bates et al., 2013; Dumack et al., 2016b; Dupont et al., 2016; Foissner, 1993; Geisen et al., 2015b; Grossmann et al., 2016; Mitchell, 2015). These dominating protist parasites also potentially contribute to the high animal diversity in the rainforests by the same mechanisms that other parasites contribute to high tree diversity as hypothesized in the Janzen-Connell model (Connell, 1970; Janzen, 1970). Apicomplexa infect animals (e.g., (Asghar et al., 2015; Bouwma et al., 2005)) and are usually host-specific, although host switching is known (Desportes and Schrével, 2013; Ellis et al., 2015; Lee et al., 2000; Rueckert et al., 2011). They therefore potentially limit population growth of species that become locally abundant. In this tentative model, we expect: Apicomplexa to dominate both the below- and above-ground protist communities in tropical forests worldwide, most animal populations in these forests will have apicomplexan parasites, and animals that become locally dominant will have escaped from their apicomplexan enemies or at least evolved to be able to cope with this burden. This relationship between animals and the Apicomplexa could also be reciprocal, with each group contributing to the diversity of the other. This top-down force by parasitic protists on tropical animal diversity may complement the bottom-up response that was proposed for plants on herbivorous insects (Basset et al., 2012).

In contrast to the Apicomplexa parasites, few parasitic Oomycota parasites were detected in the Neotropical soil protist communities (Supplementary Fig. 8). Along with fungi and insects, the Oomycota have long been thought to be one of the drivers of the Janzen-Connell model. However, the degree of host-specificity is unknown for most Oomycota species (Freckleton and Lewis, 2006), and a fungicide study in Belize documented a non-significant effect by these protists (Bagchi et al., 2014). If the Oomycota broadly drive the Janzen-Connell

model, then we can expect their diversity to be very high, mirroring tree diversity. We found too few Oomycota OTUs, though, to be host-specific plant parasites under this model (but as mentioned above, OTUs can mask functional diversity).

Just as there is high species richness in Neotropical rainforests at the macro-organismic scale, the high OTU numbers of both free-living and parasitic protists (Supplementary Table 1) and the high heterogeneity between samples (Supplementary Fig. 12) show that there is a similar but greater pattern of hyperdiversity at the microbial eukaryotic scale. Given the low similarity of protist compositions between samples and the six-fold ratio of protist to animal OTUs, a concerted and comparable count will likely show that protists are more diverse than arthropods in tropical rainforests. This would certainly be the case if, as we suggest, every arthropod species has at least one apicomplexan parasite, and the Apicomplexa are only one part of total protist hyperdiversity (with fungi being the second most diverse group). If protists are the most diverse eukaryotes in tropical rainforests it would not be due to inordinate speciation in just a few clades since the mid-Phanerozoic (e.g., the beetles (Farrell, 1998)), but because of the diversification of rich and functionally complex protist lineages beginning in the early Proterozoic (Knoll, 2014) that built up the multifaceted and interactive unseen foundation of these now familiar macroscopic terrestrial ecosystems.

Acknowledgements

We thank Thorsten Stoeck, Laura Katz, Håvard Kauserud, and Owen T. Lewis, and David Montagnes for helpful comments. Funding primarily came from the: Deutsche Forschungsgemeinschaft grant DU1319/1-1 to M.D.; Klaus Tschira Foundation to L.C., A.S and A.K; and French Government "Investissements d'Avenir" program OCEANOMICS (ANR-11- BTBR- 0008) to C.d.V, C.B., and S.R. Funding also came from: Natural Environment Research Council (NERC) grants NE/H000887/1 and NE/H009426/1 to D.B., the National Science Foundation's International Research Fellowship Program (OISE-1012703) and the Smithsonian Tropical Research Institute's Fellowship program to J.M., the Swiss National Science Foundation (310003A 143960) and the Swiss Federal Office for the Environment and Swiss National Science Foundation to E.L. and E.A.D.M. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de).

Author Contributions

F.M. and M.D. conceived the project. F.M., C.dV., J.M., T.S., S.R. and M.D. collected the samples. Sequencing was carried out by I.T. and S.R. Data analysis was done by F.M., D.B., L.C., A.S., E.L., D.S., J.B., S.S., I.T., C.B., A.K., E.E., and M.D. The first draft of the manuscript was written by F.M., C.dV., D.B., L.C., A.S. and M.D., and all authors contributed to discussing the results and editing the manuscript.

Appendix F

Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling

Christophe V.W. Seppey[1], David Singer[1], Kenneth Dumack[2], Lassaâd Belbahri[1], Edward A.D. Mitchell[1,3], Enrique Lara[1]

[1]Laboratory of Soil Biology, University of Neuchâtel, Rue Émile-Argand 11, 2000 Neuchâtel, Switzerland
[2]Department of Terrestrial Ecology, University of Cologne, Zülpicher Strasse 47b, Köln, Germany
[3]Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

Soil Biology and Biochemistry, accepted

Abstract

Micro-eukaryotes play many key roles in soils as they function as decomposers, parasites, predators and primary producers, but their diversity and functional ecology are still poorly known. High-throughput sequencing methods, allow now assessing the full diversity of soil micro-eukaryotes. The resulting operational taxonomic units (OTUs) can be assigned to taxonomic and functional identities using increasingly complete reference database.

Using Illumina sequencing of the V9 region of the SSU rRNA gene, we investigated the taxonomic and functional diversities of soil micro-eukaryotes in three categories of land-use: forests, meadows and croplands located in Switzerland. We assigned each OTU to a broad functional category (e.g. phototrophs, phagotrophs, osmotrophs, and parasites). We then compared the distribution patterns of these categories to infer trophic linkages beyond the classical model of the "soil microbial loop" in which protists prey on bacteria and release nutrients to be taken up by vascular plants.

OTU richness was similar for every taxon in the three land-use categories. However, the proportion of fungal sequences was highest in forests and thus Shannon diversity lowest as compared to meadows and croplands, and community structure was also different between forests and the other two land-use types. Phototroph sequences were less abundant in forests than in meadows and croplands while the proportion of different phototroph higher taxa was similar among the land-uses. Several OTUs representing phagotrophic protists, together accounting for more than 25% of all phagotroph sequences, were significantly correlated to the total number of phototroph sequences (i.e. soil microalgae). At least four of these corresponded to acknowledged algal predators and this was further confirmed by microscopic observation of isolated taxa.

These results suggest that beyond plants, microalgae represent a functionally significant but rarely considered input of carbon in soils that should be taken into account when modelling soil nutrient cycling.

Keywords: eukaryotic micro-algae, phagotrophic protists, carbon cycling, V9 region of the SSU rRNA gene, high-throughput sequencing

Introduction

Soil microbial eukaryotes, including protists and fungi, are key actors of biogeochemical cycling, are involved in numerous biotic interactions (van Hannen et al., 1999; Verni and Gualtieri, 1997), and are thus considered a key element in soil fertility. Their first recognised functional role was grazers of bacteria leading to the "soil microbial loop" paradigm, according to which protozoan grazing on soil bacteria releases labile compounds such as ammonium that stimulate plant growth (Bonkowski and Clarholm, 2012; Clarholm, 1985). This concept has since been refined by suggesting an antagonistic role of mycorrhizae and the possible selection of Plant Growth Promoting Rhizobacteria (PGPR) by protists (Bonkowski, 2004; Bonkowski and Brandt, 2002). However, while feeding on bacteria is unquestionably widespread in phagotrophic microbial eukaryotes, eukaryovory (i.e. the act of feeding partially or exclusively on other eukaryotes) is also common (Dumack et al., 2016a,b; Geisen et al., 2016). Other protists are highly specialised predators of eukaryotes. For example, Grossglockneriid ciliates feed exclusively on fungi (Petz et al., 1985). Parasitoids are also frequent in soils, including the widespread but still poorly studied Rozella group (also known as "Rozellida"; (Lara et al., 2009a) or Cryptomycota (Jones et al., 2011)) which prey on chytrids, oomycetes and green algae and also include endo-nuclear parasites of Amoebozoa that ultimately cause cell death and lysis (Corsaro et al., 2014a). In those cases, nutrient release by protists does not rely on bacterivory, implying pathways for nutrient cycling alternative to the microbial loop. It is unclear how quantitatively relevant this pathway is but one way to assess this is to study the diversity and abundance of taxa involved in these trophic relationships using the now available data from massive sequencing of soil environmental DNA.

The true diversity of soil protists has long been poorly know, mainly due to methodological limitations for their isolation, culture and subsequent identification (Ekelund and Ronn, 1995; Foissner, 1999b). Metabarcoding (environmental DNA amplicon based identification) of high throughput sequencing data is now the golden standard for environmental screening of microbial diversity (Pawlowski et al., 2016), and also provide information on the functioning of ecosystems based on the genetic identification of the organisms and knowledge on their lifestyles (De Vargas et al., 2015; Lara et al., 2015; Massana et al., 2014). The next step is to infer the trophic relationships between these organisms. In practice, the nature of these relationships (i.e. trophic, but also symbiotic, exploitation of a similar resource, etc.) is not known, and it is thus difficult to interpret in biological terms. Examples of known relationships taken from the literature can however illustrate well supported co-occurrence and clarify the true nature of these relationships between organisms (Dumack et al., 2016b; Hess and Melkonian, 2013; Hess et al., 2012). Putative relationships inferred from metabarcoding studies can also be explored by conducting new observations and experiments.

Phototrophic protists (i.e. eukaryotic algae) in soils include mostly exclusive phototrophs (e.g. Bacillariophyta, Chrysophyceae, Xanthophyceae) and photosymbiont (e.g. Trebouxiophyceae). Eukaryotic algae constitute an important part of the so-called cryptogamic crusts which represent a significant carbon input in arid ecosystems (Elbert et al., 2012; Freeman et al., 2009; Frey et al., 2013). They are however also widespread in more humid soils but their functional role there is less well known and consequently, have not been considered in the classical models of soil microbial loop (Berard et al., 2005).

Our aim was to 1) compare the overall diversity and community structure of soil micro-eukaryotes in forest, meadow and cropland soils from 44 sites in Switzerland based on Illumina sequencing of the v9 region of SSU rRNA gene, and 2) compare the abundance of eukaryotic algae to that of individual phagotrophic taxa (OTUs) in order to infer potential trophic linkages.

Materials and methods

Sampling

We collected 44 soil samples in permanent plots of the Swiss Biodiversity Monitoring program http://www.biodiv ersitymonitoring.ch/en/home.html. The sites included 16 forests, 16 meadows and 12 croplands (fig. F.1, Table S1). Each sampled habitat was characterized using the typology of Swiss natural habitats (Delarze et al., 2015) (Table S1). Forests included both coniferous (e.g. Picea abies), or broadleaf (e.g. Fagus sylvatica). Most of meadows were amended and used to produce fodder. Croplands were used for maize, cereals or tobacco cultivation. Meadows and croplands were designed as open habitats as much more light reach their soil surface than in forests. Sampling was performed over one month between September 27th 2012 and October 31th 2012. At each site, three topsoil cores (5cm diameter x 5cm depth, without the litter layer) were taken along a circle of 1m radius in the same land-use and pooled. Soil samples were kept cool (in an icebox) and DNA was extracted within 2-3 days.

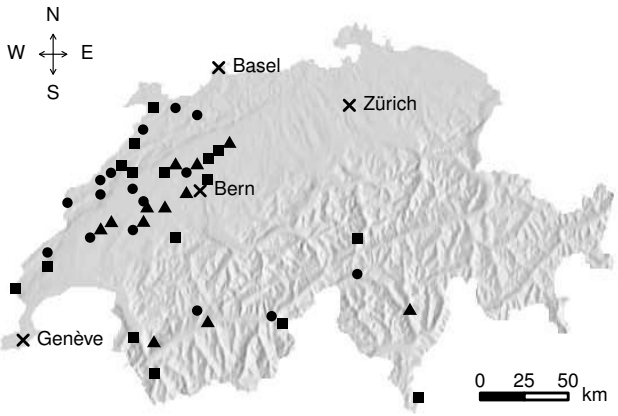

Figure F.1 – Location of the 44 sampling sites in Switzerland. The squares, circles and triangles indicate forests, meadows, and croplands, respectively.

DNA extraction, amplification and sequencing

DNA was extracted using the MoBio PowerSoil extraction kit (Carlsbad, CA, USA) according to the manufacturer instructions. The SSU rRNA V9 region was amplified using the broad spectrum eukaryotic primers 1380F/1510R (CCCTGCCHTTTGTACACAC / CCTTCYGCAGGTTCACCTAC) (Amaral-Zettler et al., 2009). PCR reactions were run in triplicates with a PTC-200 Peltier Thermo Cycler (BioConcept, Allswill, Switzerland) with 1ng of environmental DNA, 6µL of 10 x PCR buffer, 0.6µL of each primer, 0.6µL of each dNTP 400µM (Promega, Dübendorf, Switzerland) and 0.2µL of 0.05 U/µL GoTaq (Promega, Dübendorf, Switzerland). The volume was adjusted to 30µL with ultra-pure water. Amplification was conducted with the following conditions denaturation at 94°C for 3 min, 30 cycles at 94°C for 30 s, 57°C for 60 s and 72°C for 90 s and final extension at 72°C for 10 min (Amaral-Zettler et al., 2009). PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and pooled together at the same concentration prior to sequencing. A DNA library was prepared using the New England Biolabs's kit NEBNext DNA Sample Prep Master Mix Set 1. Illumina HiSeq sequencing was done by Fasteris (Geneva, Switzerland) using an Illumina HiSeq 2000 technology to obtain paired-end reads (2 X 100bp).

Bioinformatic analyses

The PR^2 database (Guillou et al., 2013) was used as the reference database for a first taxonomic assignation of reads and OTUs; only sequences containing complete forward and reverse primers described above were retained. SSU sequences from bacteria and archaea were also added to the PR^2 database from the Silva database (Pruesse et al., 2007) in order to identify and remove eventual prokaryotic sequences from the analysis. Prokaryotic SSU sequences were truncated from the general primer 1389F (TTGTACACACCGCCC) (Amaral-Zettler et al.,

2009) to the end of the SSU rRNA sequence and kept as ortholog of the eukaryotic V9 fragment. The truncated prokaryotic sequences were then de-replicated before being added to the PR^2 database.

Reads were merged using the program Flash v. 1.2.9 (Magoc and Salzberg, 2011) and demultiplexed into samples using the program Sabre `https://github.com/najoshi/sabre`. Good quality sequences were selected with a custom script inspired from the method used in (De Vargas et al., 2015). Chimeric sequences were then discarded using the software Usearch v. 7.0.1090 (**?**) by comparing reads against the PR^2 (Guillou et al., 2013) and Silva (Pruesse et al., 2007) databases and against reads within the sample. Only sequences containing complete forward and reverse primers described above were kept. In order to remove artefactual sequences, we kept only those that were found at least three times in two samples (De Vargas et al., 2015).

OTUs were clustered using the software Swarm v. 1.2.5 (Mahé et al., 2014) with the default set-up. OTUs were then taxonomically assigned by aligning the dominant sequence of every OTU against the PR^2 database using Ggsearch (Fasta package v. 36.3.6: `http://faculty.virginia.edu/wrpearson/fasta/CURRENT/`). The OTUs were considered as undetermined eukaryotes if their percentage identity with sequences of PR^2 was lower than 80% as in (De Vargas et al., 2015). We also removed sequences belonging to prokaryote, Metazoa or Embryophyceae. In order to homogenize the number of reads present in all samples for further numerical analyses, we randomly selected 50'000 for each sample using a custom script.

Assignation to functional groups and numerical analyses

We selected 41 taxa characterized by having the same trophic function (i.e. 5 osmotrophs, 5 parasites, 6 phototrophs, 25 phagotrophs) in the list of division, class and order the PR^2 assignation for the diversity analyses (Table S1).

As a first comparison of community composition, we calculated the Shannon index and performed a non-metric multidimensional scaling (NMDS) analysis on the OTUs abundances. We assessed the difference in diversity among land-use with a non-parametric multiple comparison Nemenyi test (Hollander and Wolfe, 1999), and also calculated NMDSs for each pairs of land-use and tested the community difference by a permutation test (envfit function vegan package v. 2.0-10 (Oksanen et al., 2015)). P-values were multiplied by three to take into account multiple tests adjustment (Holm, 1979).

We then assessed in which environment sequences belonging to phototroph organisms were most abundant using a Nemenwi test (posthoc.kruskal.nemenyi.test function, package PMCMR v. 4.1 (Pohlert, 2014)). To retrieve putative algae consumers, we measured the correlation between each of the 100 most dominant phagotroph OTUs and the total abundance of phototrophs, taking also into account land-use as second environmental variable in linear models (LM). To normalize the distribution of both phagotroph OTUs and total phototroph abundance we log transformed (x' = log10 x) their sequences abundances. The correlation of the two environmental variables (i.e. total phototroph abundance, land-use) were kept independent as none of the model tested showed significant interaction. We finally adjusted the p-values of the two environmental variables for the 100 models according to Holm, 1979. We also verified if each of the LM respected conditions of residuals normality and homoscedasticity by performing a Shapiro test on the models residuals and non-constant variance test (function shapiro.test and ncvTest; packages car v. 2.0-20; (Fox and Weisberg, 2010), and stats v. 3.1-0; (R Core Team, 2013) respectively). OTUs respecting the LM conditions and showing a significant correlation with phototroph abundance, were selected as putative algae consumer and their taxonomy was verified on GenBank by using Blast with the default parameters.

Isolation of protists and microscopic observation on algivorous behaviour

To backup our data by microscopical observations, we documented organisms from the same genus/species as the OTUs whose abundances were positively and significantly correlated with those of algae. *Rhogostoma* sp. was isolated from leaf surfaces (Cologne, Germany), Leptophrys vorax was isolated from a freshwater puddle (Cologne, Germany), and the undetermined chrysophyte and Trinema sp. appeared as a contamination in such protist cultures. All protists were morphologically determined.

The pictures of Leptophrys vorax were obtained from an individual directly taken from the natural sample. Other organisms were cultured in Waris-H (McFadden and Melkonian, 1986) at room temperature on a window bench and enriched with *Characium* sp. and a not further determined coccal green alga. The cultures were checked for potential algal ingestion after three days of incubation, using an inverted microscope (Nikon Eclipse TS-100, Japan) at 100x and 400x magnification. Pictures were taken with a Nikon digital sight DS-U2 camera (program: NIS-Elements V4.13.04) and a Nikon Eclipse 90i (DIC, up to 600x magnification).

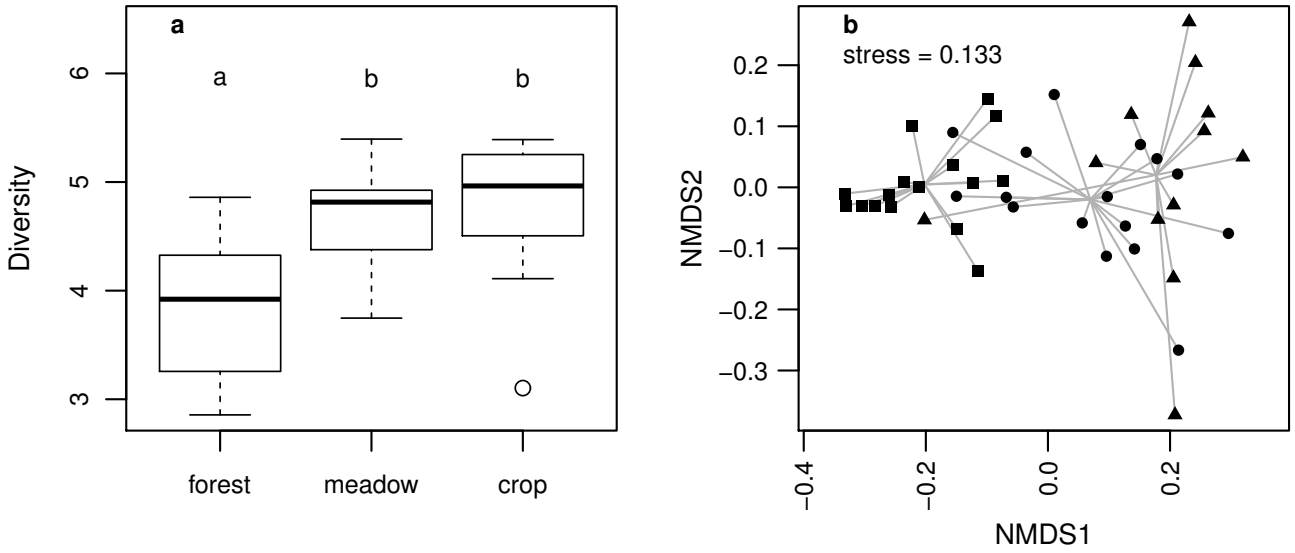

Figure F.2 – Distribution of OTUs Shannon diversity for each environments (a) and non-metric multidimentional scaling (NMDS) based on Bray-Curtis dissimilarities of 44 soil samples from Switzerland. The letters above the boxplots represent groups of environments expressing significant different diversity distribution according to a Nemenyi test ($P < 0.05$). The three land-uses are denoted by squares (forests), circles (meadows) and triangles (croplands).

Results

Data quality and overall diversity

The full dataset contained 15'365'116 raw reads, of which 93.9% passed the quality check, 87.5% were found at least three times in two samples, 87.4% were not considered as chimeras, and 77.4% were not considered as Metazoa, Embryophyceae or prokaryotes. Therefore, a total of 11'893'592 reads were left for further analyses. In the dataset adjusted to 50'000 sequences by sample, we retrieved a total of 18'586 OTUs, of which 87% could be taxonomically assigned unambiguously according to the assignation threshold; altogether, representing 97% of the reads and 75% of the OTUs (Fig. S1, Fig. S2). The most abundant supergroup of eukaryotes in all samples were Fungi, followed by Rhizaria and Stramenopiles.

The most noticeable difference in relative abundance of taxa could be observed between open and forest habitat, and was mostly due to a divergence in the abundance of Basidiomycota. These Fungi were strongly dominant in forest ecosystems with almost 50% of all sequences, mostly due to the presence of a single OTU (X3), which represented 38% of the totality of all reads in forest (Fig. S1). In contrast, richness did not differ deeply between land-use types, and varied between 2371 and 3516 OTUs. Richness was dominated by both Fungi and Rhizaria, more or less in equal proportions, followed by Stramenopiles (Fig. S2).

Shannon diversity and micro-eukaryotic communities differed significantly between forest and open habitats (meadows and croplands) (Nemenyi test, and permutation test on NMDS after correction, $P < 0.001$) while diversity and communities did not differ significantly between meadows and croplands ($P > 0.05$; fig. F.2 and fig. S3).

Diversity and abundance patterns of phototrophs

Sequences belonging to OTUs assigned to phototroph organisms accounted for 1.9% (42'475 sequences) of all sequences. This proportion was highest in open habitats, representing 3.5, 1.8 and 0.9% of the sequences found in croplands, meadows and forests, respectively (fig. F.3). As for overall community patterns, this difference was statistically significant between forests and the two other land-uses (Nemenwi test after correction, $P < 0.01$; fig. F.3). The diversity of phototrophic micro-eukaryotes was largely dominated by Chlorophyceae, followed by diatoms (Bacillariophyta), and Trebouxiophyceae or Xanthophyceae; the rest being shared by other typical subaerial algae like Ulvophyceae and other Archaeplastida (fig. F.3).

On their side, OTU assigned to phagotroph organisms accounted for 32% (704'308) of all sequences. Seven of the 100 most dominant phagotroph OTUs showed a positive correlation to total phototroph sequence abundance, and respected the conditions of residuals normality and homoscedasticity (fig. F.4 Fig. S4 and Table S4). These OTUs belong to Cercozoa (X2, X117, X64, X54), Ciliophora (X321) and Stramenopiles (X12, X343) and together account for 26% of the phagotroph sequence abundance and 8% of the sequences of the dataset (Table S3). Apart from the Labyrinthulea OTU (X343), all other OTU sequences obtained a good match ($\geq$ 97%) with

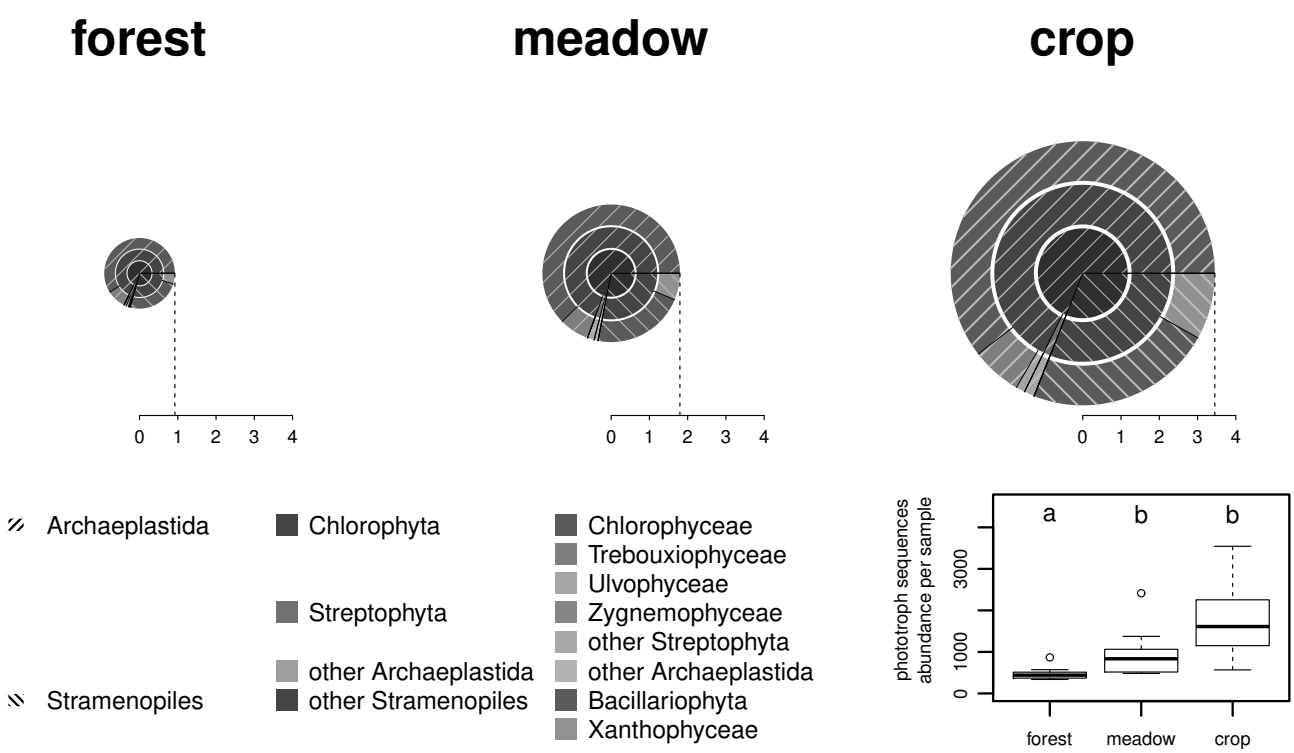

Figure F.3 – In the upper part, the relative abundance of phagotroph taxon sequences in the three land-use types. The radius of the pie-chart represents the percentage of phototroph sequences in each land-use. Taxa representing less than 1% of the land-use are represented in the "other Streptophyta" or "other Archaeplastida" section. On the bottom right part, boxplots representing the abundance of phototroph sequences according to the land-use. Letters on the top part of the boxplots represent the groups of land-use formed according to the Nemenyi test ($P < 0.05$) on the phototroph sequences abundances.

GenBank database sequences (Table S3). In addition to these seven OTUs an eighth one, assigned to the Glissomonadida (Viridiraptor), retained our attention (fig. F.4, Table S3 and S4, and fig. S4). Indeed, despite the fact this OTU did not respect the conditions of homoscedasticity, it was significantly correlated with total phototroph abundance in addition to being among the ten dominant phagotroph OTUs.

Morphological analyses

We successfully screened natural samples for four of the eight correlating phagotrophic protist taxa. All four taxa were observed with most likely ingested algal material, either in the natural samples (*Leptophrys vorax*), or when incubated with algal cells (*Rhogostoma* sp., *Trinema* sp.) (fig. F.5).

Discussion

Overall diversity and community patterns

The high abundance of Fungi in soil environmental DNA surveys reconfirms previous observations (Behnke et al., 2011; Geisen et al., 2015a; Glaser et al., 2015; Lesaulnier et al., 2008); in forests, this dominance is the consequence of a single OTU called here X3, which represented 38% of the totality of all reads in forest. This phylotype shared 100% identity with a wide array of Fungi that build ectomycorrhiza with trees (e.g. *Leucopaxillus*, *Ampulloclitocybe*) (Cairney and Chambers, 1999). It is therefore likely that OTU X3 represents a whole array of mycorrhizal fungi that established symbiontic relationships with the trees. The next most represented groups (Stramenopiles, Rhizaria) comprise also organisms that can be encountered often in soils, such as oomycetes, cercomonads and chrysophytes (Lesaulnier et al., 2008). Fungi, despite of being by far, the most frequent microbial eukaryotes in soils, had a richness which was comparable to Rhizaria (Fig. S2). This is most probably due to the highly ramified hyphae, which have an important biomass, in comparison to the mostly small rhizarian unicells.

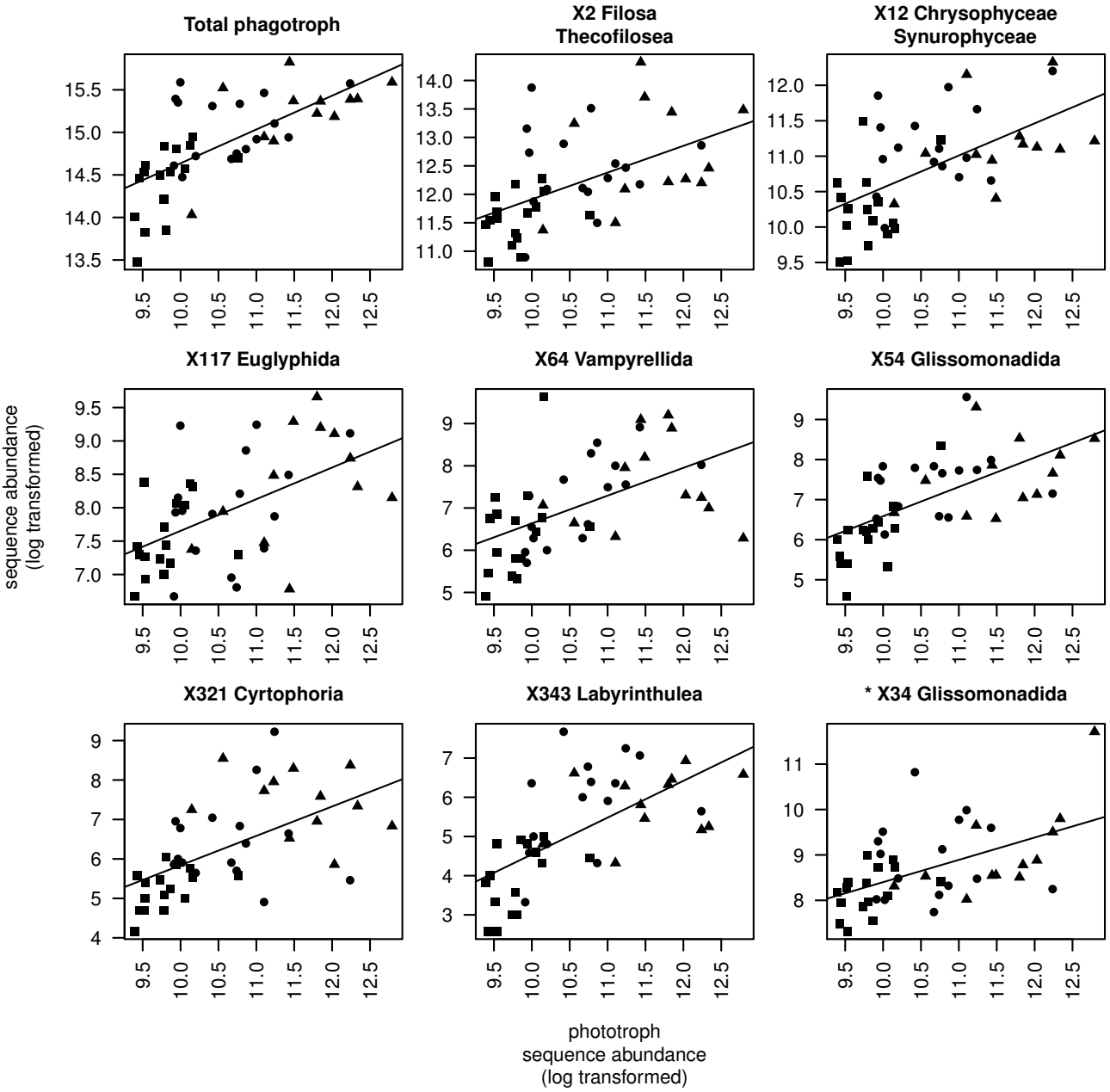

Figure F.4 – Biplots showing the regression between the abundance of total phagotroph and eight OTUs and phototroph abundance. The identifiant of each OTU is shown on the top of each graphic. The asterisk indicate the OTU which belong to the ten most dominant phagotroph OTU and respond significantly to the phototroph despite an heteroscedastic distribution.

Phagotroph vs phototroph

The abundance of seven OTUs representing phagotrophs was correlated to the total abundance of phototroph OTUs (fig. F.4). Six of these could be assigned taxonomically to genus level according to a threshold of 97% identity (Table S3). Amongst these species, four (*Rhogostoma*, *Platyreta*, *Trinema* and *Pseudochilodonopsis*) are large sized protists - and thus potential predators of micro-eukaryotes, including phototrophs. Our pictures illustrate three of these species in the act of predating algae (fig. F.5). *Rhogostoma* spp. (Thecofilosea) (X2) is closely related to recently characterized fungi- and algivores that have been shown to despise bacteria (Dumack et al., 2016b). Although *Rhogostoma* spp. are known to feed on bacteria (Howe et al., 2009), we could show *Rhogostoma* to feed on algae as well. *Platyreta* (X64), which is described to feed on fungi is a member of the exclusively eukaryvorous Vampyrellida, which are best known as algal predators (see fig. F.5, represented by the closely related *Leptophrys vorax*) (Bass et al., 2009a; Hess et al., 2012). Our results indicate that the terrestrial Vampyrellida, although described as predominantly fungivorous, might indeed feed on algae. *Trinema* spp. (Euglyphida) (X117) is known to feed on bacteria, but, for the larger taxa, fungi and micro-algae

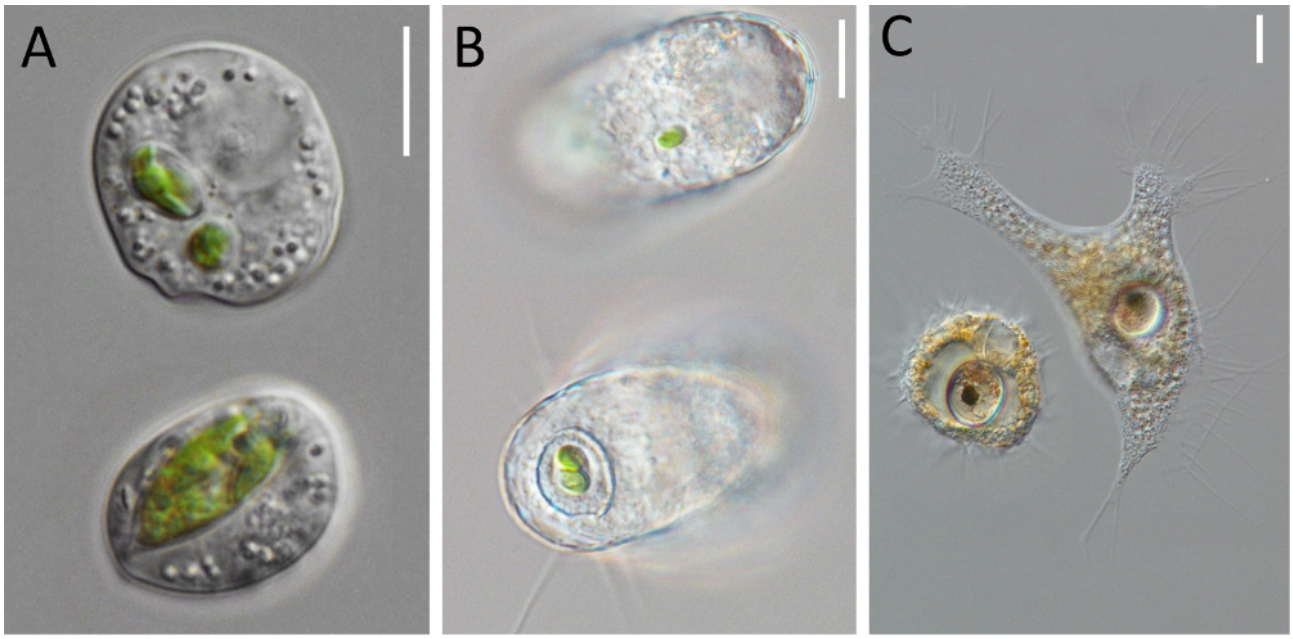

Figure F.5 – Pictures of three selected organisms, closely related to the found OTUs correlating to the phototroph sequence abundances (*Rhogostoma* sp. (a), *Trinema* sp. (b), *Leptophrys vorax* (c)). The scale bar represent 10µm.

(cyanobacteria and/or pigmented eukaryotes) (Meisterfeld, 2000a). Our observations confirmed the ingestion of algal material (fig. F.5). *Pseudochilodonopsis* (X321) are considered as exclusive algivores specialized on diatoms (Hamels et al., 2004). Labyrinthulomycetes from the Amphifilidae clade (X343) are a diverse group including bacterivores such as Sorodiplophrys stercoraria and Amphifila marina (Anderson et al., 2011). The taxonomic as well as functional diversity of this group is however only marginally documented, and the existence of algivorous forms is thus possible. The group of *Spumella*-like chrysophyte (X12) is composed of small phagotrophic flagellates having lost their photosynthetic abilities secondarily. However, it has been shown that transitions between phagotrophic and phototrophic strategies occurred often in the evolutionary history of chrysophytes. It is possible therefore that the *Spumella*-like chrysophyte X12 is actually mixotrophic like many chrysophyceae (Boenigk et al., 2005b), and therefore shares higher light requirements with other phototrophs. Alternatively, it is possible that the *Spumella*-like chrysophyte X12 feeds preferentially on bacteria that are associated to phototrophs and their exudates. Bacterial communities associated to algae are highly influenced by the host in aquatic systems (Sapp et al., 2007). A similar explanation could possibly be given for *Allapsa* (X54), a genus of small cercozoan flagellates formerly collectively classified under the name "*Heteromita globosa*" (Howe et al., 2009).

To the contrary, OTU X34 is assigned to the Viridiraptoridae, a family of highly specialised organisms feeding as yet known exclusively on phototrophic organisms (Hess and Melkonian, 2013). The LM obtained for this OTU did not respect the conditions of homoscedasticity because of its high abundance in two samples. Such high sequence abundance may correspond to local blooms of these small flagellates, which are reported as frequent (Hess and Melkonian, 2013).

Altogether, phagotrophic sequences belonging to an OTU co-occurring with phototrophs reaches 26.9% of all phagotrophs (28.1% if X34 is considered). Under a pessimistic scenario, if only those organisms that have been observed eating algae are playing actually that role, 19.8% of all phagotrophs actually correlate with phototrophs in a putative trophic way. The total amount is estimated between one fifth and one third of all phagotrophic sequences, an amount which is far from being negligible. This fifth to third of phagotroph sequences (139'453 to 197'911 sequences) represent from three to five time the amount of phototrophic sequences. This ratio suggests that the standing biomass of soil microalgae is lower than that of their predators (Giner et al., 2016). By analogy to aquatic ecosystems, this can be explained by the faster turnover of phototrophs. It is therefore possible that the phagotrophs whose abundance is significantly correlated to that of phototrophs indeed primarily feed on them, although this correlation does not constitute a formal proof.

Our data suggest that eukaryotic phototrophs represent an alternative nutrient source for phagotrophic protists in soils, in addition to the traditional "soil microbial loop". As in marine systems (De Vargas et al., 2015), metabarcoding of soil micro-eukaryotes shows that trophic interactions are much more complex than previously thought, and integrate components that have hardly been studied before. These trophic relationships inferred from correlative analysis of metabarcoding data need to be further explored. Combining traditional

culture-based approaches, direct microscopic observation, high throughput sequencing of bulk soil samples or isolated phagotrophs, as well as new statistical tools will be the next steps for unravelling other potential relationships between protists. Nevertheless, we argue that what is now most needed is to characterise the many unknown OTUs, and conducting good observations on these organisms to provide useful natural history background needed for sound interpretation of HTS data. As suggested by our study, we believe, that future studies providing exact identities of the huge amount of unknown OTUs and suggesting their life styles and ecology, will provide sound interpretation of the ever-increasing massive sequencing data.

Acknowledgements

We would like to thank Kathleen Hasler for the field work, and Marion Cartier and Christophe Paul for the laboratory analysis. We would also want to thank the BDM program for providing basic data about the sites. This study was partly funded by Swiss National Science Foundation projects no. 310003A 143960 to EL. LB is supported by the European Union's Seventh Framework Programme under grant agreement 245268 (ISEFOR). The authors declare no conflict of interest.

Appendix G

Curriculum Vitae: David Singer 03.01.2017

SINGER David

Date of birth :	26.10.1985	Phone :	+41 79 218 24 26
Nationality:	Switzerland	Email:	david.singer@unine.com
Family status :	Single	Website :	http://www.david-singer-biologist.com/
Address :	Rue Daniel Jeanrichard 35 2400 Le Locle Switzerland	Skype:	david.singer85

Biologist specialised in protistology

Research Interest

My research areas are soil biodiversity, community and ecosystem ecology. On-going projects focus on metabarcoding and phylogeny approach of micro-eukaryotes to understand ecological interactions, biogeography patterns on different ecosystems.

Education

2012-Present **PhD Student**, Laboratory of soil biology, University of Neuchâtel, Switzerland (Project SNF: Is everything everywhere? A metabarcoding approach to protist ecology and biogeography.)

2009-2011 **Master's degree,** Master in Biogeosciences, University of Neuchâtel, Switzerland

2005-2008 **Bachelor's degree**, Bachelor in Biology, University of Neuchâtel

Skills and qualifications

Language French: Native speaker, English: Professional working proficiency

Laboratory Environmental DNA extraction, PCR, Cloning/sequencing technics, Elemental analysis (AAS, XRF, XRD, CHN), Soil Analysis (pH, CEC, $CaCO_3$, laser granulometry...)

Informatics Linux, NGS Data analysis, Unix shell scripting, Filemaker, Microsoft Office, R, Latex.

Reviewer for: Journal of Eukaryotic Microbiology, Microbial ecology and Protist

Presentations to scientific meetings

2017

22-24.02.2017 36rd Annual Meeting of the German Society for Protozoology/ *Meissen, Germany/* Born in the USA: a molecular phylogeography of *Hyalosphenia papilio*. **Singer D**, Blandenier Q, Duckert C, Fernandez LD, Mitchell EAD, Granath G, Rydin H, Bragazza L, Gałka M, Gennadievna KN, Goia I, Kajukało K,, Lamentowicz M, Pavlovna KN, Payne RJ, Vellak K , Lara E

2016

12-15.09.2016 International Symposium on Testate Amoebae (ISTA8)/ Ilhabela, Brazil/ Born in the USA: a molecular phylogeography of *Hyalosphenia papilio*. **Singer D**, Fernandez LD, Blandenier Q, Mitchell EAD, Lara E.

23-26.02.2016 35rd Annual Meeting of the German Society for Protozoology/ Saignelegier, Switzerland/ Environmental diversity of cryptic species from the *Nebela collaris* complex is strongly correlated with environmental filters. **Singer D**, Kosakyan A, Fernandez LD, Seppey CVW, Mitchell EAD, Lara E

2015

05-10.11.2015 European Congress Of Protistology (ECOP 7)/ Seville, Spain/ Environmental diversity of cryptic species from the *Nebela collaris* complex is strongly correlated with environmental filters. **Singer D**, Kosakyan A, Fernandez L, Seppey CVW, Mitchell EAD, Lara E

2014

14.11.2014 Swiss Systematics Society/ Geneva, Switzerland/ Strong niche separation among species of the *Nebela collaris* complex: a tool for bioindication? **Singer D,** Kosakyan A, Seppey CVW, Mitchell EAD, Lara E

31.10.2014 SwissBOL Conference/ Bern, Switzerland/Earthworm diversity in Switzerland: focus on potential cryptic species. **Singer D**, Luiz L, Al-Dourobi A, Lara E, Le Bayon C

08-12.09.2014 International Symposium on Testate Amoebae (ISTA7)/ Poznan Poland/ Strong niche separation among species of the *Nebela collaris complex*: a tool for bioindication? **Singer D**, Kosakyan A, Seppey CVW, Mitchell EAD, Lara E

15-16.04.2014 Fundamental and applied protistology/ Neuchâtel, Switzerland/ Cryptic testate amoeba species occupy different niches in apeatland: the case of the *Nebela collaris* complex **Singer D,** Kosakyan A, Mulot M, Mitchell EAD, Lara E

List of peer reviewed articles by David Singer (03.05.2017)

In Prep/submitted

1) **Singer D**, Kosakyan A, Seppey CVW, Pillonel A, Fernández LD, Fontaneto D, Mitchell EAD, Lara E. (in prep) Environmental filtering and phylogenetic clustering correlate with the distribution patterns of cryptic microeukaryotic species in peatlands *Molecular ecology*
2) **Singer D**, Fernandez LD, Blandenier Q, Duckert C, Mitchell EAD, Lara E. (in prep) Born in the USA: a molecular phylogeography of *Hyalosphenia papilio* *Global Ecology and Biogeography*
3) **Singer D,** Metz S, Mitchell EAD, Lara E. (in prep) Micro-eukaryotic functional diversity associated with *Sphagnum* mosses in Tropical, Subtropical and Temperate climatic zones *Scientific report*
4) **Singer D,** Mitchell EAD, Lara E. (in prep) The "Sphagnosphere": A unique component of earth's Biosphere to assess microeukaryotes diversity, ecology and biogeography *Journal of Eukaryotic Microbiology*
5) Reczuga MK, Seppey CVW, Szelecz I, Fournier B, **Singer D**, Lara E., Mulot M, Mitchell EAD. (in prep) Temporal patterns of soil micro-eukaryotic diversity beneath decomposing pig cadavers as assessed by high throughput sequencing *ISME journal*
6) Szelecz I, Lösch S, Seppey CVW, Lara E, **Singer D**, Sorge F, Tschui J, Perotti MA, Mitchell EAD. (submitted) Comparative analysis of bones, mites, soil chemistry, nematodes and soil micro-eukaryotic communities of a suspected homicide to estimate a long post-mortem interval *scientific report*
7) Fernández LD, Seppey CVW, **Singer D**, Fournier B, Tatti D, Mitchell EAD, Lara E. (in prep) Elevational diversity patterns in free-living soil unicellular eukaryotes are driven by evolutionary constraints to water and energy availability *ISME journal*

In press/published

2017

8) Seppey C, **Singer D**, Dumack K, Belbahri L, Mitchell EAD, Lara E (submitted) Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling *Soil Biology and Biochemistry*
9) *Geisen S, Mitchell EAD, Wilkinson DM, SAdl S, Bonkowski M, Brown MW, Fiore-Donno AM, Heger TJ, Jassey VEJ, Krashevska V, Lahr DJG, Marcisz K, Mulot M, Payne R,* ***Singer D****, Anderson OR, Charman DJ, Ekelund F, Griffiths BS, Rønn R, Smirnov A, Bass D, Belbahri L, Berney C, Blandenier Q, Chatzinotas A, Clarholm M, Dunthorn M, Feest A, Fernandez-Parra LD, Foissner W, Fournier B, Gentekaki E, Hajek M, Helder J, Jousset A, Koller R, Kumar S, La Terza A, Lamentowicz M, Mazei Y, Santos SS, Seppey CVW, Spiegel FW, Walochnik J, Winding A, Lara E. (submitted)* Soil protistology rebooted: 30 fundamental questions to start with *Soil Biology and Biochemistry* DOI: 10.1016/j.soilbio.2017.04.001
10) Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, **Singer D**, Mayor J, BungeJ, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell EAD, Seppey CVW, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M. Soil Protists in Three Neotropical Rainforests are Hyperdiverse and Dominated by Parasites. *Nature Ecology & Evolution* DOI:10.1038/s41559-017-0091

2016

11) Schiaffino MR, Lara E, Fernández LD, Balagué V, **Singer D,** Seppey CVW, Massana R, Izaguirre I. Microbial eukaryote communities from Patagonian-Antarctic gradient of lakes evidence of a biogeographical pattern *Environmental microbiology* DOI: 10.1111/1462-2920.13566
12) Blandenier Q, Seppey CVW, **Singer D**, Vlimant M, Simon A, Duckert C, Lara E. 2016. *Mycamoeba gemmipara* nov. gen., nov. sp., the First Cultured Member of the Environmental Dermamoebidae Clade LKM74 and its Unusual Life Cycle *Journal of Eukaryotic Microbiology* DOI: 10.1111/jeu.12357
13) **Singer D**, Lara E, Steciow MM, Seppey CVW, Paredes N, Pillonel A, Oszako T, Belbahri L. 2016. High-throughput sequencing reveals diverse oomycete communities in oligotrophic peat bog micro-habitat. *Fungal Ecology*, 23, 42-47 DOI: 10.1016/j.funeco.2016.05.009

2015

14) Lara E, Seppey CVW, Garraza GG, **Singer D**, Quiroga MV, Mataloni G. 2015, Planktonic eukaryote molecular diversity: discrimination of minerotrophic and ombrotrophic peatland pools in Tierra del Fuego (Argentina), *Journal of Plankton Research* 37 (3), 645-655, DOI: 10.1093/plankt/fbv016

15) Seppey CVW, Fournier B, Szelecz I, **Singer D**, Mitchell EAD, Lara E. 2015. Response of forest soil euglyphid testate amoebae (Rhizaria: Cercozoa) to pig cadavers assessed by high-throughput sequencing. *International journal of legal medicine*, 130 (2), 551–562 DOI: 10.1007/s00414-015-1149-7

16) **Singer D**, Kosakyan A, Pillonel A, Mitchell EAD, Lara E. 2015. Eight species in the *Nebela collaris* complex: *Nebela gimlii* (Arcellinida, Hyalospheniidae), a new species described from a Swiss raised bog. *European Journal of Protistology* 51: 79-85. DOI:10.1016/j.ejop.2014.11.004

Awards

23-26.02.2016 35rd Annual Meeting of the German Society for Protozoology, Saignelegier, Switzerland, Second best presentation

12-15.02.2014 33rd Annual Meeting of the German Society for Protozoology, Essen, Germany Second best poster presentation

Activities and Interests

Volunteering: INSIEME (2012-2016), In charge of a summer camp with persons with intellectual disabilities
Music: Guitar (in a band for 10 years)
Photography and Films: Animals, macro, Nature and night

References

Dr. Enrique Lara Senior Scientist of the Soil Biodiversity laboratory of the University of Neuchâtel (Switzerland) Email: enrique.lara@unine.ch

Prof. Edward Mitchell Director of the Soil Biodiversity laboratory of the University of Neuchâtel (Switzerland) Email: edward.mitchell@unine.ch

Dr. Micah Dunthorn Independent Junior Research Group Leader, University of Kaiserslautern, (Germany) Email: dunthorn@rhrk.uni-kl.de

Dr. Diego Fontaneto Permanent researcher, National research council, Verbania (Italy) Email: d.fontaneto@ise.cnr.it

Dr. Thomas Posch Groupleader department of Plant and Microbial Biology, University of Zürich (Switzerland) Email: posch@limnol.uzh.ch

Appendix H

Bibliography

Abad Z.G., Abad J.A., Cacciola S.O., Pane A., Faedda R., Moralejo E., et al. *Phytophthora niederhauserii sp nov., a polyphagous species associated with ornamentals, fruit trees and native plants in 13 countries.* Mycologia, 106(3):431–447, 2014. (Cited on page 28.)

Ackerly D.D. *Community assembly, niche conservatism, and adaptive evolution in changing environments.* International Journal of Plant Sciences, 164(S3):S165–S184, 2003. (Cited on page 6.)

Adl M.S., and Gupta V.V.S.R. *Protists in soil ecology and forest nutrient cycling.* Canadian Journal of Forest Research, 36(7): 1805–1817, 2006. (Cited on pages 44 and 107.)

Adl S.M., Simpson A.G.B., Lane C.E., Lukeš J., Bass D., Bowser S.S., et al. *The revised classification of eukaryotes.* Journal of Eukaryotic Microbiology, 59(5):429–514, 2012. (Cited on pages 36, 44, 102, and 106.)

Adl S.M., Habura A, and Eglit Y. *Amplification primers of ssu rdna for soil protists.* Soil Biology and Biochemistry, 69:328–342, 2014. (Cited on pages 83 and 107.)

Aguilar M., and Lado C. *Ecological niche models reveal the importance of climate variability for the biogeography of protosteloid amoebae.* Isme Journal, 6(8):1506–1514, 2012. (Cited on page 36.)

Aguilar M., Fiore-Donno A.M., Lado C., and Cavalier-Smith T. *Using environmental niche models to test the everything is everywhere hypothesis for badhamia.* The ISME journal, 8:737–745, 2013. (Cited on page 49.)

Ali S.S., Yu Y., Pfosser M., and Wetschnig W. *Inferences of biogeographical histories within subfamily hyacinthoideae using s-diva and bayesian binary mcmc analysis implemented in rasp (reconstruct ancestral state in phylogenies).* Annals of Botany, 109(1):95–107, 2012. (Cited on page 49.)

Allison S.D., and Martiny J.B.H. *Resistance, resilience, and redundancy in microbial communities.* Proceedings of the National Academy of Sciences of the United States of America, 105:11512–11519, 2008. (Cited on page 88.)

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J.H., Zhang Z., Miller W., and Lipman D.J. *Gapped blast and psi-blast: a new generation of protein database search programs.* Nucleic Acids Research, 25(17):3389–3402, 1997. (Cited on pages 84 and 103.)

Amaral-Zettler L.A., McCliment E.A., Ducklow H.W., and Huse S.M. *A method for studying protistan diversity using massively parallel sequencing of v9 hypervariable regions of small-subunit ribosomal rna genes.* Plos One, 4(7), 2009. (Cited on pages 5, 28, 29, 35, 58, 59, 79, 83, 84, 101, and 111.)

Amendt J., Krettek R., and Zehner R. *Forensic entomology.* Naturwissenschaften, 91(2):51–65, 2004. (Cited on page 82.)

Amesbury M.J., Swindles G.T., Bobrov A., Charman D.J., Holden J., Lamentowicz M., et al. *Development of a new pan-european testate amoeba transfer function for reconstructing peatland palaeohydrology.* Quaternary Science Reviews, 152:132–151, 2016. (Cited on page 52.)

Andersen R., Poulin M., Borcard D., Laiho R., Laine J., Vasander H., and Tuittila E.T. *Environmental control and spatial structures in peatland vegetation.* Journal of Vegetation Science, 22 (5):878–890, 2011. (Cited on page 21.)

Anderson M.J., Crist T.O., Chase J.M., Vellend M., Inouye B.D., Freestone A.L., et al. *Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist.* Ecology Letters, 14(1):19–28, 2011. (Cited on pages 68 and 116.)

Asghar M., Hasselquist D., Hansson B., Zehtindjiev P., Westerdahl H., and Bensch S. *Hidden costs of infection: chronic malaria accelerates telomere degradation and senescence in wild birds.* Science, 347(6220):436–438, 2015. (Cited on page 107.)

Azam F., Fenchel T., Field J.G., Gray J.S., Meyerreil L.A., and Thingstad F. *The ecological role of water-column microbes in the sea.* Marine Ecology Progress Series, 10(3):257–263, 1983. (Cited on pages 6, 34, 44, 51, 68, and 92.)

Azovsky A., and Mazei Y. *Do microbes have macroecology? large-scale patterns in the diversity and distribution of marine benthic ciliates.* Global Ecology and Biogeography, 22(2):163–172, 2013. (Cited on page 38.)

Baas-Becking L.G.M. *Geobiologie; of inleiding tot de milieukunde.* 1934. (Cited on pages 6 and 44.)

Bachy C., Dolan J.R., Lopez-Garcia P., Deschamps P., and Moreira D. *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(2):244–255, 2013. (Cited on page 58.)

Bagchi R., Gallery R.E., Gripenberg S., Gurr S.J., Narayan L., Addis C.E., et al. *Pathogens and insect herbivores drive rainforest plant diversity and composition.* Nature, 506(7486):85–88, 2014. (Cited on pages 100 and 107.)

Bahramisharif A., Lamprecht S.C., Spies C.F.J., Botha W.J., Calitz F.J., and McLeod A. *Pythium spp. associated with rooibos seedlings, and their pathogenicity toward rooibos, lupin, and oat.* Plant Disease, 98(2):223–232, 2014. (Cited on page 31.)

Baldauf S.L. *The deep roots of eukaryotes.* Science, 300(5626): 1703–1706, 2003. (Cited on page 4.)

Barta J.R., Ogedengbe J.D., Martin D.S., and Smith T.G. *Phylogenetic position of the adeleorinid coccidia (myzozoa, apicomplexa, coccidia, eucoccidiorida, adeleorina) inferred using 18s rdna sequences.* Journal of Eukaryotic Microbiology, 59(2): 171–180, 2012. (Cited on page 102.)

Barton P.S., Cunningham S.A., Lindenmayer D.B., and Manning A.D. *The role of carrion in maintaining biodiversity and ecological processes in terrestrial ecosystems.* Oecologia, 171(4): 761–772, 2013. (Cited on pages 23, 82, and 88.)

Baselga A. *Partitioning the turnover and nestedness components of beta diversity.* Global Ecology and Biogeography, 19(1):134–143, 2010. (Cited on page 68.)

Baselga A. *The relationship between species replacement, dissimilarity derived from nestedness, and nestedness.* Global Ecology and Biogeography, 21(12):1223–1232, 2012. (Cited on page 80.)

Baselga A., and Orme C.D.L. *betapart: an r package for the study of beta diversity.* Methods in Ecology and Evolution, 3(5):808–812, 2012. (Cited on pages 68 and 80.)

Bass D., and Cavalier-Smith T. *Phylum-specific environmental dna analysis reveals remarkably high global biodiversity of cercozoa (protozoa).* International Journal of Systematic and Evolutionary Microbiology, 54(6):2393–2404, 2004. (Cited on page 102.)

Bass D., Chao E.E.Y., Nikolaev S., Yabuki A., Ishida K., Berney C., et al. *Phylogeny of Novel Naked Filose and Reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea Revised.* Protist, 160(1):75–109, 2009a. (Cited on pages 60 and 115.)

Bass D., Howe A.T., Mylnikov A.P., Vickerman K., Chao E.E., Smallbone J.E., et al. *Phylogeny and classification of cercomonadida (protozoa, cercozoa): Cercomonas, eocercomonas, paracercomonas, and cavernomonas gen. nov.* Protist, 160(4): 483–521, 2009b. (Cited on page 5.)

Basset Y., Cizek L., Cuénoud P., Didham R.K., Guilhaumon F., Missa O., et al. *Arthropod diversity in a tropical forest.* Science, 338(6113):1481–1484, 2012. (Cited on pages 100 and 107.)

Bates S.T., Clemente J.C., Flores G.E., Walters W.A., Parfrey L.W., Knight R., and Fierer N. *Global biogeography of highly diverse protistan communities in soil.* Isme Journal, 7(3):652–659, 2013. (Cited on pages 38, 44, 102, and 107.)

Batzer D.P., and Baldwin A.H. *Wetland Habitats of North America: Ecology and Conservation Concerns.* Stephen Bechtel Fund imprint in ecology and the environment. University of California Press, 2012. ISBN 9780520271647. (Cited on page 29.)

Behnke A., Engel M., Christen R., Nebel M., Klein R.R., and Stoeck T. *Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions.* Environmental Microbiology, 13(2):340–349, 2011. (Cited on pages 58 and 114.)

Beltran E., Valiente-Banuet A., and Verdu M. *Trait divergence and indirect interactions allow facilitation of congeneric species.* Annals of Botany, 110(7):1369–1376, 2012. (Cited on page 21.)

Ben Ali A., De Baere R., De Wachter R., and Van de Peer Y. *Evolutionary relationships among heterokont algae (the autotrophic stramenopiles) based on combined analyses of small and large subunit ribosomal rna.* Protist, 153(2):123–132, 2002. (Cited on page 28.)

Benhamou N., le Floch G., Vallance J., Gerbore J., Grizard D., and Rey P.". *Pythium oligandrum: an example of opportunistic success.* Microbiology, 158(11):2679–2694, 2012. (Cited on page 32.)

Berard A., Dorigo U., Humbert J.F., and Martin-Laurent F. *Microalgae community structure analysis based on 18s rdna amplification from dna extracted directly from soil as a potential soil bioindicator.* AGRONOMY FOR SUSTAINABLE DEVELOPMENT, 25(2):285–291, 2005. (Cited on page 111.)

Berger S.A., and Stamatakis A. *Aligning short reads to reference alignments and trees.* Bioinformatics, 27(15):2068–2075, 2011. (Cited on pages 102 and 104.)

Berger S.A., Krompass D., and Stamatakis A. *Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood.* Systematic biology, pages 391–302, 2011. (Cited on page 102.)

Berney C., Geisen S., Van Wichelen J., Nitsche F., Vanormelingen P., Bonkowski M., and Bass D. *Expansion of the 'reticulosphere': Diversity of novel branching and network-forming amoebae helps to define variosea (amoebozoa).* Protist, 166(2): 271–295, 2015. (Cited on page 5.)

Beyens L., Ledeganck P., Graae B.J., and Nijs I. *Are soil biota buffered against climatic extremes? an experimental test on testate amoebae in arctic tundra (qeqertarsuaq, west greenland).* Polar Biology, 32(3):453–462, 2009. (Cited on page 82.)

Blandenier Q., Seppey C.V.W., Singer D., Vlimant M., Simon A., Duckert C., and Lara E. *Mycamoeba gemmipara nov. gen., nov. sp., the first cultured member of the environmental dermamoebidae clade lkm74 and its unusual life cycle.* Journal of Eukaryotic Microbiology, 2016. (Cited on pages 5 and 51.)

Blaxter M., Mann J., Chapman T., Thomas F., Whitton C., Floyd R., and Abebe E. *Defining operational taxonomic units using dna barcode data.* Philosophical Transactions of the Royal Society B-Biological Sciences, 360(1462):1935–1943, 2005. (Cited on page 5.)

Bobrov A.A., Charman D.J., and Warner B.G. *Ecology of testate amoebae (protozoa : Rhizopoda) on peatlands in western russia with special attention to niche separation in closely related taxa.* Protist, 150(2):125–136, 1999. (Cited on page 82.)

Boenigk J., Pfandl K., Stadler P., and Chatzinotas A. *High diversity of the 'Spumella-like' flagellates: an investigation based on the SSU rRNA gene sequences of isolates from habitats located in six different geographic regions.* Environmental Microbiology, 7(5):685–697, 2005a. (Cited on pages 58 and 64.)

Boenigk J., Pfandl K., Stadler P., and Chatzinotas A. *High diversity of the spumella like flagellates an investigation based on the ssu rrna gene sequences of isolates from habitats located in six different geographic regions.* Environmental Microbiology, 7 (5):685–697, 2005b. (Cited on pages 36 and 116.)

Boenigk J., Jost S., Stoeck T., and Garstecki T. *Differential thermal adaptation of clonal strains of a protist morphospecies originating from different climatic zones.* Environmental Microbiology, 9(3):593–602, 2007. (Cited on page 77.)

Bold H.C. *Morphology of plants.* Harper and Row, 1973. (Cited on page 34.)

Boltovskoy D., Kling S.A., Takahashi K., and Bjørklund K. *World atlas of distribution of recent polycystina (radiolaria).* Palaeontologia Electronica, 13(3):230, 2010. (Cited on page 38.)

Bonkowski M. *Protozoa and plant growth: the microbial loop in soil revisited.* New Phytologist, 162(3):617–631, 2004. (Cited on pages 6, 44, and 110.)

Bonkowski M., and Brandt F. *Do soil protozoa enhance plant growth by hormonal effects?* Soil Biology & Biochemestry, 34 (11):1709–1715, 2002. (Cited on page 110.)

Bonkowski M., and Clarholm M. *Stimulation of plant growth through interactions of bacteria and protozoa: Testing the auxiliary microbial loop hypothesis.* Acta Protozoologica, 51(3): 237–247, 2012. (Cited on pages 6 and 110.)

Booth R.K. *Ecology of testate amoebae (protozoa) in two lake superior coastal wetlands: Implications for paleoecology and environmental monitoring.* Wetlands, 21(4):564–576, 2001. (Cited on page 82.)

Booth R.K. *Testate amoebae as proxies for mean annual waitertable depth in sphagnum-dominated peatlands of north america.* Journal of Quaternary Science, 23(1):43–57, 2008. (Cited on page 82.)

Borcard D., and Legendre P. *All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices.* Ecological Modelling, 153(1-2):51–68, 2002. (Cited on page 80.)

Borcard D., Gillet F., and Legendre P. *Numerical Ecology with R.* Use R! Springer New York, 2011. ISBN 9781441979766. (Cited on page 85.)

Bornemissza G.F. *An analysis of arthropod succession in carrion and the effect of its decomposiion on the soil fauna.* Australian Journal of Zoology, 5(1):1–12, 1957. (Cited on page 82.)

Bouwma A.M., Howard K.J., and Jeanne R.L. *Parasitism in a social wasp: effect of gregarines on foraging behavior, colony productivity, and adult mortality.* Behavioral Ecology and Sociobiology, 59(2):222–233, 2005. (Cited on page 107.)

Bragazza L., Freeman C., Jones T., Rydin H., Limpens J., Fenner N., et al. *Atmospheric nitrogen deposition promotes carbon loss from peat bogs.* Proceedings of the National Academy of Sciences, 103(51):19386–19389, 2006. (Cited on page 53.)

Brate J., Logares R., Berney C., Ree D.K., Klaveness D., Jakobsen K.S., and Shalchian-Tabrizi K. *Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA.* Isme Journal, 4(9): 1144–1153, 2010. (Cited on pages 58 and 64.)

Breiman L. *Random forests.* Machine Learning, 45(1):5–32, 2001. (Cited on pages 60 and 64.)

Bridgham S.D., Pastor J., Janssens J.A., Chapin C., and Malterer T.J. *Multiple limiting gradients in peatlands: A call for a new paradigm.* Wetlands, 16:45–65, 1996. (Cited on pages 22 and 26.)

Bridgham S.D., Updegraff K., and Pastor J. *A comparison of nutrient availability indices along an ombrotrophic-minerotrophic gradient in minnesota wetlands.* Soil Science Society of America Journal, 65(1):259–269, 2001. (Cited on page 22.)

Brown J.H., and Lomolino M.V. *Biogeography.* Sinauer Associates, 1998. ISBN 9780878930739. (Cited on page 70.)

Brown M.W., Spiegel F.W., and Silberman J.D. *Amoeba at attention: Phylogenetic affinity of sappinia pedata.* Journal of Eukaryotic Microbiology, 54(6):511–519, 2007. (Cited on page 97.)

Brown M.W., Spiegel F.W., and Silberman J.D. *Phylogeny of the "forgotten" cellular slime mold, fonticula alba, reveals a key evolutionary branch within opisthokonta.* Molecular Biology and Evolution, 26(12):2699–2709, 2009. (Cited on page 98.)

Bu Z-J., Zheng X-X., Rydin H., Moore T., and Ma J. *Facilitation vs. competition: Does interspecific interaction affect drought responses in sphagnum?* Basic and applied ecology, 14(7):574–584, 2013. (Cited on page 34.)

Bunge J., Woodard L., Böhning D., Foster J.A., Connolly S., and Allen H.K. *Estimating population diversity with catchall.* Bioinformatics, 28(7):1045–1047, 2012. (Cited on page 103.)

Burki F. *The eukaryotic tree of life from a global phylogenomic perspective.* Cold Spring Harbor perspectives in biology, 6(5): a016147, 2014. (Cited on pages 4 and 102.)

Burnham K.P., and Anderson D.R. *Information and likelihood theory: a basis for model selection and inference.* Model selection and multimodel inference: a practical information-theoretic approach, pages 49–97, 2002. (Cited on page 23.)

Buttler A., Warner B.G., Grosvernier P., and Matthey Y. *Vertical patterns of testate amoebae (protozoa: Rhizopoda) and peat forming vegetation on cutover bogs in the jura, switzerland.* New Phytologist, 134:371–382, 1996. (Cited on page 20.)

Buée M., Reich M., Murat C., Morin E., Nilsson R.H., Uroz S., and Martin F. *454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity.* New Phytologist, 184(2):449–456, 2009. (Cited on pages 5 and 28.)

Cairney J.W.G., and Chambers S.M. *Ectomycorrhizal Fungi: Key Genera in Profile.* Springer-Verlag, Berlin Heidelberg, 1st edition, 1999. (Cited on page 114.)

Caron D.A. *Past president's address: protistan biogeography: why all the fuss?* Journal of Eukaryotic Microbiology, 56(2):105–112, 2009. (Cited on page 6.)

Caron D.A., Peele E.R., Lim E.L., and Dennett M.R. *Picoplankton and nanoplankton and their trophic coupling in surface waters of the sargasso sea south of bermuda.* Limnology and Oceanography, 44(2):259–272, 1999. (Cited on page 68.)

Caron D.A., Gast R.J., Countway P.D., and Heidelberg K.B. *Microbial eukaryote diversity and biogeography.* Microbe, 4(2): 71–77, 2009. (Cited on page 78.)

Carter D.O., and Tibbett M. *Taphonomic mycota: Fungi with forensic potential.* Journal of Forensic Sciences, 48(1):168–171, 2003. (Cited on page 82.)

Carter D.O., Yellowlees D., and Tibbett M. *Cadaver decomposition in terrestrial ecosystems.* Naturwissenschaften, 94(1): 12–24, 2007. (Cited on page 82.)

Carter R., and Gregorich E.G. *Soil Sampling and Methods of Analysis, Second Edition.* CRC Press, 2007. ISBN 9781420005271. (Cited on page 21.)

Casamayor E.O., Schafer H., Baneras L., Pedros-Alio C., and Muyzer G. *Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: Comparison by microscopy and denaturing gradient gel electrophoresis.* Applied and Environmental Microbiology, 66 (2):499–508, 2000. (Cited on page 79.)

Cavalier-Smith T., and Chao E.E.Y. *Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista).* Journal of Molecular Evolution, 62(4):388–420, 2006. (Cited on page 58.)

Cavalier-Smith T., Chao E.E.Y., and Oates B. *Molecular phylogeny of amoebozoa and the evolutionary significance of the unikont phalansterium.* European Journal of Protistology, 40 (1):21–48, 2004. (Cited on page 97.)

Chambers J.M., and Hastie T.J. *Statistical models in S.* CRC Press, Inc., 1991. ISBN 0412052911. (Cited on page 36.)

Chao A. *Nonparametric estimation of the number of classes in a population.* Scandinavian Journal of Statistics, 11(4):265–270, 1984. (Cited on page 80.)

Chapman S., Buttler A., Francez A-J., Laggoun-Défarge F., Vasander H., Schloter M., et al. *Exploitation of northern peatlands and biodiversity maintenance: a conflict between economy and ecology.* Frontiers in Ecology and the Environment, 1 (10):525–532, 2003. (Cited on page 53.)

Charman D.J. *Biostratigraphic and palaeoenvironmental applications of testate amoebae.* Quaternary Science Reviews, 20 (16-17):1753–1764, 2001. (Cited on pages 14, 18, and 82.)

Charman D.J., Hendon D., Woodland W.A., Britain) Natural Environment Research Council (Great, and Britain). Quaternary Research Association (Great. *The Identification of Testate Amoebae (Protozoa: Rhizopoda) in Peats.* Technical guide. Quaternary Research Association, 2000. ISBN 9780907780489. (Cited on pages 14, 20, 25, and 52.)

Charriere F., Pavillon N., Colomb T., Depeursinge C., Heger T.J., Mitchell E.A.D., et al. *Living specimen tomography by digital holographic microscopy: morphometry of testate amoeba.* Optics Express, 14(16):7005–7013, 2006. (Cited on page 14.)

Charvet S., Vincent W.F., and Lovejoy C. *Chrysophytes and other protists in High Arctic lakes: molecular gene surveys, pigment signatures and microscopy.* Polar Biology, 35(5):733–748, 2012. (Cited on pages 64 and 77.)

Charvet S., Vincent W.F., and Lovejoy C. *Effects of light and prey availability on arctic freshwater protist communities examined by high-throughput dna and rna sequencing.* Fems Microbiology Ecology, 88(3):550–564, 2014. (Cited on pages 75 and 77.)

Chase J.M., and Leibold M.A. *Ecological niches: linking classical and contemporary approaches.* University of Chicago Press, 2003. ISBN 0226101800. (Cited on pages 5 and 6.)

Chase J.M., and Ryberg W.A. *Connectivity, scale-dependence, and the productivity-diversity relationship.* Ecology Letters, 7 (8):676–683, 2004. (Cited on page 70.)

Chatelain A.P., Meisterfeld R., Roussel-Delif L., and Lara E. *Sphenoderiidae (fam. nov.), a new clade of euglyphid testate amoebae characterized by small, round scales surrounding the aperture.* Protist, 164(6):782–792, 2013. (Cited on pages 5, 83, and 86.)

Cheema Z.A., Farooq M., and Wahid A. *Allelopathy: current trends and future applications.* Springer Science & Business Media, 2012. (Cited on page 51.)

Chen J., Bittinger K., Charlson E.S., Hoffmann C., Lewis J., Wu G.D., et al. *Associating microbiome composition with environmental covariates using generalized UniFrac distances.* Bioinformatics, 28(16):2106–2113, 2012. (Cited on page 60.)

Chiu G., Lockhart R., and Routledge R. *Bent-cable regression theory and applications.* Journal of the American Statistical Association, 101(474):542–553, 2006. (Cited on page 59.)

Choi Y-J., Klosterman S.J., Kummer V., Voglmayr H., Shin H-D., and Thines M. *Multi-locus tree and species tree approaches toward resolving a complex clade of downy mildews (straminipila, oomycota), including pathogens of beet and spinach.* Molecular Phylogenetics and Evolution, 86:24–34, 2015. (Cited on page 32.)

Chomczynski P., and Sacchi N. *Single-step method of rna isolation by acid guanidinium thiocyanate–phenol–chloroform extraction.* Analytical Biochemistry, 162(1):156–159, 1987. (Cited on pages 15, 45, and 93.)

Clarholm M. *Possible roles for roots, bacteria, protozoa and fungi in supplying nitrogen to plants.* Blackwell Scientific Publication, Oxford, 1985. (Cited on pages 6, 34, 44, and 110.)

Clymo R.S. *The growth of sphagnum: some effects of the environment.* Journal of Ecology, 61:849–869, 1973. (Cited on page 26.)

Clymo R.S., and Hayward P.M. *The ecology of Sphagnum*, pages 229–289. Springer, 1982. (Cited on page 9.)

Connell J.H. *On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees, dynamics of population*. Ed. PJ Den Boer and GR Gradwell. Wageningen: Pudoc, 1970. (Cited on pages 100 and 107.)

Corsaro D., and Venditti D. *Molecular phylogenetics evidence for a novel lineage of amoebae within discosea (amoebozoa: Lobosa)*. Acta Protozoologica, 52(4):309–316, 2013. (Cited on pages 92, 95, 96, and 97.)

Corsaro D., Walochnik J., Venditti D., Mueller K.D., Hauroeder B., and Michel R. *Rediscovery of nucleophaga amoebae, a novel member of the rozellomycota*. Parasitology reasearch, 113(12): 4491–4498, 2014a. (Cited on page 110.)

Corsaro D., Walochnik J., Venditti D., Steinmann J., Mueller K-D., and Michel R. *Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota*. Parasitology Research, 113(5):1909–1918, 2014b. (Cited on page 64.)

Cottenie K. *Integrating environmental and spatial processes in ecological community dynamics*. Ecology Letters, 8(11):1175–1182, 2005. (Cited on page 77.)

Couradeau E., Benzerara K., Moreira D., Gerard E., Kazmierczak J., Tavera R., and Lopez-Garcia P. *Prokaryotic and Eukaryotic Community Structure in Field and Cultured Microbialites from the Alkaline Lake Alchichica (Mexico)*. Plos One, 6(12), 2011. (Cited on page 65.)

Crane K.W., and Grover J.P. *Coexistence of mixotrophs, autotrophs, and heterotrophs in planktonic microbial communities*. Journal of Theoretical Biology, 262(3):517–527, 2010. (Cited on pages 6 and 34.)

Crawley M.J. *Mixed-effects models*. The R Book, Second Edition, pages 681–714, 2012. (Cited on page 23.)

Cray J.A., Bell A.N.W., Bhaganna P., Mswaka A.Y., Timson D.J., and Hallsworth J.E. *The biology of habitat dominance; can microbes behave as weeds?* Microbial Biotechnology, 6(5):453–492, 2013. (Cited on page 65.)

Crow G.E. *Species diversity in aquatic angiosperms: latitudinal patterns*. Aquatic Botany, 44(2-3):229–258, 1993. (Cited on page 75.)

Currie D.J., Mittelbach G.G., Cornell H.V., Field R., Guegan J.F., Hawkins B.A., et al. *Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness*. Ecology Letters, 7(12):1121–1134, 2004. (Cited on pages 35 and 70.)

Daniel G., Amblard C., Bourdier G., André-Jean F., and Mitchell E.A.D. *Le régime alimentaire des thécamoebiens (protista, sarcodina)*. L'Année Biologique, 39(2):57–68, 2000. (Cited on pages 82 and 88.)

Darling K.F., Kucera M., and Wade C.M. *Global molecular phylogeography reveals persistent arctic circumpolar isolation in a marine planktonic protist*. Proceedings of the National Academy of Sciences, 104(12):5002–5007, 2007. (Cited on page 49.)

Darwin C. *The origin of species*. Lulu.com, 1872. ISBN 130473787X. (Cited on page 3.)

Davis A.K., Yabsley M.J., Keel M.K., and Maerz J.C. *Discovery of a novel alveolate pathogen affecting southern leopard frogs in Georgia: Description of the disease and host effects*. Ecohealth, 4(3):310–317, 2007. (Cited on page 64.)

Dawson S.C., and Hagen K.D. *Mapping the protistan'rare biosphere'*. Journal of biology, 8(12):1, 2009. (Cited on page 68.)

Dawson S.C., and Pace N.R. *Novel kingdom-level eukaryotic diversity in anoxic environments*. Proceedings of the National Academy of Sciences, 99(12):8324–8329, 2002. (Cited on page 106.)

de Caceres M., and Legendre P. *Associations between species and groups of sites: indices and statistical inference*. Ecology, 90 (12):3566–3574, 2009. (Cited on page 36.)

de la Bastide P.Y., Leung W.L., and Hintz W.E. *Species composition of the genus saprolegnia in fin fish aquaculture environments, as determined by nucleotide sequence analysis of the nuclear rdna its regions*. Fungal Biology, 119(1):27–43, 2015. (Cited on page 28.)

De Vargas C., Audic S., Henry N., Decelle J., Mahé F., Logares R., et al. *Eukaryotic plankton diversity in the sunlit ocean*. Science, 348(6237):1261605, 2015. (Cited on pages 4, 5, 20, 28, 29, 31, 34, 35, 51, 52, 79, 94, 100, 101, 104, 110, 112, and 116.)

Debroas D., Hugoni M., and Domaizon I. *Evidence for an active rare biosphere within freshwater protists community*. Molecular Ecology, 24(6):1236–1247, 2015. (Cited on pages 69 and 75.)

Dedysh S.N., Pankratov T.A., Belova S.E., Kulichevskaya I.S., and Liesack W. *Phylogenetic analysis and in situ identification of bacteria community composition in an acidic sphagnum peat bog*. Applied and Environmental Microbiology, 72(3):2110–2117, 2006. (Cited on pages 8, 20, 26, and 35.)

del Campo J., Sieracki M.E., Molestina R., Keeling P., Massana R., and Ruiz-Trillo I. *The others: our biased perspective of eukaryotic genomes*. Trends in ecology & evolution, 29(5):252–259, 2014. (Cited on page 102.)

del Campo J., Mallo D., Massana R., de Vargas C., Richards T.A., and Ruiz-Trillo I. *Diversity and distribution of unicellular opisthokonts along the european coast analysed using high-throughput sequencing*. Environmental Microbiology, 17 (9):3195–3207, 2015. (Cited on pages 5 and 92.)

Delarze R., Gonseth Y., Eggenberg S., and Vust M. *Guide des milieux naturels de Suisse: Écologie, menaces, espèces caractéristiques*. Bussigny, Switzerland, 3e edition, 2015. (Cited on page 111.)

Desportes I., and Schrével J. *Treatise on Zoology-Anatomy, Taxonomy, Biology. The Gregarines (2 vols) The Early Branching Apicomplexa*. Brill, 2013. (Cited on pages 104, 106, and 107.)

Di Filippo M.M., Santoro M., Lovreglio P., Monno R., Capolongo C., Calia C., et al. *Isolation and molecular characterization of free-living amoebae from different water sources in italy*. International Journal of Environmental Research and Public Health, 12(4):3417–3427, 2015. (Cited on page 92.)

Diamond J.M., and Cody M.L. *Ecology and evolution of communities / Martin L. Cody and Jared M. Diamond, editors*. Belknap Press of Harvard University Press, Cambridge, Mass, 1975. ISBN 0674224442. (Cited on page 22.)

Diez B., Pedros-Alio C., and Massana R. *Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rrna gene cloning and sequencing*. Applied and Environmental Microbiology, 67(7):2932–2941, 2001. (Cited on pages 5, 51, 77, and 79.)

Dise N.B. *Peatland response to global change*. Science, 326(5954): 810–811, 2009. (Cited on page 54.)

Dixon P. *Vegan, a package of r functions for community ecology*. Journal of Vegetation Science, 14(6):927–930, 2003. (Cited on page 31.)

Doolittle D.F., Li W.K.W., and Wood A.M. *Wintertime abundance of picoplankton in the atlantic sector of the southern ocean*. Nova Hedwigia, pages 147–160, 2008. (Cited on page 75.)

Dorrepaal E., Toet S., van Logtestijn R.S.P., Swart E., van de Weg M.J., Callaghan T.V., and Aerts R. *Carbon respiration from subsurface peat accelerated by climate warming in the subarctic*. Nature, 460(7255):616–U79, 2009. (Cited on pages 8 and 35.)

Drummond A.J., Suchard M.A., Xie D., and Rambaut A. *Bayesian phylogenetics with beauti and the beast 1.7*. Molecular biology and evolution, 29(8):1969–1973, 2012. (Cited on pages 45 and 46.)

Duffy M.A., James T.Y., and Longworth A. *Ecology, virulence, and phylogeny of blastulidium paedophthorum, a widespread brood parasite of daphnia spp*. Applied and environmental microbiology, 81(16):5486–5496, 2015. (Cited on page 28.)

Dufrene M., and Legendre P. *Species assemblages and indicator species: The need for a flexible asymmetrical approach*. Ecological Monographs, 67(3):345–366, 1997. (Cited on page 60.)

Dumack K., Baumann C., and Bonkowski M. *A bowl with marbles: Revision of the thecate amoeba genus Lecythiu (chlamydophryidae, tectofilosida, cercozoa, rhizaria) includinga description of four new species and an identification key*. Protist, 167(5):440–459, 2016a. (Cited on page 110.)

Dumack K., Müller M.E.H., and Bonkowski M. *Description of lecythium terrestris sp. nov. (chlamydophryidae, cercozoa), a soil dwelling protist feeding on fungi and algae*. Protist, 167 (2):93–105, 2016b. (Cited on pages 6, 34, 107, 110, and 115.)

Dunthorn M., Klier J., Bunge J., and Stoeck T. *Comparing the hyper-variable v4 and v9 regions of the small subunit rdna for assessment of ciliate environmental diversity.* Journal of Eukaryotic Microbiology, 59(2):185–187, 2012. (Cited on page 5.)

Dunthorn M., Otto J., Berger S. A., Stamatakis A., Mahe F., Romac S., et al. *Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context.* Molecular Biology and Evolution, 31(4):993–1009, 2014. (Cited on pages 86 and 104.)

Dupont A., Griffiths R.I., Bell T., and Bass D. *Differences in soil micro-eukaryotic communities over soil ph gradients are strongly driven by parasites and saprotrophs.* Environmental microbiology, 18:2010–2024, 2016. (Cited on pages 102 and 107.)

Duran A., Gryzenhout M., Drenth A, Slippers B., Ahumada E., Wingfield B.D., and Wingfield M.J. *Aflp analysis reveals a clonal population of phytophthora pinifolia in chile.* Fungal Biology, 114(9):746–752, 2010. (Cited on page 28.)

Dykstra M.J., and Keller H.W. Class mycetozoa de bary, 1859. In Lee JJ, Leedale GF, and Bradbury P, editors, *An Illustrated Guide to the Protozoa*, page 952–981. Society of Protozoologists, Lawrence, Kansas, 2000. (Cited on page 97.)

Edgar R.C. *Muscle: a multiple sequence alignment method with reduced time and space complexity.* Bmc Bioinformatics, 5:1–19, 2004. (Cited on page 86.)

Edvardsen B., Egge E.S., and Vaulot D. *Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding a review.* Perspectives in Phycology, 3:77–91, 2016. (Cited on pages 102 and 106.)

Egge E.S., Eikrem W., and Edvardsen B. *Deep-branching novel lineages and high diversity of haptophytes in the skagerrak (norway) uncovered by 454 pyrosequencing.* Journal of Eukaryotic Microbiology, 62(1):121–140, 2015. (Cited on page 102.)

Ekelund L., and Ronn R. *Notes on protozoa in agricultural soil with emphasis on heterotrophic flagellates and naked amoebae and their ecology.* FEMS Microbiology review, 15(4):321–353, 1995. (Cited on page 110.)

Elbert W., Weber B., Burrows S., Steinkamp J., Büdel B., Andreae M.O., and Pöschl U. *Contribution of cryptogamic covers to the global cycles of carbon and nitrogen.* Nature Geoscience, 5(7):459–462, 2012. (Cited on pages 106 and 111.)

Ellis V. A, Collins M.D., Medeiros M.C.I., Sari E.H.R., Coffey E.D., Dickerson R.C., et al. *Local host specialization, host-switching, and dispersal shape the regional distributions of avian haemosporidian parasites.* Proceedings of the National Academy of Sciences, 112(36):11294–11299, 2015. (Cited on page 107.)

Elton C.S. *Animal ecology, by Charles Elton; with an introduction by Julian S. Huxley.* Macmillan Co., New York, 1927. doi: 10.5962/bhl.title.7435. (Cited on pages 5 and 22.)

Emerson B.C., and Gillespie R.G. *Phylogenetic analysis of community assembly and structure over space and time.* Trends in Ecology and Evolution, 23(11):619–630, 2008. (Cited on page 21.)

Epstein S., and Lopez-Garcia P. *"missing" protists: a molecular prospective.* Biodiversity and Conservation, 17(2):261–276, 2008. (Cited on page 68.)

Erwin T.L. *Tropical forests: their richness in coleoptera and other arthropod species.* Coleopterists Bulletin, 36(1):74–75, 1982. (Cited on page 100.)

Evans K.M., Chepurnov V.A., Sluiman H.J., Thomas S.J., Spears B.M., and Mann D.G. *Highly differentiated populations of the freshwater diatom sellaphora capitata suggest limited dispersal and opportunities for allopatric speciation.* Protist, 160(3):386–396, 2009. (Cited on pages 7 and 53.)

Falkowski P.G. *The ocean's invisible forest.* Scientific American, 287(2):38–45, 2002. (Cited on pages 6 and 44.)

Farrell B.D. *"inordinate fondness" explained why are there so many beetles?* Science, 281(5376):555–559, 1998. (Cited on page 108.)

Fenchel T. *Cosmopolitan microbes and their 'cryptic'species.* Aquatic Microbial Ecology, 41(1):49–54, 2005. (Cited on page 6.)

Fernández L.D., Fournier B., Rivera R., Lara E., Mitchell E.A.D., and Hernández C.E. *Water-energy balance, past ecological perturbations and evolutionary constraints shape the latitudinal diversity gradient of soil testate amoebae in south-western south america.* Global Ecology and Biogeography, 2016. (Cited on page 75.)

Field K.J., Rimington W.R., Bidartondo M.I., Allinson K.E., Beerling D.J., Cameron D.D., et al. *First evidence of mutualism between ancient plant lineages (haplomitriopsida liverworts) and mucoromycotina fungi and its response to simulated palaeozoic changes in atmospheric co2.* New Phytologist, 205(2):743–756, 2015. (Cited on page 40.)

Fierer N., Lauber C.L., Ramirez K.S., Zaneveld J., Bradford M.A., and Knight R. *Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients.* Isme Journal, 6(5):1007–1017, 2012. (Cited on page 88.)

Filker S., Gimmler A., Dunthorn M., Mahé F., and Stoeck T. *Deep sequencing uncovers protistan plankton diversity in the portuguese ria formosa solar saltern ponds.* Extremophiles, 19(2):283–295, 2015. (Cited on page 107.)

Filker S., Sommaruga R., Vila I., and Stoeck T. *Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns.* Molecular Ecology, 25(10):2286–2301, 2016. (Cited on page 69.)

Finlay B.J. *The global diversity of protozoa and other small species.* International journal for parasitology, 28(1):29–48, 1998. (Cited on pages 6 and 7.)

Finlay B.J. *Global dispersal of free-living microbial eukaryote species.* Science, 296(5570):1061–1063, 2002. (Cited on pages 6, 44, 53, and 69.)

Finlay B.J. *Protist taxonomy: an ecological perspective.* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 359(1444):599–610, 2004. (Cited on page 4.)

Fisher R.A., Corbet A.S., and Williams C.B. *The relation between the number of species and the number of individuals in a random sample of an animal population.* Journal of Animal Ecology, 12:42–58, 1943. (Cited on page 69.)

Fiz-Palacios O., Leander B.S., and Heger T.J. *Old lineages in a new ecosystem: diversification of arcellinid amoebae (amoebozoa) and peatland mosses.* PloS one, 9(4):e95238, 2014. (Cited on page 49.)

Foissner W. *Soil protozoa: fundamental problems, ecological significance, adaptations in ciliates and testaceans, bioindicators, and guide to the literature.* Progress in protistology, 2:69–212, 1987. (Cited on page 82.)

Foissner W. *Class colpodea (ciliophora).* Protozoenfauna. Bd, 4 (1), 1993. (Cited on page 107.)

Foissner W. *Protist diversity: Estimates of the near-imponderable.* Protist, 150(4):363–368, 1999a. (Cited on pages 7 and 68.)

Foissner W. *Soil protozoa as bioindicators: pros and cons, methods, diversity, representative examples.* Agriculture Ecosystems & Environment, 74(1-3):95–112, 1999b. (Cited on pages 82 and 110.)

Foissner W. *Biogeography and dispersal of micro-organisms: a review emphasizing protists.* Acta Protozoologica, 45(2):111–136, 2006. (Cited on pages 7, 44, and 53.)

Folmer O., Black M., Hoeh W., Lutz R., and Vrijenhoek R. *Dna primers for amplification of mitochondrial cytochrome c oxidase subunit i from diverse metazoan invertebrates.* Molecular marine biology and biotechnology, 3(5):294–9, 1994. (Cited on pages 15, 21, and 45.)

Fontaneto D. *Biogeography of microscopic organisms: is everything small everywhere?* Cambridge University Press, 2011. ISBN 1139496581. (Cited on page 7.)

Fontaneto D., Herniou E.A., Boschetti C., Caprioli M., Melone G., Ricci C., and Barraclough T.G. *Independently evolving species in asexual bdelloid rotifers.* PLoS Biol, 5(4):e87, 2007. (Cited on pages 5 and 45.)

Fontaneto D., Flot J.F., and Tang C.Q. *Guidelines for dna taxonomy, with a focus on the meiofauna.* Marine Biodiversity, 45 (3):433–451, 2015. (Cited on pages 5 and 45.)

Forster D., Dunthorn M., Mahé F., Dolan J.R., Audic S., Bass D., et al. *Benthic protists: the under-charted majority.* FEMS microbiology ecology, page fiw120, 2016a. (Cited on page 107.)

Forster D., Dunthorn M., Stoeck T., and Mahé F. *Comparison of three clustering approaches for detecting novel environmental microbial diversity.* PeerJ, 4:e1692, 2016b. (Cited on page 107.)

Fox J., and Weisberg S. *An R companion to applied regression.* Sage, 2010. (Cited on page 112.)

Freckleton R.P., and Lewis O.T. *Pathogens, density dependence and the coexistence of tropical trees.* Proceedings of the Royal Society of London B: Biological Sciences, 273(1604):2909–2916, 2006. (Cited on page 107.)

Freeman K.R., Pescador M.Y., Reed S.C., Costello E.K., Robeson M.S., and Schmidt S.K. *Soil co2 flux and photoautotrophic community composition in high-elevation,'barren'soil.* Environmental microbiology, 11(3):674–686, 2009. (Cited on page 111.)

Frey B., Buehler L., Schmutz S., Zumsteg A., and Furrer G. *Molecular characterization of phototrophic microorganisms in the forefield of a receding glacier in the swiss alps.* Environmental reasearch letters, 8(1), 2013. (Cited on page 111.)

Fuhrman J.A. *Microbial community structure and its functional implications.* Nature, 459(7244):193–199, 2009. (Cited on page 79.)

Fujisawa T., and Barraclough T.G. *Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets.* Systematic Biology, 62(5):707–724, 2013. (Cited on pages 5 and 45.)

Fuller M.S., and Jaworski A. *Zoosporic Fungi in Teaching & Research.* Palfrey contributions in botany. Southeastern Publishing Corporation, 1987. ISBN 9780935460025. (Cited on page 28.)

Futuyma D.J., and Moreno G. *The evolution of ecological specialization.* Annual Review of Ecology and Systematics, 19:207–233, 1988. (Cited on page 20.)

Galand P.E., Casamayor E.O., Kirchman D.L., and Lovejoy C. *Ecology of the rare microbial biosphere of the arctic ocean.* Proceedings of the National Academy of Sciences of the United States of America, 106(52):22427–22432, 2009. (Cited on page 69.)

Gaston K.J. *Global patterns in biodiversity.* Nature, 405(6783): 220–227, 2000. (Cited on page 75.)

Gaston K.J., and Blackburn T.M. *Pattern and process in macroecology.* 2000. (Cited on page 70.)

Geisen S. *Thorough high-throughput sequencing analyses unravels huge diversities of soil parasitic protists.* Environmental Microbiology, 18(6):1669–1672, 2016. (Cited on page 34.)

Geisen S., Laros I., Vizcaino A., Bonkowski M., and De Groot G.A. *Not all are free-living: high-throughput dna metabarcoding reveals a diverse community of protists parasitizing soil metazoa.* Molecular Ecology, 24(17):4556–4569, 2015a. (Cited on page 114.)

Geisen S., Tveit A.T., Clark I.M., Richter A., Svenning M.M., Bonkowski M., and Urich T. *Metatranscriptomic census of active protists in soils.* Isme Journal, 9(10):2178–2190, 2015b. (Cited on pages 5, 34, 92, 102, and 107.)

Geisen S., Koller R., Huenninghaus M., Dumack K., Urich T., and Bonkowski M. *The soil food web revisited: Diverse and widespread mycophagous soil protists.* Saoil biology & biochemestry, 94:10–18, 2016. (Cited on pages 6 and 110.)

Gentry A.H. *Four Neotropical Rainforests.* Yale University Press, 1993. ISBN 9780300054484. (Cited on page 107.)

Gibbons S.M., Caporaso J.G., Pirrung M., Field D., Knight R., and Gilbert J.A. *Evidence for a persistent microbial seed bank throughout the global ocean.* Proceedings of the National Academy of Sciences, 110(12):4651–4655, 2013. (Cited on page 69.)

Gilbert D. *Les communautés microbiennes à la surface des tourbières à sphaignes: structure, fonctionnement et impact des apports de fertilisants.* PhD thesis, Université de Clermont-Ferrand 2, Clermont-Ferrand, France, 1998. (Cited on page 14.)

Gilbert D., Amblard C., Bourdier G., and Francez A.J. *Short-term effect of nitrogen enrichment on the microbial communities of a peatland.* Hydrobiologia, 374:111–119, 1998a. (Cited on page 88.)

Gilbert D., Amblard C., Bourdier G., and Francez A.J. *The microbial loop at the surface of a peatland: Structure, function, and impact of nutrient input.* Microbial Ecology, 35(1):83–93, 1998b. (Cited on pages 14, 82, and 88.)

Gilbert D., Mitchell E.A.D., Amblard C., Bourdier G., and Francez A.J. *Population dynamics and food preferences of the testate amoeba nebela tincta major-bohemica-collaris complex (protozoa) in a sphagnum peatland.* Acta Protozoologica, 42:99–104, 2003. (Cited on page 20.)

Gilbert D., Mitchell E.A.D., Martini I.P., Martínez-Cortizas A., and Chesworth W. Microbial diversity in sphagnum peatlands. In *Peatlands: Evolution and Records of Environmental and Climate Changes*, pages 287–318. Elsevier Amsterdam, 2006. (Cited on pages 9, 14, 20, and 35.)

Giner C.R., Forn I., Romac S., Logares R., de Vargas C., and Massana R. *Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes.* Applied and environmental microbiology, pages AEM. 00560–16, 2016. (Cited on pages 34 and 116.)

Giovannoni S.J., Britschgi T.B., Moyer C.L., and Field K.G. *Genetic diversity in sargasso sea bacterioplankton.* Nature, 345 (6270):60–63, 1990. (Cited on page 5.)

Glaser K., Kuppardt A., Boenigk J., Harms H., Fetzer I., and Chatzinotas A. *The influence of environmental factors on protistan microorganisms in grassland soils along a land-use gradient.* Science of the total environment, 537:33–42, 2015. (Cited on page 114.)

Glockling S.L., and Serpell L.C. *A new species of aplanosporic haptoglossa, h. beakesii, with vesiculate spore release.* Botany, 88(1):93–101, 2010. (Cited on page 32.)

Gomaa F., Mitchell E.A.D., and Lara E. *Amphitremida (poche, 1913) is a new major, ubiquitous labyrinthulomycete clade.* Plos One, 8(1), 2013. (Cited on pages 15 and 45.)

Gomaa F., Kosakyan A., Heger T.J., Corsaro D., Mitchell E.A.D., and Lara E. *One alga to rule them all: unrelated mixotrophic testate amoebae (amoebozoa, rhizaria and stramenopiles) share the same symbiont (trebouxiophyceae).* Protist, 165(2): 161–176, 2014. (Cited on page 45.)

Gomez F. *A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (dinoflagellata, alveolata).* Systematics and Biodiversity, 10(3):267–275, 2012. (Cited on page 36.)

Gómez F., López-García P., Nowaczyk A., and Moreira D. *The crustacean parasites ellobiopsis caullery, 1910 and thalassomyces niezabitowski, 1913 form a monophyletic divergent clade within the alveolata.* Systematic parasitology, 74(1):65–74, 2009. (Cited on page 102.)

Gonzalez M.A., Coleman A.W., Gomez P.I., and Montoya R. *Phylogenetic relationship among various strains of dunaliella (chlorophyceae) based on nuclear its rdna sequences.* Journal of Phycology, 37(4):604–611, 2001. (Cited on page 20.)

Gonzalez Garraza G., Mataloni G., Iturraspe R., Lombardo R., Camargo S., and Quiroga M.V. *The limnological character of bog pools in relation to meteorological and hydrological features.* Mires and Peat, 10, 2012. (Cited on page 58.)

González Garraza G. *Fracciones de tamano del fitoplancton de las lagunas de la turbera de Rancho Hambre (Tierra del Fuego): caracterización y relación con los parametros bióticos y abióticos.* PhD thesis, Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires, Argentina, 2012. (Cited on pages 59 and 61.)

Goodfellow M., and Stackebrandt E. *Nucleic acid techniques in bacterial systematics*, volume 5. J. Wiley, 1991. (Cited on page 104.)

Gorham E. *The development of peat lands.* Quarterly Review of Biology, pages 145–166, 1957. (Cited on page 34.)

Gorham E. *Northern peatlands: role in the carbon cycle and probable responses to climatic warming.* Ecological applications, 1 (2):182–195, 1991. (Cited on pages 8 and 35.)

Gorham E. *The biogeochemistry of northern peatlands and its possible response to global warming.* Biotic Feedback in the Global Climatic System, GM Woodwell and FT Mackenzie (eds.), Oxford University Press, New York, pages 169–187, 1995. (Cited on page 8.)

Gotelli N.J. *Null model analysis of species co-occurrence patterns.* Ecology, 81(9):2606–2621, 2000. (Cited on page 22.)

Gotelli N.J., and Ellison A.M. *Ecosimr 1.00.* See http://www.uvm. edu/~ ngotelli/EcoSim/EcoSim. html, 2013. (Cited on pages 22 and 23.)

Gotelli N.J., and Graves G.R. *Null models in ecology.* Smithsonian Institution Press, 1996. ISBN 9781560986454. (Cited on page 22.)

Gottsberger G., and Nannenga.N.E . *New species of didymium from brazil.* Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen Series C-Biological and Medical Sciences, 74(3):264–+, 1971. (Cited on page 97.)

Grabda Jadwiga. *Marine fish parasitology: An outline.* VCH Verlagsgesellschaft mbH, 1991. ISBN 3527268987. (Cited on page 40.)

Green J.L., Holmes A.J., Westoby M., Oliver I., Briscoe D., Dangerfield M., et al. *Spatial scaling of microbial eukaryote diversity.* Nature, 432(7018):747–750, 2004. (Cited on page 78.)

Grinnell J. *The niche-relationships of the california thrasher.* The Auk, 34(4):427–433, 1917. (Cited on pages 5 and 22.)

Grossmann L., Jensen M., Heider D., Jost S., Glücksman E., Hartikainen H., et al. *Protistan community analysis: key findings of a large-scale molecular sampling.* The ISME journal, 10: 269–279, 2016. (Cited on pages 5, 51, and 107.)

Guillou L., Viprey M., Chambouvet A., Welsh R.M., Kirkham A.R., Massana R., et al. *Widespread occurrence and genetic diversity of marine parasitoids belonging to syndiniales (alveolata).* Environmental Microbiology, 10(12):3349–3365, 2008. (Cited on pages 5 and 92.)

Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., et al. *The protist ribosomal reference database (pr2): a catalog of unicellular eukaryote small sub-unit rrna sequences with curated taxonomy.* Nucleic Acids Research, 41(D1):D597–D604, 2013. (Cited on pages 29, 35, 59, 79, 102, 104, 105, 111, 112, and 144.)

Hadziavdic K., Lekang K., Lanzen A., Jonassen I., Thompson E.M., and Troedsson C. *Characterization of the 18s rrna gene for designing universal eukaryote specific primers.* PLoS One, 9(2): e87624, 2014. (Cited on page 78.)

Hagen F., Khayhan K., Theelen B., Kolecka A., Polacheck I., Sionov E., et al. *Recognition of seven species in the cryptococcus gattii/cryptococcus neoformans species complex.* Fungal Genetics and Biology, 78:16–48, 2015. (Cited on page 5.)

Hall T.A. Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. In *Nucleic acids symposium series*, volume 41, pages 95–98, 1999. (Cited on pages 15 and 45.)

Hamels I., Mussche H., Sabbe K., Muylaert K., and Vyverman W. *Evidence for constant and highly specific active food selection by benthic ciliates in mixed diatoms assemblages.* Limonology and oceonography, 49(1):58–68, 2004. (Cited on page 116.)

Harris D., Horwath W.R., and van Kessel C. *Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis.* Soil Science Society of America Journal, 65(6):1853–1856, 2001. (Cited on page 84.)

Hawkins B.A., Field R., Cornell H.V., Currie D.J., Guegan J.F., Kaufman D.M., et al. *Energy, water, and broad-scale geographic patterns of species richness.* Ecology, 84(12):3105–3117, 2003. (Cited on page 70.)

Hawksworth D.L., and Wiltshire P.E.J. *Forensic mycology: the use of fungi in criminal investigations.* Forensic Science International, 206(1-3):1–11, 2011. (Cited on page 82.)

Heal O.W. *Observations on the seasonal and spatial distribution of testacea (protozoa: Rhizopoda) in sphagnum.* Journal of Animal Ecology, 33(3):395–412, 1964. (Cited on page 83.)

Hebert P.D.N., Cywinska A., and Ball S.L. *Biological identifications through dna barcodes.* Proceedings of the Royal Society of London B: Biological Sciences, 270(1512):313–321, 2003. (Cited on page 3.)

Hedwall P., Brunet J., and Rydin H. *Peatland plant communities under global change: negative feedback loops counteract shifts in species composition.* Ecology, 98(1):150–161, 2017. (Cited on page 54.)

Heesch S., Pažoutová M., Moniz M.B.J., and Rindi F. *Prasiolales (trebouxiophyceae, chlorophyta) of the svalbard archipelago: diversity, biogeography and description of the new genera prasionella and prasionema.* European Journal of Phycology, 51 (2):171–187, 2016. (Cited on page 51.)

Heger T.J., Mitchell E.A.D., Todorov M., Golemansky V., Lara E., Leander B.S., and Pawlowski J. *Molecular phylogeny of euglyphid testate amoebae (cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent.* Molecular Phylogenetics and Evolution, 55(1):113–122, 2010. (Cited on pages 44 and 87.)

Heger T.J., Lara E., and Mitchell E.A.D. Arcellinida testate amoebae (amoebozoa: Arcellinida): model of organisms for assessing microbial biogeography. In Fontaneto Diego, editor, *Biogeography of Microscopic Organisms*, pages 111–129. Cambridge University Press, 2011a. ISBN 9780511974878. Cambridge Books Online. (Cited on pages 15, 17, 22, and 52.)

Heger T.J., Pawlowski J., Lara E., Leander B.S., Todorov M., Golemansky V., and Mitchell E.A.D. *Comparing potential coi and ssu rdna barcodes for assessing the diversity and phylogenetic relationships of cyphoderiid testate amoebae (rhizaria: Euglyphida).* Protist, 162(1):131–141, 2011b. (Cited on page 83.)

Heger T.J., Mitchell E.A.D., and Leander B.S. *Holarctic phylogeography of the testate amoeba hyalosphenia papilio (amoebozoa: Arcellinida) reveals extensive genetic diversity explained more by environment than dispersal limitation.* Molecular Ecology, 22(20):5172–5184, 2013. (Cited on pages 5, 7, 14, 18, 44, 45, 46, 48, 49, 53, and 141.)

Heger T.J., Derungs N., Theurillat J-P., and Mitchell E.A.D. *Testate amoebae like it hot: Species richness decreases along a subalpine-alpine altitudinal gradient in both natural calluna vulgaris litter and transplanted minuartia sedoides cushions.* Microbial ecology, 71(3):725–734, 2016. (Cited on page 35.)

Held A.A. *Rozella and Rozellopsis: Naked endoparasitic fungi which dress-up as their hosts.* Botanical Review, 47(4):451–515, 1981. (Cited on page 64.)

Henssge C., Althaus L., Bolt J., Freislederer A., Haffner H.T., Henssge C.A., et al. *Experiences with a compound method for estimating the time since death. i. rectal temperature nomogram for time since death.* International Journal of Legal Medicine, 113(6):303–319, 2000. (Cited on page 82.)

Hess S., and Melkonian M. *The mystery of clade x: Orciraptor gen. nov and Viridiraptor gen. nov are highly specialised, algivorous amoeboflagellates (glissomonadida, cercozoa).* Protist, 164(5):706–747, 2013. (Cited on pages 110 and 116.)

Hess S., Sausen N., and Melkonian M. *Shedding light on vampires: The phylogeny of vampyrellid amoebae revisited.* plos one, 7(2), 2012. (Cited on pages 110 and 115.)

Hidy A.J., Gosse J.C., Froese D.G., Bond J.D., and Rood D.H. *A latest pliocene age for the earliest and most extensive cordilleran ice sheet in northwestern canada.* Quaternary Science Reviews, 61:77–84, 2013. (Cited on page 50.)

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., and Jarvis A. *Very high resolution interpolated climate surfaces for global land areas.* International Journal of Climatology, 25(15):1965–1978, 2005. (Cited on page 36.)

Hillebrand H. *On the generality of the latitudinal diversity gradient.* The American Naturalist, 163(2):192–211, 2004. (Cited on pages 38, 69, and 75.)

Hillebrand H., and Azovsky A.I. *Body size determines the strength of the latitudinal diversity gradient.* Ecography, 24(3):251–256, 2001. (Cited on page 75.)

Hollander M., and Wolfe D.A. *Nonparametric statistical methods.* Wiley, 1999. (Cited on page 112.)

Holm S. *A simple sequentially rejective multiple test procedure.* Scandinavian journal of statistics, 6(2):65–70, 1979. (Cited on page 112.)

Horner N.R., Grenville-Briggs L.J., and van West P. *The oomycete pythium oligandrum expresses putative effectors during mycoparasitism of phytophthora infestans and is amenable to transformation.* Fungal Biology, 116(1):24–41, 2012. (Cited on page 32.)

Horner-Devine M.C., and Bohannan B.J.M. *Phylogenetic clustering and overdispersion in bacterial communities.* Ecology, 87 (7):S100–S108, 2006. (Cited on page 23.)

Horswell J., Cordiner S.J., Maas E.W., Martin T.M., Sutherland B.W., Speir T.W., et al. *Forensic comparison of soils by bacterial community dna profiling.* Journal of Forensic Sciences, 47 (2):350–353, 2002. (Cited on page 82.)

Howard G.T., Duos B., and Watson-Horzelski E.J. *Characterization of the soil microbial community associated with the decomposition of a swine carcass.* International Biodeterioration & Biodegradation, 64(4):300–304, 2010. (Cited on pages 82 and 88.)

Howe A.T., Bass D., Vickerman K., Chao E.E., and Cavalier-Smith T. *Phylogeny, taxonomy, and astounding genetic diversity of glissomonadida ord. nov., the dominant gliding zooflagellates in soil (protozoa: Cercozoa).* Protist, 160(2):159–189, 2009. (Cited on pages 115 and 116.)

Hu S.K., Liu Z.F., Lie A.A.Y., Countway P.D., Kim D.Y., Jones A.C., et al. *Estimating protistan diversity using high-throughput sequencing.* Journal of Eukaryotic Microbiology, 62 (5):688–693, 2015. (Cited on pages 78 and 107.)

Hu S.K., Campbell V., Connell P., Gellene A.G., Liu Z., Terrado R., and Caron D.A. *Protistan diversity and activity inferred from rna and dna at a coastal ocean site in the eastern north pacific.* FEMS microbiology ecology, 92(4):fiw050, 2016. (Cited on page 107.)

Hubert N., and Hanner R. *Dna barcoding, species delineation and taxonomy: a historical perspective.* DNA Barcodes, 3(1):44–58, 2015. (Cited on page 3.)

Hurlbert S.H. *The nonconcept of species diversity: a critique and alternative parameters.* Ecology, 52(4):577–586, 1971. (Cited on pages 68 and 75.)

Huss V.A.R., Frank C., Hartmann E.C., Hirmer M., Kloboucek A., Seidel B.M., et al. *Biochemical taxonomy and molecular phylogeny of the genus Chlorella sensu lato (Chlorophyta).* Journal of Phycology, 35(3):587–598, 1999. (Cited on page 58.)

Hutchinson G.E. *Cold spring harbor symposium on quantitative biology.* Concluding remarks, 22:415–427, 1957. (Cited on page 6.)

Hutchinson G.E. *The paradox of the plankton.* American Naturalist, 95(882):137–145, 1961. (Cited on pages 6 and 20.)

Hájek T., and Beckett R.P. *Effect of water content components on desiccation and recovery in sphagnum mosses.* Annals of Botany, 101(1):165–173, 2008. (Cited on page 8.)

Imparato V., Santos S.S., Johansen A., Geisen S., and Winding A. *Stimulation of bacteria and protists in rhizosphere of glyphosate-treated barley.* Applied Soil Ecology, 98:47–55, 2016. (Cited on page 26.)

Izaguirre I., Vinocur A., Mataloni G., and Pose M. *Phytoplankton communities in relation to trophic status in lakes from hope bay (antarctic peninsula).* Hydrobiologia, 369:73–87, 1998. (Cited on page 78.)

Izaguirre I., Saad J.F., Schiaffino M.R., Vinocur A., Tell G., Sanchez M.L., et al. *Drivers of phytoplankton diversity in patagonian and antarctic lakes across a latitudinal gradient (2150 km): the importance of spatial and environmental factors.* Hydrobiologia, 764(1):157–170, 2016. (Cited on pages 34, 38, 69, 75, and 77.)

Janzen D.H. *Herbivores and the number of tree species in tropical forests.* American naturalist, pages 501–528, 1970. (Cited on pages 100 and 107.)

Jassey V. *Impact d'un réchauffement climatique sur le fonctionnement de la sphagnosphère: relations polyphenols-communautés microbiennes.* Thesis, 2011. (Cited on pages 9 and 53.)

Jassey V.E.J., Gilbert D., Binet P., Toussaint M-L., and Chiapusio G. *Effect of a temperature gradient on sphagnum fallax and its associated living microbial communities: a study under controlled conditions.* Canadian Journal of Microbiology, 57(3):226–235, 2011. (Cited on pages 6, 17, 34, and 35.)

Jassey V.E.J., Chiapusio G., Binet P., Buttler A., Laggoun-Defarge F., Delarue F., et al. *Above- and belowground linkages in sphagnum peatland: climate warming affects plant-microbial interactions.* Global Change Biology, 19(3):811–823, 2013. (Cited on pages 41 and 44.)

Jassey V.E.J., Lamentowicz L., Robroek B.J.M., Gazbka M., Rusinska A., and Lamentowicz M. *Plant functional diversity drives niche-size-structure of dominant microbial consumers along a poor to extremely rich fen gradient.* Journal of Ecology, 102(5):1150–1162, 2014. (Cited on page 26.)

Jassey V.E.J., Signarbieux C., Hattenschwiler S., Bragazza L., Buttler A., Delarue F., et al. *An unexpected role for mixotrophs in the response of peatland carbon cycling to climate warming.* Scientific Reports, 5, 2015. (Cited on pages 6, 34, 35, 41, and 53.)

Jassey V.E.J., Lamentowicz M., Bragazza L., Hofsommer M.L., Mills R.T.E., Buttler A., et al. *Loss of testate amoeba functional diversity with increasing frost intensity across a continental gradient reduces microbial activity in peatlands.* European journal of protistology, 2016. (Cited on page 53.)

Jones Meredith D. M., Forn I., Gadelha C., Egan M.J., Bass D., Massana R., and Richards T.A. *Discovery of novel intermediate forms redefines the fungal tree of life.* Nature, 474(7350): 200–U234, 2011. (Cited on pages 32, 64, and 110.)

Jones R.I. *Mixotrophy in planktonic protists: an overview.* Freshwater Biology, 45(2):219–226, 2000. (Cited on pages 6 and 34.)

Joosten H. *The global peatland co2 picture: peatland status and drainage related emissions in all countries of the world.* The Global Peatland CO2 Picture: peatland status and drainage related emissions in all countries of the world, 2009. (Cited on page 53.)

Kamoun S., Furzer O., Jones J.D.G., Judelson H.S., Ali G.S., Dalio R.J.D., et al. *The top 10 oomycete pathogens in molecular plant pathology.* Molecular Plant Pathology, 16(4):413–434, 2015. (Cited on page 28.)

Kappel T., and Anken R.H. *An aqauarium myxomycete: Didymium nigripes.* Mycologist, 6(3):106 – 107, 1992. (Cited on page 97.)

Karpov S.A., Bass D., Mylnikov A.P., and Cavalier-Smith T. *Molecular phylogeny of cercomonadidae and kinetid patterns of cercomonas and eocercomonas gen. nov (cercomonadida, cercozoa).* Protist, 157(2):125–158, 2006. (Cited on page 5.)

Karpov S.A., Mikhailov K.V., Mirzaeva G.S., Mirabdullaev I.M., Mamkaeva K.A., Titova N.N., and Aleoshin V.V. *Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi.* Protist, 164(2):195–205, 2013. (Cited on page 92.)

Katoh K., Misawa K., Kuma K-I., and Miyata T. *Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform.* Nucleic acids research, 30(14):3059–3066, 2002. (Cited on page 102.)

Keeling P.J., Burki F., Wilcox H.M., Allam B., Allen E.E., Amaral-Zettler L.A., et al. *The marine microbial eukaryote transcriptome sequencing project (mmetsp): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing.* PLoS Biol, 12(6):e1001889, 2014. (Cited on pages 4 and 141.)

Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D., et al. *Picante: R tools for integrating phylogenies and ecology.* Bioinformatics, 26(11):1463–1464, 2010. (Cited on page 23.)

Ketcheson S.J., and Price J.S. *Characterization of the fluxes and stores of water within newly formed sphagnum moss cushions and their environment.* Ecohydrology, 7(2):771–782, 2014. (Cited on page 40.)

Kettridge N., and Waddington J.M. *Towards quantifying the negative feedback regulation of peatland evaporation to drought.* Hydrological Processes, 28(11):3728–3740, 2014. (Cited on page 40.)

Kimura M. *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.* Journal of molecular evolution, 16(2):111–120, 1980. (Cited on pages 5 and 45.)

Kivinen E., and Pakarinen P. *Geographical Distribution of Peat Resources and Major Peatland Complex Types in the World.* Suomalainen Tiedeakatemia, 1981. ISBN 9789514104145. (Cited on page 35.)

Knoll A.H. *Paleobiological perspectives on early eukaryotic evolution.* Cold Spring Harbor perspectives in biology, 6(1):a016121, 2014. (Cited on page 108.)

Koch T.A., and Ekelund F. *Strains of the heterotrophic flagellate bodo designis from different environments vary considerably with respect to salinity preference and ssu rrna gene composition.* Protist, 156(1):97–112, 2005. (Cited on page 5.)

Koeppel A.F., and Wu M. *Species matter the role of competition in the assembly of congeneric bacteria.* Isme Journal, 8 (3):531–540, 2014. (Cited on page 23.)

Kooistra W.H.C.F., Hargraves P., Andersen R.A., Balzano S., Zingone A., and Sarno D. Pseudo-cryptic diversity and phylogeography in the centric diatom skeletonema. In *Phycologia*, volume 44, pages 56–57. INT PHYCOLOGICAL SOC NEW BUSINESS OFFICE, PO BOX 1897, LAWRENCE, KS 66044-8897 USA, 2005. (Cited on page 49.)

Kosakyan A., Heger T.J., Leander B.S., Todorov M., Mitchell E.A.D., and Lara E. *Coi barcoding of nebelid testate amoebae (amoebozoa: Arcellinida): Extensive cryptic diversity and redefinition of the hyalospheniidae schultze.* Protist, 163(3):415–434, 2012. (Cited on pages 5, 14, 15, 16, 17, 22, 25, 44, 45, 48, and 52.)

Kosakyan A., Gomaa F., Mitchell E.A.D., Heger T.J., and Lara E. *Using dna-barcoding for sorting out protist species complexes: A case study of the nebela tincta-collaris-bohemica group (amoebozoa; arcellinida, hyalospheniidae).* European Journal of Protistology, 49(2):222–237, 2013. (Cited on pages 4, 14, 15, 17, 21, 22, 25, 26, 44, 48, and 52.)

Kosakyan A., Mulot M., Mitchell E.A.D., and Lara E. *Environmental dna coi barcoding for quantitative analysis of protists communities: A test using the nebela collaris complex (amoebozoa; arcellinida; hyalospheniidae).* European Journal of Protistology, 51(4):311–320, 2015. (Cited on pages 21, 22, 23, 26, 34, and 141.)

Kosakyan A., Lahr D.J.G., Mulot M., Meisterfeld R., Mitchell E.A.D., and Lara E. *Phylogenetic reconstruction based on coi reshuffles the taxonomy of hyalosphenid shelled (testate) amoebae and reveals the convoluted evolution of shell plate shapes.* Cladistics, 32(6):606–623, 2016. (Cited on pages 9 and 52.)

Kreutz Martin, and Foissner Wilhelm. *The sphagnum ponds of simmelried in germany: a biodiversity hot-spot for microscopic organisms.* 2006. (Cited on page 52.)

Krienitz L., Hegewald E.H., Hepperle D., Huss V.A.R., Rohrs T., and Wolf M. *Phylogenetic relationship of Chlorella and Parachlorella gen. nov (Chlorophyta, Trebouxiophyceae).* Phycologia, 43(5):529–542, 2004. (Cited on page 58.)

Kristiansen J. *16. dispersal of freshwater algae—a review.* Hydrobiologia, 336(1-3):151–157, 1996. (Cited on page 48.)

Kudryavtsev A., and Pawlowski J. *Cunea n. g. (amoebozoa, dactylopodida) with two cryptic species isolated from different areas of the ocean.* European Journal of Protistology, 51(3):197–209, 2015. (Cited on pages 92 and 96.)

Kudryavtsev A., Bernhard D., Schlegel M., Chao E.E.Y., and Cavalier-Smith T. *18s ribosomal rna gene sequences of cochliopodium (himatismenida) and the phylogeny of amoebozoa.* Protist, 156(2):215–224, 2005. (Cited on page 97.)

Kudryavtsev A., Wylezich C., and Pawlowski J. *Ovalopodium desertum n. sp and the phylogenetic relationships of cochliopodiidae (amoebozoa).* Protist, 162(4):571–589, 2011. (Cited on pages 96 and 97.)

Kunin V., Engelbrektson A., Ochman H., and Hugenholtz P. *Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates.* Environmental microbiology, 12(1):118–123, 2010. (Cited on page 5.)

Kwasna H., Bateman G.L., and Ward E. *Microbiota in wheat roots evaluated by cloning of its1/2 rdna and sequencing.* Journal of Phytopathology, 158(4):278–287, 2010. (Cited on page 32.)

Lahr D.J.G., Bergmann P.J., and Lopes S. *Taxonomic identity in microbial eukaryotes: A practical approach using the testate amoeba centropyxis to resolve conflicts between old and new taxonomic descriptions.* Journal of Eukaryotic Microbiology, 55(5):409–416, 2008. (Cited on page 17.)

Lahr D.J.G., Grant J., Molestina R., Katz L.A., and Anderson O.R. *Sapocribrum chincoteaguense n. gen. n. sp.: A small, scale-bearing amoebozoan with flabellinid affinities.* Journal of Eukaryotic Microbiology, 62(4):444–453, 2015. (Cited on page 96.)

Lamentowicz M., van der Knaap W.O., Lamentowicz Ł., van Leeuwen J.F.N., Mitchell E.A.D., Goslar T., and Kamenik C. *A near-annual palaeohydrological study based on testate amoebae from a sub-alpine mire: surface wetness and the role of climate during the instrumental period.* Journal of Quaternary Science, 25(2):190–202, 2010. (Cited on page 20.)

Lande R. *Genetic variation and phenotypic evolution during allopatric speciation.* American Naturalist, pages 463–479, 1980. (Cited on page 3.)

Lara E., and Belbahri L. *Ssu rrna reveals major trends in oomycete evolution.* Fungal Diversity, 49(1):93–100, 2011. (Cited on pages 28, 52, and 103.)

Lara E., Chatzinotas A., and Simpson A.G.B. *Andalucia (n. gen.) - the deepest branch within jakobids (jakobida; excavata), based on morphological and molecular study of a new flagellate from soil.* Journal of Eukaryotic Microbiology, 53(2): 112–120, 2006. (Cited on page 93.)

Lara E., Berney C., Ekelund F., Harms H., and Chatzinotas A. *Molecular comparison of cultivable protozoa from a pristine and a polycyclic aromatic hydrocarbon polluted site.* Soil Biology and Biochemistry, 39(1):139–148, 2007a. (Cited on pages 93 and 101.)

Lara E., Berney C., Harms H., and Chatzinotas A. *Cultivation-independent analysis reveals a shift in ciliate 18s rrna gene diversity in a polycyclic aromatic hydrocarbon-polluted soil.* FEMS microbiology ecology, 62(3):365–373, 2007b. (Cited on page 44.)

Lara E., Heger T.J., Mitchell E.A.D., Meisterfeld R., and Ekelund F. *Ssu rrna reveals a sequential increase in shell complexity among the euglyphid testate amoebae (rhizaria : Euglyphida).* Protist, 158(2):229–237, 2007c. (Cited on pages 85 and 93.)

Lara E., Heger T.J., Ekelund F., Lamentowicz M., and Mitchell E.A.D. *Ribosomal rna genes challenge the monophyly of the hyalospheniidae (amoebozoa: Arcellinida).* Protist, 159(2):165–176, 2008. (Cited on page 5.)

Lara E., Moreira D., and Lopez-Garcia P. *The environmental clade lkm11 and rozella form the deepest branching clade of fungi.* Protist, 161(1):116–121, 2009a. (Cited on pages 5, 65, 92, and 110.)

Lara E., Moreira D., Vereshchaka A., and Lopez-Garcia P. *Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids.* Environmental Microbiology, 11(1):47–55, 2009b. (Cited on pages 6, 59, and 92.)

Lara E., Seppey C.V.W., Gonzalez Garraza G., Singer D., Victoria Quiroga M., and Mataloni G. *Planktonic eukaryote molecular diversity: discrimination of minerotrophic and ombrotrophic peatland pools in tierra del fuego (argentina).* Journal of Plankton Research, 37(3):645–655, 2015. (Cited on pages 34, 79, and 110.)

Lara E., Roussel-Delif L., Fournier B., Wilkinson D.M., and Mitchell E.A.D. *Soil microorganisms behave like macroscopic organisms: patterns in the global distribution of soil euglyphid testate amoebae.* Journal of Biogeography, 43(3):520–532, 2016. (Cited on pages 7, 34, 36, 38, 53, and 75.)

Lara Enrique, Mitchell Edward A. D., Moreira David, and Garcia Purificacion Lopez. *Highly diverse and seasonally dynamic protist community in a pristine peat bog.* Protist, 162(1):14–32, 2011. (Cited on pages 9, 28, 31, 32, 35, 49, 51, 52, 64, 65, 69, and 92.)

Laybourn-Parry J., and Pearce D.A. *The biodiversity and ecology of antarctic lakes: models for evolution.* Philosophical Transactions of the Royal Society B-Biological Sciences, 362(1488): 2273–2289, 2007. (Cited on page 34.)

Lecroq B., Lejzerowicz F., Bachar D., Christen R., Esling P., Baerlocher L., et al. *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(32): 13177–13182, 2011. (Cited on page 28.)

Lee J.J., Leedale G.F., and Bradbury P.C. *An illustrated guide to the protozoa: organisms traditionally referred to as protozoa, or newly discovered groups.* Number vol. 2 in An Illustrated Guide to the Protozoa: Organisms Traditionally Referred to as Protozoa, Or Newly Discovered Groups. Society of Protozoologists, 2000. (Cited on pages 104 and 107.)

Lefevre E., Roussel B., Amblard C., and Sime-Ngando T. *The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton.* Plos One, 3(6), 2008. (Cited on pages 64 and 65.)

Lefranc M., Thenot A., Lepere U., and Debroas D. *Genetic diversity of small eukaryotes in lakes differing by their trophic status.* Applied and Environmental Microbiology, 71(10):5935–5942, 2005. (Cited on page 64.)

Legendre P., and Birks H.J.B. From classical to canonical ordination. In *Tracking environmental change using lake sediments*, pages 201–248. Springer, 2012. (Cited on page 80.)

Legendre P., and Gallagher E.D. *Ecologically meaningful transformations for ordination of species data.* Oecologia, 129(2): 271–280, 2001. (Cited on pages 36, 80, and 85.)

Legendre P., and Legendre L.F.J. *Numerical Ecology.* Developments in Environmental Modelling. Elsevier Science, 1998. ISBN 9780080523170. (Cited on pages 68 and 75.)

Legendre P., Borcard D., and Peres-Neto P.R. *Analyzing beta diversity: Partitioning the spatial variation of community composition data.* Ecological Monographs, 75(4):435–450, 2005. (Cited on page 80.)

Leigh J.W., and Bryant D. *popart: full-feature software for haplotype network construction.* Methods in Ecology and Evolution, 6(9):1110–1116, 2015. (Cited on page 45.)

Leliaert F., Verbruggen H., Vanormelingen P., Steen F., López-Bautista J.M., Zuccarello G.C., and De Clerck O. *Dna-based species delimitation in algae.* European journal of phycology, 49(2):179–196, 2014. (Cited on page 5.)

Lennon J.J., Beale C.M., Reid C.L., Kent M., and Pakeman R.J. *Are richness patterns of common and rare species equally well explained by environmental variables?* Ecography, 34(4):529–539, 2011. (Cited on pages 65 and 69.)

Lentendu G., Hübschmann T., Müller S., Dunker S., Buscot F., and Wilhelm C. *Recovery of soil unicellular eukaryotes: An efficiency and activity analysis on the single cell level.* Journal of Microbiological Methods, 95(3):463–469, 2013. (Cited on page 28.)

Lentendu G., Wubet T., Chatzinotas A., Wilhelm C., Buscot F., and Schlegel M. *Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach.* Molecular Ecology, 23(13):3341–3355, 2014. (Cited on pages 92 and 107.)

Lepere C., Boucher D., Jardillier L., Domaizon I., and Debroas D. *Succession and regulation factors of small eukaryote community. composition in a lacustrine ecosystem (Lake pavin).* Applied and Environmental Microbiology, 72(4):2971–2981, 2006. (Cited on page 64.)

Lepere C., Domaizon I., Taib N., Mangot J.F., Bronner G., Boucher D., and Debroas D. *Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems.* Fems Microbiology Ecology, 85(1):85–94, 2013. (Cited on pages 68, 69, and 77.)

Lesaulnier C., Papamichail D., McCorkle S., Ollivier B., Skiena S., Taghavi S., et al. *Elevated atmospheric co2 affects soil microbial diversity associated with trembling aspen.* Environmental Microbiology, 10(4):926–941, 2008. (Cited on pages 51, 92, and 114.)

Levin S.A. *The problem of pattern and scale in ecology.* Ecology, 73(6):1943–1967, 1992. (Cited on page 78.)

Li W.K.W., Rao D.V.S., Harrison W.G., Smith J.C., Cullen J.J., Irwin B., and Platt T. *Autotrophic picoplankton in the tropical ocean.* Science, 219(4582):292–295, 1983. (Cited on page 68.)

Limpens J., Berendse F., Blodau C., Canadell J.G., Freeman C., Holden J., et al. *Peatlands and the carbon cycle: from local processes to global implications–a synthesis.* Biogeosciences, 5 (5):1475–1491, 2008. (Cited on page 8.)

Lin X., Green S., Tfaily M.M., Prakash O., Konstantinidis K.T., Corbett J.E., et al. *Microbial community structure and activity linked to contrasting biogeochemical gradients in bog and fen environments of the glacial lake agassiz peatland.* Applied and environmental microbiology, 78(19):7023–7031, 2012. (Cited on page 52.)

Lindley L.A., Stephenson S.L., and Spiegel F.W. *Protostelids and myxomycetes isolated from aquatic habitats.* Mycologia, 99(4): 504–509, 2007. (Cited on page 97.)

Lindo Z., and Gonzalez A. *The bryosphere: An integral and influential component of the earth's biosphere.* Ecosystems, 13 (4):612–627, 2010. (Cited on pages 9, 26, and 35.)

Linnaeus C. *Systema naturae per regna tria naturae secundum classes, ordines, genera, species*, volume 1. impensis Georg Emanuel Beer, 1788. (Cited on page 3.)

Liu Y., de Bruijn I., Jack A.L.H., Drynan K., van den Berg A.H., Thoen E., et al. *Deciphering microbial landscapes of fish eggs to mitigate emerging diseases.* The ISME journal, 8(10):2002–2014, 2014. (Cited on page 32.)

Logares R., Lindstrom E.S., Langenheder S., Logue J.B., Paterson H., Laybourn-Parry J., et al. *Biogeography of bacterial communities exposed to progressive long-term environmental change.* Isme Journal, 7(5):937–948, 2013. (Cited on page 69.)

Logares R., Audic S., Bass D., Bittner L., Boutte C., Christen R., et al. *Patterns of rare and abundant marine microbial eukaryotes.* Current Biology, 24(8):813–821, 2014. (Cited on pages 68, 69, 77, 101, and 107.)

Losos J.B. *Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species.* Ecology Letters, 11(10):995–1003, 2008. (Cited on page 26.)

Lotze F., and Hölder H. *Handbuch der stratigraphischen Geologie Bd. IV Jura.* Ferdinand Enke, 1964. (Cited on page 21.)

Lovejoy C., and Potvin M. *Microbial eukaryotic distribution in a dynamic beaufort sea and the arctic ocean.* Journal of Plankton Research, 33(3):431–444, 2011. (Cited on page 77.)

Lukes J., Flegontova O., and Horak A. *Quick guide diplonemids.* Current Biology, 25(16):R702–R704, 2015. (Cited on page 92.)

Lynch M.D.J., and Neufeld J.D. *Ecology and exploration of the rare biosphere.* Nature Reviews Microbiology, 13(4):217–229, 2015. (Cited on page 69.)

López-García P., Rodriguez-Valera F., Pedrós-Alió C., and Moreira D. *Unexpected diversity of small eukaryotes in deep-sea antarctic plankton.* Nature, 409(6820):603–607, 2001. (Cited on pages 5, 51, and 92.)

Ma M. *Species richness vs evenness: independent relationship and different responses to edaphic factors.* Oikos, 111(1):192–198, 2005. (Cited on pages 68 and 75.)

MacArthur R.H., and Wilson E.O. The theory of island biogeography, 1967. (Cited on page 68.)

Maddison W.P., and Maddison D.R. *Mesquite: a modular system for evolutionary analysis.* 2001. (Cited on page 46.)

Magoc T., and Salzberg S.L. *Flash: fast length adjustment of short reads to improve genome assemblies.* Bioinformatics, 27 (21):2957–2963, 2011. (Cited on page 112.)

Magoč T., and Salzberg S.L. *Flash: fast length adjustment of short reads to improve genome assemblies.* Bioinformatics, 27 (21):2957–2963, 2011. (Cited on page 101.)

Magurran A.E. Why diversity? In *Ecological diversity and its measurement*, pages 1–5. Springer, 1988. (Cited on pages 68 and 75.)

Magurran A.E. *Measuring Biological Diversity.* Wiley, 2004. ISBN 9780632056330. (Cited on pages 68, 69, and 75.)

Magurran A.E., and Henderson P.A. *Explaining the excess of rare species in natural species abundance distributions.* Nature, 422 (6933):714–716, 2003. (Cited on pages 68 and 69.)

Mahé F., Rognes T., Quince C., De Vargas C., and Dunthorn M. *Swarm: robust and fast clustering method for amplicon-based studies.* PeerJ, 2:e593, 2014. (Cited on pages 5, 29, 35, 79, 94, 101, 104, 107, and 112.)

Mahé F., Mayor J., Bunge J., Chi J., Siemensmeyer T., Stoeck T., et al. *Comparing high-throughput platforms for sequencing the v4 region of ssu-rdna in environmental microbial eukaryotic diversity surveys.* Journal of Eukaryotic Microbiology, 62(3): 338–345, 2015. (Cited on page 101.)

Mahe F., de Vargas C., Bass D., Czech L., Stamatakis A., Lara E., et al. *Soil protists communities in three neotropical rainforests are hyperdiverse and dominated by parasites.* Nature Evolution and Ecology, 2017. (Cited on pages 4, 40, and 51.)

Maidana N.I., Izaguirre I., Vinocur A., Mataloni G., , and Pizarro H. *Diatomeas en una transecta patagónieo-antártica.* Ecología austral, 15:159–176, 2005. (Cited on page 69.)

Mallet J. *A species definition for the modern synthesis.* Trends in Ecology & Evolution, 10(7):294–299, 1995. (Cited on page 3.)

Mangot J-F., Debroas D., and Domaizon I. *Perkinsozoa, a well-known marine protozoan flagellate parasite group, newly identified in lacustrine systems: a review.* Hydrobiologia, 659(1): 37–48, 2011. (Cited on pages 58 and 64.)

Mangot J.F., Domaizon I., Taib N., Marouni N., Duffaud E., Bronner G., and Debroas D. *Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes.* Environmental Microbiology, 15(6): 1745–1758, 2013. (Cited on page 69.)

Manneville O. *des tourbières et des marais.* Delachaux et Niestlé, 1999. (Cited on page 8.)

Marcisz K., Fournier B., Gilbert D., Lamentowicz M., and Mitchell E.A.D. *Response of sphagnum peatland testate amoebae to a 1-year transplantation experiment along an artificial hydrological gradient.* Microbial ecology, 67(4):810–818, 2014a. (Cited on page 29.)

Marcisz K., Lamentowicz L., Słowińska S., Słowiński M., Muszak W., and Lamentowicz M. *Seasonal changes in sphagnum peatland testate amoeba communities along a hydrological gradient.* European journal of protistology, 50(5):445–455, 2014b. (Cited on pages 44 and 52.)

Martin M. *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet. journal, 17(1):pp–10, 2011. (Cited on page 101.)

Martiny J.B.H., Bohannan B.J.M., Brown J.H., Colwell R.K., Fuhrman J.A., Green J.L., et al. *Microbial biogeography: putting microorganisms on the map.* Nature Reviews Microbiology, 4(2):102–112, 2006. (Cited on page 80.)

Massana R. *Eukaryotic picoplankton in surface oceans.* Annual Review of Microbiology, Vol 65, 65:91–110, 2011. (Cited on page 68.)

Massana R., Balague V., Guillou L., and Pedros-Alio C. *Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches.* Fems Microbiology Ecology, 50(3):231–243, 2004a. (Cited on page 79.)

Massana R., Castresana J., Balague V., Guillou L., Romari K., Groisillier A., et al. *Phylogenetic and ecological analysis of novel marine stramenopiles.* Applied and Environmental Microbiology, 70(6):3528–3534, 2004b. (Cited on page 92.)

Massana R., del Campo J., Sieracki M.E., Audic S., and Logares R. *Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles.* Isme Journal, 8(4):854–866, 2014. (Cited on pages 92 and 110.)

Massana R., Gobet A., Audic S., Bass D., Bittner L., Boutte C., et al. *Marine protist diversity in european coastal waters and sediments as revealed by high-throughput sequencing.* Environmental microbiology, 17(10):4035–4049, 2015. (Cited on page 101.)

Masters J.C., and Spencer H.G. *Why we need a new genetic species concept.* Systematic Zoology, 38(3):270–279, 1989. (Cited on page 3.)

Mataloni G., Gonzalez Garraza G., and Vinocur A. *Landscape-driven environmental variability largely determines abiotic characteristics and phytoplankton patterns in peat bog pools (Tierra del Fuego, Argentina).* Hydrobiologia, 751(1):105–125, 2015. (Cited on pages 59 and 143.)

Mayr E. *Systematics and the origin of species, from the viewpoint of a zoologist.* Harvard University Press, 1942. ISBN 0674862503. (Cited on pages 3 and 7.)

Mayr E. *Animal species and evolution*, volume 797. Belknap Press of Harvard University Press Cambridge, Massachusetts, 1963. (Cited on page 3.)

Mayr E. *What is a species, and what is not?* Philosophy of Science, pages 262–277, 1996. (Cited on page 3.)

Mazei Yu A., and Tikhonenkov D.V. *Heterotrophic flagellates in the littoral and sublittoral zones of the southeast part of the Pechora Sea.* Oceanology, 46(3):368–375, 2006. (Cited on page 65.)

McFadden G.I., and Melkonian M. *Use of hepes buffer for microalgal culture media and fixation for electron-microscopy.* Phycologia, 25(4):551–557, 1986. (Cited on page 112.)

McNaughton S.J. *Diversity and stability of ecological communities: a comment on the role of empiricism in ecology.* American Naturalist, pages 515–525, 1977. (Cited on pages 68 and 75.)

McWilliam H., Li W.Z., Uludag M., Squizzato S., Park Y.M., Buso N., et al. *Analysis tool web services from the embl-ebi.* Nucleic Acids Research, 41(W1):597–600, 2013. (Cited on page 79.)

Medlin L., Elwood H.J., Stickel S., and Sogin M.L. *The characterization of enzymatically amplified eukaryotic 16s-like rrna-coding regions.* Gene, 71(2):491–499, 1988. (Cited on page 104.)

Meisterfeld R. Order arcellinida kent, 1880. In Lee JJ, Leedale GF, and Bradbury P, editors, *An Illustrated Guide to the Protozoa*, pages 827–859. Society of Protozoologists, Lawrence, Kansas, 2000a. (Cited on pages 82 and 116.)

Meisterfeld R. Testate amoebae with filopodia. In Lee JJ, Leedale GF, and Bradbury P, editors, *An Illustrated Guide to the Protozoa*, pages 1054–1084. Society of Protozoologists, Lawrence, Kansas, 2000b. (Cited on page 82.)

Meisterfeld R. *Order arcellinida kent, 1880.* The illustrated guide to the Protozoa, 2:827–860, 2002. (Cited on page 44.)

Melis C., Selva N., Teurlings I., Skarpe C., Linnell J.D.C., and Andersen R. *Soil and vegetation nutrient response to bison carcasses in bialeowieza primeval forest, poland.* Ecological Research, 22(5):807–813, 2007. (Cited on page 82.)

Mendoza L., Taylor J.W., and Ajello L. *The class mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary.* Annual Reviews in Microbiology, 56(1):315–344, 2002. (Cited on page 40.)

Mitchell E.A.D. *Response of testate amoebae (protozoa) to n and p fertilization in an arctic wet sedge tundra.* Arctic Antarctic and Alpine Research, 36(1):78–83, 2004. (Cited on pages 82 and 88.)

Mitchell E.A.D. *Pack hunting by minute soil testate amoebae: nematode hell is a naturalist's paradise.* Microbiol Rev, 15: 321–353, 2015. (Cited on page 107.)

Mitchell E.A.D., and Meisterfeld R. *Taxonomic confusion blurs the debate on cosmopolitanism versus local endemism of free-living protists.* Protist, 156(3):263–267, 2005. (Cited on pages 7 and 44.)

Mitchell E.A.D., Buttler A.J., Warner B.G., and Gobat J.M. *Ecology of testate amoebae (protozoa : Rhizopoda) in sphagnum peatlands in the jura mountains, switzerland and france.* Ecoscience, 6(4):565–576, 1999. (Cited on page 82.)

Mitchell E.A.D., Gilbert D., Buttler A., Amblard C., Grosvernier P., and Gobat J.M. *Structure of microbial communities in sphagnum peatlands and effect of atmospheric carbon dioxide enrichment.* Microbial Ecology, 46(2):187–199, 2003. (Cited on page 14.)

Mitchell E.A.D., Charman D.J., and Warner B.G. *Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future.* Biodiversity and Conservation, 17(9): 2115–2137, 2008. (Cited on pages 14, 25, and 82.)

Monchy S., Sanciu G., Jobard M., Rasconi S., Gerphagnon M., Chabe M., et al. *Exploring and quantifying fungal diversity in freshwater lake ecosystems using rdna cloning/sequencing and ssu tag pyrosequencing.* Environmental Microbiology, 13 (6):1433–1453, 2011. (Cited on page 68.)

Moon-van der Staay S.Y., De Wachter R., and Vaulot D. *Oceanic 18s rdna sequences from picoplankton reveal unsuspected eukaryotic diversity.* Nature, 409(6820):607–610, 2001. (Cited on page 51.)

Moore P.D. *The future of cool temperate bogs.* Environmental Conservation, 29(1):3–20, 2002. (Cited on pages 8 and 35.)

Moore T.R., and Dalva M. *The influence of temperature and water table position on carbon dioxide and methane emissions from laboratory columns of peatland soils.* Journal of Soil Science, 44(4):651–664, 1993. (Cited on page 54.)

Mora C., Tittensor D.P., Adl S., Simpson A.G.B., and Worm B. *How many species are there on earth and in the ocean?* PLoS Biol, 9(8):e1001127, 2011. (Cited on page 4.)

Moreno L.I., Mills D-E., Fetscher J., John-Williams K., Meadows-Jantz L., and McCord B. *The application of amplicon length heterogeneity pcr (lh-pcr) for monitoring the dynamics of soil microbial communities associated with cadaver decomposition.* Journal of Microbiological Methods, 84(3):388–393, 2011. (Cited on pages 82 and 88.)

Muyzer G., Dewaal E.C., and Uitterlinden A.G. *Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16s rrna.* Applied and Environmental Microbiology, 59(3):695–700, 1993. (Cited on page 79.)

Nassonova E., Smirnov A., Fahrni J., and Pawlowski J. *Barcoding amoebae: Comparison of ssu, its and coi genes as tools for molecular identification of naked lobose amoebae.* Protist, 161 (1):102–115, 2010. (Cited on pages 17 and 45.)

Nebel M., Pfabel C., Stock A., Dunthorn M., and Stoeck T. *Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rrna gene sequences.* Environmental Microbiology Reports, 3(2):154–158, 2011. (Cited on page 5.)

Needleman S.B., and Wunsch C.D. *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* Journal of Molecular Biology, 48(3):443–453, 1970. (Cited on page 29.)

Nolte V., Pandey R.V., Jost S., Medinger R., Ottenwalder B., Boenigk J., and Schlotterer C. *Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity.* Molecular Ecology, 19(14):2908–2915, 2010. (Cited on pages 75 and 77.)

Okland R.H., Rydgren K., and Okland T. *Species richness in boreal swamp forests of se norway the role of surface microtopography.* Journal of Vegetation Science, 19(1):67–74, 2008. (Cited on page 20.)

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., et al. *vegan: Community Ecology Package. R package version 2.0-2. 2011.* 2012. (Cited on pages 59 and 85.)

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., et al. *vegan: Community Ecology Package. R package version 2.0-2. 2011.* 2015. (Cited on pages 21, 36, 103, and 112.)

Olive L.S. *The protostelida - a new order of the mycetozoa.* Mycologia, 59(1):1, 1967. (Cited on page 97.)

Oliverio A.M., Lahr D.J.G., Grant J., and Katz L.A. *Are microbes fundamentally different than macroorganisms? convergence and a possible case for neutral phenotypic evolution in testate amoeba (amoebozoa: Arcellinida).* Royal Society open science, 2(12):150414, 2015. (Cited on page 45.)

Orsi W., Biddle J.F., and Edgcomb V. *Deep sequencing of subseafloor eukaryotic rrna reveals active fungi across marine subsurface provinces.* PLoS One, 8(2):e56335, 2013. (Cited on page 101.)

Pagni M., Niculita-Hirzel H., Pellissier L., Dubuis A., Xenarios I., Guisan A., et al. *Density-based hierarchical clustering of pyrosequences on a large scale-the case of fungal ITS1.* Bioinformatics, 29(10):1268–1274, 2013. (Cited on pages 5 and 59.)

Paradis E., Claude J., and Strimmer K. *Ape: analyses of phylogenetics and evolution in r language.* Bioinformatics, 20(2): 289–290, 2004. (Cited on page 45.)

Passy S.I. *A distinct latitudinal gradient of diatom diversity is linked to resource supply.* Ecology, 91(1):36–41, 2010. (Cited on pages 38 and 75.)

Pawlowski J. *The new micro-kingdoms of eukaryotes.* Bmc Biology, 11, 2013. (Cited on page 4.)

Pawlowski J., and Burki F. *Untangling the phylogeny of amoeboid protists.* Journal of Eukaryotic Microbiology, 56(1):16–25, 2009. (Cited on page 97.)

Pawlowski J., Bolivar I., Fahrni J., de Vargas C., and Bowser S.S. *Naked foraminiferans revealed.* Nature, 399(6731):27–27, 1999. (Cited on page 89.)

Pawlowski J., Fahrni J., Lecroq B., Longet D., Cornelius N., Excoffier L., et al. *Bipolar gene flow in deep-sea benthic foraminifera.* Molecular ecology, 16(19):4089–4096, 2007. (Cited on pages 7, 49, and 53.)

Pawlowski J., Audic S., Adl S., Bass D., Belbahri L., Berney C., et al. *Cbol protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms.* Plos Biology, 10(11), 2012. (Cited on pages 3, 5, 18, 44, and 100.)

Pawlowski J., Esling P., Lejzerowicz F., Cedhagen T., and Wilding T.A. *Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities.* Molecular ecology resources, 14(6):1129–1140, 2014. (Cited on page 101.)

Pawlowski J., Lejzerowicz F., Apotheloz-Perret-Gentil L., Visco J., and Esling P. *Protist metabarcoding and environmental biomonitoring: Time for change.* European journal of protistology, 55(A, SI):12–25, 2016. (Cited on pages 5, 51, and 110.)

Payne J.A. *A summer carrion study of the baby pig sus scrofa linnaeus.* Ecology, 46(5):592–602, 1965. (Cited on pages 82 and 84.)

Payne R., Gauci V., and Charman D.J. *The impact of simulated sulfate deposition on peatland testate amoebae.* Microbial Ecology, 59(1):76–83, 2010. (Cited on page 83.)

Pearson W.R. *Blast and fasta similarity searching for multiple sequence alignment.* Multiple Sequence Alignment Methods, pages 75–101, 2014. (Cited on pages 29 and 35.)

Pedros-Alio C. *Marine microbial diversity: can it be determined?* Trends in Microbiology, 14(6):257–263, 2006. (Cited on pages 69, 78, and 79.)

Perrigo A.L., Baldauf S.L., and Romeralo M. *Diversity of dictyostelid social amoebae in high latitude habitats of northern sweden.* Fungal Diversity, 58(1):185–198, 2013. (Cited on page 38.)

Petsch D.K., Pinha G.D., Dias J.D., and Takeda A.M. *Temporal nestedness in Chironomidae and the importance of environmental and spatial factors in species rarity.* Hydrobiologia, 745(1):181–193, 2015. (Cited on page 65.)

Petz W., and Foissner W. *The effects of mancozeb and lindane on the soil microfauna of a spruce forest: A field study using a completely randomized block design.* Biology and Fertility of Soils, 7(3):225–231, 1989. (Cited on page 82.)

Petz W., Foissner W., and Adam H. *Culture, food selection and growth-rate in the mycophagous ciliate Grossglockneria-acuta foissner, 1980 - 1st evidence of autochthonous soil ciliates.* Soil biology & biochemestry, 17(6):871–875, 1985. (Cited on page 110.)

Philippot L., Raaijmakers J.M., Lemanceau P., and van der Putten W.H. *Going back to the roots: the microbial ecology of the rhizosphere.* Nature Reviews Microbiology, 11(11):789–799, 2013. (Cited on page 26.)

Phillips A.J., Anderson V.L., Robertson E.J., Secombes C.J., and van West P. *New insights into animal pathogenic oomycetes.* Trends in microbiology, 16(1):13–19, 2008. (Cited on page 28.)

Pianka E.R. *Latitudinal gradients in species diversity: A review of concepts.* American Naturalist, 100(910):33, 1966. (Cited on page 75.)

Pianka E.R. *The structure of lizard communities.* Annual Review of Ecology and Systematics, 4(1):53–74, 1973. (Cited on page 22.)

Pick U. *Dunaliella—a model extremophilic alga.* Israel Journal of plant sciences, 46(2):131–139, 1998. (Cited on page 20.)

Pocheville A. *The ecological niche: history and recent controversies*, pages 547–586. Springer, 2015. (Cited on page 5.)

Pohlert T. *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)*, 2014. R package. (Cited on page 112.)

Pomeroy L.R. *The ocean's food web, a changing paradigm.* Bioscience, 24(9):499–504, 1974. (Cited on pages 6 and 68.)

Pommier T., Neal P.R., Gasol J.M., Coll M., Acinas S.G., and Pedros-Alio C. *Spatial patterns of bacterial richness and evenness in the nw mediterranean sea explored by pyrosequencing of the 16s rrna.* Aquatic Microbial Ecology, 61(3):212–224, 2010. (Cited on page 78.)

Pontarp M., Canback B., Tunlid A., and Lundberg P. *Phylogenetic analysis suggests that habitat filtering is structuring marine bacterial communities across the globe.* Microbial Ecology, 64 (1):8–17, 2012. (Cited on page 23.)

Porter S.M. Tiny vampires in ancient seas: evidence for predation via perforation in fossils from the 780–740 million-year-old chuar group, grand canyon, usa. In *Proc. R. Soc. B*, volume 283, page 20160221. The Royal Society, 2016. ISBN 0962-8452. (Cited on page 34.)

Posada D., and Crandall K.A. *Modeltest: testing the model of dna substitution.* Bioinformatics, 14(9):817–818, 1998. (Cited on page 46.)

Post W.M., Emanuel W.R., Zinke P.J., and Stangenberger A.G. *Soil carbon pools and world life zones.* Nature, 298(5870):156–159, 1982. (Cited on page 41.)

Pouliot R., Rochefort L., and Karofeld E. *Initiation of microtopography in revegetated cutover peatlands.* Applied Vegetation Science, 14(2):158–171, 2011. (Cited on page 29.)

Prangnell J., and McGowan G. *Soil temperature calculation for burial site analysis.* Forensic Science International, 191(1-3): 104–109, 2009. (Cited on page 82.)

Preston F.W. *The commonness, and rarity, of species.* Ecology, 29(3):254–283, 1948. (Cited on page 69.)

Pruesse E., Quast C., Knittel K., Fuchs B.M., Ludwig W., Peplies J., and Gloeckner F.O. *Silva: a comprehensive online resource for quality checked and aligned ribosomal rna sequence data compatible with arb.* Nucleic acid reasearch, 35(21):7188–7196, 2007. (Cited on pages 111 and 112.)

Puillandre N., Lambert A., Brouillet S., and Achaz G. *Abgd, automatic barcode gap discovery for primary species delimitation.* Mol Ecol, 21(8):1864–77, 2012. (Cited on pages 5, 22, and 45.)

Quiroga M.V., Unrein F., Gonzalez Garraza G., Kueppers G., Lombardo R., Marinone M.C., et al. *The plankton communities from peat bog pools: structure, temporal variation and environmental factors.* Journal of Plankton Research, 35(6):1234–1253, 2013. (Cited on pages 58 and 59.)

Quirós R., and Cuch S. Relaciones entre visibilidad, fósforo total y concentración de clorofila en 32 lago patagónicos, argentina. In *Actas XII Congreso Nacional del Agua, Mendoza. I, Estudios de aguas superficiales y subterráneas*, pages 1–30, 1985. (Cited on page 78.)

Quirós R., and Drago E. *The environmental state of argentinean lakes: an overview.* Lakes & Reservoirs: Research & Management, 4(1-2):55–64, 1999. (Cited on page 78.)

R Core Team . *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013. (Cited on pages 36, 45, 85, 103, and 112.)

Ramette A. *Multivariate analyses in microbial ecology.* Fems Microbiology Ecology, 62(2):142–160, 2007. (Cited on page 80.)

Rengefors K., Logares R., and Laybourn-Parry J. *Polar lakes may act as ecological islands to aquatic protists.* Molecular Ecology, 21(13):3200–3209, 2012. (Cited on page 75.)

Rezanezhad F., Price J.S., Quinton W.L., Lennartz B., Milojevic T., and Van Cappellen P. *Structure of peat soils and implications for water storage, flow and solute transport: A review update for geochemists.* Chemical Geology, 429:75–84, 2016. (Cited on page 8.)

Richards T.A., Vepritskiy A.A., Gouliamova D.E., and Nierzwicki-Bauer S.A. *The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages.* Environmental Microbiology, 7(9): 1413–1425, 2005. (Cited on pages 58, 64, and 92.)

Richards T.A., Leonard G., Mahé F., del Campo J., Romac S., Jones M.D.M., et al. Molecular diversity and distribution of marine fungi across 130 european environmental samples. In *Proc. R. Soc. B*, volume 282, page 20152243. The Royal Society, 2015. (Cited on page 107.)

Rippon J.W. *Medical Mycology: The Pathogenic Fungi and the Pathogenic Actinomycetes.* Saunders, 1982. ISBN 9780721675862. (Cited on page 97.)

Rizzo D.M., Garbelotto M., and Hansen E.M. *Phytophthora ramorum: integrative research and management of an emerging pathogen in california and oregon forests.* Annu. Rev. Phytopathol., 43:309–335, 2005. (Cited on page 28.)

Rochera C., Toro M., Rico E., Fernandez-Valiente E., Villaescusa J.A., Picazo A., et al. *Structure of planktonic microbial communities along a trophic gradient in lakes of byers peninsula, south shetland islands.* Antarctic Science, 25(2):277–287, 2013. (Cited on page 77.)

Rodriguez W.C., and Bass W.M. *Decomposition of buried bodies and methods that may aid in their location.* Journal of Forensic Sciences, 30(3):836–852, 1985. (Cited on page 82.)

Rogerson A. *Parvamoeba rugata n.g., n.sp., (gymnamoebia, thecamoebidae): An exceptionally small marine naked amoeba.* European Journal of Protistology, 29(4):446–452, 1993. (Cited on page 96.)

Rognes T. *Faster smith-waterman database searches with intersequence simd parallelisation.* Bmc Bioinformatics, 12, 2011. (Cited on page 59.)

Rognes T., Flouri T., Nichols B., Quince C., and Mahé F. *Vsearch: a versatile open source tool for metagenomics.* PeerJ, 4:e2584, 2016. (Cited on page 101.)

Ronquist F., and Huelsenbeck J.P. *Mrbayes 3: Bayesian phylogenetic inference under mixed models.* Bioinformatics, 19(12): 1572–1574, 2003. (Cited on pages 45 and 46.)

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Hohna S., et al. *Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space.* Systematic Biology, 61(3):539–542, 2012. (Cited on page 22.)

Rosenzweig M.L. *Species Diversity in Space and Time.* Cambridge University Press, 1995. ISBN 9780521499521. (Cited on page 75.)

Rueckert S., and Leander B.S. *Molecular phylogeny and surface morphology of marine archigregarines (apicomplexa), selenidium spp., filipodium phascolosomae n. sp., and platyproteum ng and comb. from north-eastern pacific peanut worms (sipuncula).* Journal of Eukaryotic Microbiology, 56(5):428–439, 2009. (Cited on page 102.)

Rueckert S., Chantangsi C., and Leander B.S. *Molecular systematics of marine gregarines (apicomplexa) from north-eastern pacific polychaetes and nemerteans, with descriptions of three novel species: Lecudina phyllochaetopteri sp. nov., difficilina tubulani sp. nov. and difficilina paranemertis sp. nov.* International journal of systematic and evolutionary microbiology, 60 (11):2681–2690, 2010. (Cited on page 102.)

Rueckert S., Villette P.M.A.H., and Leander B.S. *Species boundaries in gregarine apicomplexan parasites: A case study—comparison of morphometric and molecular variability in lecudina cf. tuzetae (eugregarinorida, lecudinidae).* Journal of Eukaryotic Microbiology, 58(4):275–283, 2011. (Cited on page 107.)

Rydin H., and Jeglum J.K. *The Biology of Peatlands.* Biology of Habitats Series. OUP Oxford, 2013. ISBN 9780191508295. (Cited on page 28.)

Saad J.F., Schiaffino M.R., Vinocur A., O'Farrell I., Tell G., and Izaguirre I. *Microbial planktonic communities of freshwater environments from tierra del fuego: dominant trophic strategies in lakes with contrasting features.* Journal of Plankton Research, 35(6):1220–1233, 2013. (Cited on pages 34 and 65.)

Sachs L. *Angewandte statistik: anwendung statistischer methoden.* Springer-Verlag, 2013. ISBN 3662057484. (Cited on page 36.)

Salinas H.F.O., Alder V.A., Puig A., and Boltovskoy D. *Latitudinal diversity patterns of diatoms in the southwestern atlantic and antarctic waters.* Journal of Plankton Research, 37(4):659–665, 2015. (Cited on page 38.)

Sanders R.W. *Mixotrophic protists in marine and freshwater ecosystems.* The Journal of protozoology, 38(1):76–81, 1991. (Cited on page 36.)

Sapp M., Schwaderer A.S., Wiltshire K.H., Hoppe H., Gerdts G., and Wichels A. *Species-specific bacterial communities in the phycosphere of microalgae?* Microbial ecology, 53(4):683–699, 2007. (Cited on page 116.)

Sarowar M.N., van den Berg A.H., McLaggan D., Young M.R., and van West P. *Reprint of: Saprolegnia strains isolated from river insects and amphipods are broad spectrum pathogens.* Fungal Biology, 118(7):579–590, 2014. (Cited on page 32.)

Scheffer M., Rinaldi S., Huisman J., and Weissing F.J. *Why plankton communities have no equilibrium solutions to the paradox.* Hydrobiologia, 491(1-3):9–18, 2003. (Cited on pages 6 and 20.)

Schiaffino M.R., Unrein F., Gasol J.M., Massana R., Balagué V., , and Izaguirre I. *Bacterial community structure in a latitudinal gradient of lakes: the roles of spatial versus environmental factors.* Freshwater Biology, 56(10):1973–1991, 2011. (Cited on pages 69, 75, and 78.)

Schiaffino M.R., Lara E., Fernández L.D., Balagué V., Singer D., Seppey C.V.W., et al. *Microbial eukaryote communities exhibit robust biogeographical patterns along a gradient of patagonian and antarctic lakes.* Environmental Microbiology, 2016. (Cited on pages 35 and 36.)

Schlegel M., and Meisterfeld R. *The species problem in protozoa revisited.* European Journal of Protistology, 39(4):349–355, 2003. (Cited on page 4.)

Schönborn W. *Comparative studies on the production biology of protozoan communities in freshwater and soil ecosystems.* Archiv für Protistenkunde, 141(3):187–214, 1992. (Cited on page 20.)

Scoble J.M., and Cavalier-Smith T. *Scale evolution in paraphysomonadida (chrysophyceae): Sequence phylogeny and revised taxonomy of paraphysomonas, new genus clathromonas, and 25 new species.* European Journal of Protistology, 50(5): 551–592, 2014. (Cited on page 65.)

Seppey C.V.W., Fournier B., Szelecz I., Singer D., Mitchell E.A.D., and Lara E. *Response of forest soil euglyphid testate amoebae (rhizaria: Cercozoa) to pig cadavers assessed by high-throughput sequencing.* International Journal of Legal Medicine, 130(2):551–562, 2016. (Cited on pages 51 and 94.)

Seppey C.V.W., Singer D., Dumack K, Belbahri L., Mitchell E.A.D., and Lara E. *Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling.* Soil Biology and Biochemistry, 2017. (Cited on page 6.)

Shadwick L.L., Spiegel F.W., Shadwick J.D.L., Brown M.W., and Silberman J.D. *Eumycetozoa = amoebozoa?: Ssurdna phylogeny of protosteloid slime molds and its significance for the amoebozoan supergroup.* Plos One, 4(8), 2009. (Cited on pages 92 and 97.)

Shalchian-Tabrizi K., Reier-Roberg K., Ree D.K., Klaveness D., and Brate J. *Marine-freshwater colonizations of haptophytes inferred from phylogeny of environmental 18s rdna sequences.* Journal of Eukaryotic Microbiology, 58(4):315–318, 2011. (Cited on page 102.)

Sharp J.H., Peltzer E.T., Alperin M.J., Cauwet G., Farrington J.W., Fry B., et al. *Measurement of dissolved organic carbon and nitrogen in natural waters: procedures subgroups report.* Marine Chemistry, 41(1-3):37–49, 1993. (Cited on page 59.)

Shiratori T., and Ishida K-I. *A new heterotrophic cryptomonad: Hemiarma marina n. g., n. sp.* Journal of Eukaryotic Microbiology, 63(6):804–812, 2016. (Cited on page 36.)

Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., et al. *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Molecular Systems Biology, 7, 2011. (Cited on page 60.)

Silvola J., Alm J., Ahlholm U., Nykanen H., and Martikainen P.J. *Co2 fluxes from peat in boreal mires under varying temperature and moisture conditions.* Journal of Ecology, pages 219–228, 1996. (Cited on page 53.)

Sime-Ngando T., and Niquil N. *Editorial: 'disregarded' microbial diversity and ecological potentials in aquatic systems: a new paradigm shift ahead.* Hydrobiologia, 659(1):1–4, 2011. (Cited on pages 58 and 64.)

Simon M., Lopez-Garcia P., Deschamps P., Moreira D., Restoux G., Bertolino P., and Jardillier L. *Marked seasonality and high spatial variability of protist communities in shallow freshwater systems.* Isme Journal, 9(9):1941–1965, 2015. (Cited on page 69.)

Simon M., Lopez-Garcia P., Deschamps P., Restoux G., Bertolino P., Moreira D., and Jardillier L. *Resilience of freshwater communities of small microbial eukaryotes undergoing severe drought events.* Frontiers in Microbiology, 7, 2016. (Cited on page 69.)

Singer D., Kosakyan A., Pillonel A., Mitchell E.A.D., and Lara E. *Eight species in the nebela collaris complex: Nebela gimlii (arcellinida, hyalospheniidae), a new species described from a swiss raised bog.* European Journal of Protistology, 51(1):79–85, 2015. (Cited on pages 5, 20, 22, 25, 44, and 52.)

Singer D., Lara E., Steciow Mca M., Seppey C.V.W., Paredes N., Pillonel A., et al. *High-throughput sequencing reveals diverse oomycete communities in oligotrophic peat bog micro-habitat.* Fungal Ecology, 23:42–47, 2016. (Cited on pages 21, 35, and 52.)

Siqueira T., Bini L.M., Roque F.O., Marques Couceiro S.R., Trivinho-Strixino S., and Cottenie K. *Common and rare species respond to similar niche processes in macroinvertebrate metacommunities.* Ecography, 35(2):183–192, 2012. (Cited on page 65.)

Slapeta J., and Linares M.C. *Combined amplicon pyrosequencing assays reveal presence of the apicomplexan "type-n"(cf. gemmocystis cylindrus) and chromera velia on the great barrier reef, australia.* PloS one, 8(9):e76095, 2013. (Cited on page 102.)

Slapeta J., Moreira D., and Lopez-Garcia P. *The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes.* Proceedings of the Royal Society B-Biological Sciences, 272(1576):2073–2081, 2005. (Cited on pages 58 and 64.)

Smirnov A., and Goodkov A. *Ultrastructure and geographic distribution of the genus paradermamoeba (gymnamoebia, thecamoebidae).* European Journal of Protistology, 40(2):113–118, 2004. (Cited on page 97.)

Smirnov A.V., Nassonova E.S., Chao E., and Cavalier-Smith T. *Phylogeny, evolution, and taxonomy of vannellid amoebae.* Protist, 158(3):295–324, 2007. (Cited on page 97.)

Smirnov A.V., Bedjagina O.M., and Goodkov A.V. *Dermamoeba algensis n. sp (amoebozoa, dermamoebidae) - an algivorous lobose amoeba with complex cell coat and unusual feeding mode.* European Journal of Protistology, 47(2):67–78, 2011a. (Cited on pages 93 and 97.)

Smirnov A.V., Chao E., Nassonova E.S., and Cavalier-Smith T. *A revised classification of naked lobose amoebae (amoebozoa: Lobosa).* Protist, 162(4):545–570, 2011b. (Cited on page 97.)

Smith H.G., and Wilkinson D.M. *Not all free-living microorganisms have cosmopolitan distributions–the case of nebela (apodera) vas certes (protozoa: Amoebozoa: Arcellinida).* Journal of Biogeography, 34(10):1822–1831, 2007. (Cited on page 7.)

Smith J.M. *Sympatric speciation.* American Naturalist, pages 637–650, 1966. (Cited on page 3.)

Sogin M.L., and Silberman J.D. *Evolution of the protists and protistan parasites from the perspective of molecular systematics.* International Journal for Parasitology, 28(1):11–20, 1998. (Cited on page 4.)

Sogin M.L., Morrison H.G., Huber J.A., Mark Welch D., Huse S.M., Neal P.R., et al. *Microbial diversity in the deep sea and the underexplored "rare biosphere"*. Proceedings of the National Academy of Sciences of the United States of America, 103(32): 12115–12120, 2006. (Cited on page 68.)

Soininen J., McDonald R., and Hillebrand H. *The distance decay of similarity in ecological communities*. Ecography, 30(1):3–12, 2007. (Cited on page 80.)

Soininen J., Korhonen J.J., Karhu J., and Vetterli A. *Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton*. Limnology and Oceanography, 56(2):508–520, 2011. (Cited on pages 68, 75, 77, and 80.)

Soininen J., Jamoneau A., Rosebery J., and Passy S.I. *Global patterns and drivers of species and trait composition in diatoms*. Global Ecology and Biogeography, 25(8):940–950, 2016. (Cited on page 38.)

Sonneborn T.M. Breeding systems, reproductive methods, and species problems in protozoa. American Association for the Advancement of Science, 1957. (Cited on page 3.)

Sonneborn T.M. *The paramecium aurelia complex of fourteen sibling species*. Transactions of the American Microscopical Society, pages 155–178, 1975. (Cited on page 3.)

Sorensen N., Daugbjerg N., and Richardson K. *Choice of pore size can introduce artefacts when filtering picoeukaryotes for molecular biodiversity studies*. Microbial Ecology, 65(4):964–968, 2013. (Cited on page 77.)

Stamatakis A. *Raxml-vi-hpc: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models*. Bioinformatics, 22(21):2688–2690, 2006. (Cited on page 86.)

Stamatakis A. *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*. Bioinformatics, 30(9): 1312–1313, 2014. (Cited on pages 60, 102, and 104.)

Stamatakis A., Hoover P., and Rougemont J. *A rapid bootstrap algorithm for the raxml web servers*. Systematic Biology, 57 (5):758–771, 2008. (Cited on pages 22, 45, and 93.)

Steciow M.M., Lara E., Pillonel A., Pelizza S.A., Lestani E.A., Rossi G.C., and Belbahri L. *Incipient loss of flagella in the genus geolegnia: the emergence of a new clade within leptolegnia?* IMA fungus, 4(2):169–175, 2013. (Cited on page 28.)

Steciow M.M., Lara E., Paul C., Pillonel A., and Belbahri L. *Multiple barcode assessment within the saprolegnia-achlya clade (saprolegniales, oomycota, straminipila) brings order in a neglected group of pathogens*. IMA fungus, 5(2):439–448, 2014. (Cited on page 28.)

Steinberg C.E.W., Kamara S., Prokhotskaya V.Y., Manusadžianas L., Karasyova T.A., Timofeyev M.A., et al. *Dissolved humic substances–ecological driving forces from the individual to the ecosystem level?* Freshwater Biology, 51(7):1189–1210, 2006. (Cited on pages 8, 20, and 35.)

Stockner J.G., and Antia N.J. *Algal picoplankton from marine and freshwater ecosystems: A multidisciplinary perspective*. Canadian Journal of Fisheries and Aquatic Sciences, 43(12):2472–2503, 1986. (Cited on page 68.)

Stoeck T., Bass D., Nebel M., Christen R., Jones M.D.M., Breiner H-W., and Richards T.A. *Multiple marker parallel tag environmental dna sequencing reveals a highly complex eukaryotic community in marine anoxic water*. Molecular Ecology, 19(s1): 21–31, 2010. (Cited on pages 104 and 107.)

Stoeck T., Breiner H-W., Filker S., Ostermaier V., Kammerlander B., and Sonntag B. *A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology*. Environmental Microbiology, 16(2):430–444, 2014. (Cited on page 58.)

Stokes K.L., Forbes S.L., and Tibbett M. *Human versus animal: Contrasting decomposition dynamics of mammalian analogues in experimental taphonomy*. Journal of Forensic Sciences, 58(3): 583–591, 2013. (Cited on page 84.)

Stomp M., Huisman J., Mittelbach G.G., Litchman E., and Klausmeier C.A. *Large-scale biodiversity patterns in freshwater phytoplankton*. Ecology, 92(11):2096–2107, 2011. (Cited on page 38.)

Stone L., and Roberts A. *The checkerboard score and species distributions*. Oecologia, 85(1):74–79, 1990. (Cited on page 22.)

Sutherland W.J., Freckleton R.P., Godfray H.C.J., Beissinger S.R., Benton T., Cameron D.D., et al. *Identification of 100 fundamental ecological questions*. Journal of ecology, 101(1):58–67, 2013. (Cited on page 5.)

Swindles G.T., and Ruffell A. *A preliminary investigation into the use of testate amoebae for the discrimination of forensic soil samples*. Science & Justice, 49(3):182–190, 2009. (Cited on page 82.)

Szelecz I., Fournier B., Seppey C.V.W., Amendt J., and Mitchell E.A.D. *Can soil testate amoebae be used for estimating the time since death? a field experiment in a deciduous forest*. Forensic Science International, 236:90–98, 2014. (Cited on pages 82, 88, 94, and 97.)

Taberlet P., Coissac E., Pompanon F., Brochmann C., and Willerslev E. *Towards next-generation biodiversity assessment using dna metabarcoding*. Molecular Ecology, 21(8):2045–2050, 2012. (Cited on page 83.)

Taib N., Mangot J.F., Domaizon I., Bronner G., and Debroas D. *Phylogenetic affiliation of ssu rrna genes generated by massively parallel sequencing: New insights into the freshwater protist diversity*. Plos One, 8(3), 2013. (Cited on page 68.)

Talavera G., Dincă V., and Vila R. *Factors affecting species delimitations with the gmyc model: insights from a butterfly survey*. Methods in Ecology and Evolution, 4(12):1101–1110, 2013. (Cited on page 5.)

Tarnawski S.E., and Lara E. *From environmental sequences to morphology: Observation and characterisation of a paulinellid testate amoeba (micropyxidiella edaphonis gen. nov. sp. nov. euglyphida, paulinellidae) from soil using fluorescent in situ hybridization*. Protist, 166(2):264–270, 2015. (Cited on pages 5 and 92.)

Tedersoo L., Bahram M., Põlme S., Kõljalg U., Yorou N.S., Wijesundera R., et al. *Global diversity and geography of soil fungi*. science, 346(6213):1256688, 2014. (Cited on page 100.)

Tekle Y.I., Anderson O.R., Katz L.A., Maurer-Alcala X.X., Romero M.A.C., and Molestina R. *Phylogenomics of 'discosea': A new molecular phylogenetic perspective on amoebozoa with flat body forms*. Molecular Phylogenetics and Evolution, 99:144–154, 2016. (Cited on page 97.)

Tell G., Izaguirre I., and Allende L. *Diversity and geographic distribution of chlorococcales (chlorophyceae) in contrasting lakes along a latitudinal transect in argentinean patagonia*. Biodiversity and Conservation, 20(4):703–727, 2011. (Cited on page 69.)

Terborgh J. *Enemies maintain hyperdiverse tropical forests*. The American Naturalist, 179(3):303–314, 2012. (Cited on page 100.)

Thines M., and Choi Y-J. *Evolution, diversity, and taxonomy of the peronosporaceae, with focus on the genus peronospora*. Phytopathology, 106(1):6–18, 2015. (Cited on page 32.)

Thompson J.D., Higgins D.G., and Gibson T.J. *Clustal-w - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Research, 22(22): 4673–4680, 1994. (Cited on page 22.)

Thormann M.N., and Rice A.V. *Fungi from peatlands*. Fungal Diversity, 24:241–299, 2007. (Cited on pages 8 and 35.)

Tice A.K., Shadwick L.L., Fiore-Donno A.M., Geisen S., Kang S., Schuler G.A., et al. *Expansion of the molecular and morphological diversity of acanthamoebidae (centramoebida, amoebozoa) and identification of a novel life cycle type within the group*. Biology Direct, 11(1):69, 2016. (Cited on page 51.)

Tittel J., Bissinger V., Zippel B., Gaedke U., Bell E., Lorke A., and Kamjunke N. *Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs*. Proceedings of the National Academy of Sciences of the United States of America, 100(22):12776–12781, 2003. (Cited on page 34.)

Titus J.E., and Wagner D.J. *Carbon balance for two sphagnum mosses: water balance resolves a physiological paradox*. Ecology, 65(6):1765–1774, 1984. (Cited on page 53.)

Tolley K.A., Townsend T.M., and Vences M. *Large-scale phylogeny of chameleons suggests african origins and eocene diversification*. Proceedings of the Royal Society of London B: Biological Sciences, 280(1759):20130184, 2013. (Cited on page 49.)

Toms J. D., and Lesperance M. L. *Piecewise regression: A tool for identifying ecological thresholds.* Ecology, 84(8):2034–2041, 2003. (Cited on page 59.)

Toro M., Camacho A., Rochera C., Rico E., Banon M., Fernandez-Valiente E., et al. *Limnological characteristics of the freshwater ecosystems of byers peninsula, livingston island, in maritime antarctica.* Polar Biology, 30(5):635–649, 2007. (Cited on page 78.)

Towne E.G. *Prairie vegetation and soil nutrient responses to ungulate carcasses.* Oecologia, 122(2):232–239, 2000. (Cited on page 82.)

Tsyganov A.N., Nijs I., and Beyens L. *Does climate warming stimulate or inhibit soil protist communities? a test on testate amoebae in high-arctic tundra with free-air temperature increase.* Protist, 162(2):237–248, 2011. (Cited on page 82.)

Tuomisto H., and Ruokolainen K. *Analyzing or explaining beta diversity? understanding the targets of different methods of analysis.* Ecology, 87(11):2697–2708, 2006. (Cited on pages 77 and 80.)

Ulrich W., and Gotelli N.J. *Null model analysis of species associations using abundance data.* Ecology, 91(11):3384–3397, 2010. (Cited on page 23.)

Unrein F., Izaguirre I., Massana R., Balague V., and Gasol J.M. *Nanoplankton assemblages in maritime antarctic lakes: characterisation and molecular fingerprinting comparison.* Aquatic Microbial Ecology, 40(3):269–282, 2005. (Cited on page 79.)

Unrein F., Massana R., Alonso-Sáez L., and Gasol J.M. *Significant year-round effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal system.* Limnology and Oceanography, 52(1):456–469, 2007. (Cited on page 34.)

Unrein F., Gasol J.M., Not F., Forn I., and Massana R. *Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters.* Isme Journal, 8(1):164–176, 2014. (Cited on page 34.)

Valencia R., Balslev H., and Miño G.P.Y. *High tree alpha-diversity in amazonian ecuador.* Biodiversity & Conservation, 3(1):21–28, 1994. (Cited on page 100.)

Valentini A., Pompanon F., and Taberlet P. *Dna barcoding for ecologists.* Trends in ecology & evolution, 24(2):110–7, 2009. (Cited on page 83.)

Valiranta M., Blundell A., Charman D.J., Karofeld E., Korhola A., Sillasoo U., and Tuittila E.S. *Reconstructing peatland water tables using transfer functions for plant macrofossils and testate amoebae: A methodological comparison.* Quaternary International, 268:34–43, 2012. (Cited on pages 14 and 20.)

Vamosi S.M. *Phylogenetic community ecology as an approach for studying old ideas on competition in the plankton opportunities and challenges.* Journal of Limnology, 73:186–192, 2014. (Cited on page 21.)

Van Breemen N. *How sphagnum bogs down other plants.* Trends in Ecology & Evolution, 10(7):270–275, 1995. (Cited on pages 9 and 26.)

Van den Brink P.J., and Ter Braak C.J.F. *Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress.* Environmental Toxicology and Chemistry, 18(2):138–148, 1999. (Cited on page 85.)

van der Wal A., Geydan T.D., Kuyper T.W., and de Boer W. *A thready affair: linking fungal diversity and community dynamics to terrestrial decomposition processes.* FEMS microbiology review, 37(4):477–494, 2013. (Cited on page 44.)

van Hannen E.J., van Agterveld M.P., Gons H.J., and Laanbroek H.J. *Revealing genetic diversity of eukaryotic microorganisms in aquatic environments by denaturing gradient gel electrophoresis.* Journal of Phycology, 34(2):206–213, 1998. (Cited on page 68.)

van Hannen E.J., Zwart G., van Agterveld M.P., Gons H.J., Ebert J., and Laanbroek H.J. *Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses.* Applied and Environmental Microbiology, 65(2):795–801, 1999. (Cited on pages 5, 65, 92, and 110.)

Van Valen L. *Ecological species, multispecies, and oaks.* Taxon, pages 233–239, 1976. (Cited on page 3.)

Vass A.A., Bass W.M., Wolt J.D., Foss J.E., and Ammons J.T. *Time since death determinations of human cadavers using soil solution.* Journal of Forensic Sciences, 37(5):1236–1253, 1992. (Cited on page 82.)

Vaulot D., Eikrem W., Viprey M., and Moreau H. *The diversity of small eukaryotic phytoplankton (<= 3 mu m) in marine ecosystems.* Fems Microbiology Reviews, 32(5):795–820, 2008. (Cited on page 79.)

Vergin K.L., Done B., Carlson C.A., and Giovannoni S.J. *Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean.* Aquatic Microbial Ecology, 71(1):1–129, 2013. (Cited on page 79.)

Verni F., and Gualtieri P. *Feeding behaviour in ciliated protists.* Micron, 28(6):487–504, 1997. (Cited on page 110.)

Vinocur A., and Unrein F. *Typology of lentic water bodies at potter peninsula (king george island, antarctica) based on physical-chemical characteristics and phytoplankton communities.* Polar Biology, 23(12):858–870, 2000. (Cited on page 78.)

Von Humboldt A., and Bonpland A. *Personal narrative of travels to the equinoctial regions of America during the years 1799-1804*, volume 13. Henry G. Bohn, 1853. (Cited on page 100.)

Wakeman K.C., and Leander B.S. *Molecular phylogeny of pacific archigregarines (apicomplexa), including descriptions of veloxidium leptosynaptae n. gen., n. sp., from the sea cucumber leptosynapta clarki (echinodermata), and two new species of selenidium.* Journal of Eukaryotic Microbiology, 59(3):232–245, 2012. (Cited on page 102.)

Wanner M., and Elmer M. *"hot spots" on a new soil surface - how do testate amoebae settle down?* Acta Protozoologica, 48 (3):281–289, 2009. (Cited on page 83.)

Wanner M., Elmer M., Kazda M., and Xylander W.E.R. *Community assembly of terrestrial testate amoebae: How is the very first beginning characterized?* Microbial Ecology, 56(1):43–54, 2008. (Cited on page 83.)

Wardle D.A. *Communities and ecosystems: linking the aboveground and belowground components*, volume 34. Princeton University Press, 2002. ISBN 0691074879. (Cited on page 35.)

Warnes G.R., Bolker B., Bonebakker L., Gentleman R., Liaw W.H.A., Lumley T., et al. Gplots: Various r programming tools for plotting data. r package version 2.16.0., 2015. (Cited on pages 31 and 36.)

Webb C.O., Ackerly D.D., McPeek M.A., and Donoghue M.J. *Phylogenies and community ecology.* Annual Review of Ecology and Systematics, 33:475–505, 2002. (Cited on page 26.)

Weiner A., Aurahs R., Kurasawa A., Kitazato H., and Kucera M. *Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist.* Molecular Ecology, 21 (16):4063–4073, 2012. (Cited on pages 6 and 20.)

Weiner A.K.M., Weinkauf M.F.G., Kurasawa A., Darling K.F., Kucera M., and Grimm G.W. *Phylogeography of the tropical planktonic foraminifera lineage globigerinella reveals isolation inconsistent with passive dispersal by ocean currents.* PloS one, 9(3):e92148, 2014. (Cited on page 49.)

Wetzel R.G. *Limnology: lake and river ecosystems.* Gulf Professional Publishing, 2001. (Cited on page 78.)

Whittaker R.H. *Vegetation of the siskiyou mountains, oregon and california.* Ecological Monographs, 30(3):280–338, 1960. (Cited on page 68.)

Wickham H. *ggplot2: elegant graphics for data analysis.* Springer Science & Business Media, 2009. (Cited on page 103.)

Wiens J.J., Ackerly D.D., Allen A.P., Anacker B.L., Buckley L.B., Cornell H.V., et al. *Niche conservatism as an emerging principle in ecology and conservation biology.* Ecology Letters, 13 (10):1310–1324, 2010. (Cited on page 26.)

Wilken S., Huisman J., Naus-Wiezer S., and Van Donk E. *Mixotrophic organisms become more heterotrophic with rising temperature.* Ecology Letters, 16(2):225–233, 2013. (Cited on page 40.)

Wilkinson D.M. *Fragments of an entangled bank: do ecologists study most of ecology?* Oikos, 82(2):393–394, 1998. (Cited on page 34.)

Wilkinson D.M. *What is the upper size limit for cosmopolitan distribution in free living microorganisms?* Journal of Biogeography, 28(3):285–291, 2001. (Cited on page 7.)

Wilkinson D.M., and Mitchell E.A.D. *Testate amoebae and nutrient cycling with particular reference to soils.* Geomicrobiology Journal, 27(6-7):520–533, 2010. (Cited on page 44.)

Wilkinson D.M., Koumoutsaris S., Mitchell E.A.D., and Bey I. *Modelling the effect of size on the aerial dispersal of microorganisms.* Journal of Biogeography, 39(1):89–97, 2012. (Cited on pages 7 and 44.)

Will K.W., Mishler B.D., and Wheeler Q.D. *The perils of dna barcoding and the need for integrative taxonomy.* Systematic biology, 54(5):844–851, 2005. (Cited on page 3.)

Williams B.R.M., Schaefer H., De Sequeira M.M., Reyes-Betancort J.A., Patiño J., and Carine M.A. *Are there any widespread endemic flowering plant species in macaronesia? phylogeography of ranunculus cortusifolius.* American journal of botany, 102 (10):1736–1746, 2015. (Cited on page 5.)

Willig M.R., Kaufman D.M., and Stevens R.D. *Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis.* Annual Review of Ecology Evolution and Systematics, 34:273–309, 2003. (Cited on page 69.)

Willis A., and Bunge J. *Estimating diversity via frequency ratios.* Biometrics, 71(4):1042–1049, 2015. (Cited on page 103.)

Willoughby L.G. *The occurrence and distribution of reproductive spores of saprolegniales in fresh water.* The Journal of Ecology, pages 733–759, 1962. (Cited on page 28.)

Willoughby L.G. *Saprolegnias of salmonid fish in windermere: a critical analysis.* Journal of Fish Diseases, 1(1):51–67, 1978. (Cited on page 28.)

Wilsey B., and Stirling G. *Species richness and evenness respond in a different manner to propagule density in developing prairie microcosm communities.* Plant Ecology, 190(2): 259–273, 2007. (Cited on page 75.)

Wilson E.O. *The arboreal ant fauna of peruvian amazon forests a first assessment.* Biotropica, pages 245–251, 1987. (Cited on page 100.)

Woese C.R., Kandler O., and Wheelis M.L. *Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya.* Proceedings of the National Academy of Sciences, 87(12):4576–4579, 1990. (Cited on page 4.)

Wolinska J., Giessler S., and Koerner H. *Molecular identification and hidden diversity of novel daphnia parasites from european lakes.* Applied and Environmental Microbiology, 75(22):7051–7059, 2009. (Cited on page 32.)

Wyss C., and Cherix D. *Traité d'entomologie forensique: les insectes sur la scène de crime.* Sciences forensiques. Presses polytechniques et universitaires romandes, 2006. ISBN 9782880746971. (Cited on page 82.)

Yang J., Smith H.G., Sherratt T.N., and Wilkinson D.M. *Is there a size limit for cosmopolitan distribution in free-living microorganisms? a biogeographical analysis of testate amoebae from polar areas.* Microbial ecology, 59(4):635–645, 2010. (Cited on page 7.)

Yu Y., Harris A.J., Blair C., and He X. *Rasp (reconstruct ancestral state in phylogenies): a tool for historical biogeography.* Molecular Phylogenetics and Evolution, 87:46–49, 2015. (Cited on page 46.)

Yu Z., Beilman D.W., Frolking S., MacDonald G.M., Roulet N.T., Camill P., and Charman D.J. *Peatlands and their role in the global carbon cycle.* Eos, Transactions American Geophysical Union, 92(12):97–98, 2011. (Cited on pages 41 and 53.)

Yubuki N., Edgcomb V.P., Bernhard J.M., and Leander B.S. *Ultrastructure and molecular phylogeny of calkinsia aureus: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria.* BMC microbiology, 9(1):1, 2009. (Cited on page 51.)

Zapata J., Yanez M., and Rudolph E. *Thecamoebians (protozoa : Rhizopoda) of the peatland from puyehue national park (40 degrees 45's; 72 degrees 19'w), chile.* Gayana, 72(1):9–17, 2008. (Cited on page 14.)

Zhang J., Kobert K., Flouri T., and Stamatakis A. *Pear: a fast and accurate illumina paired-end read merger.* Bioinformatics, 30(5):614–620, 2014. (Cited on page 101.)

Zhu F., Massana R., Not F., Marie D., and Vaulot D. *Mapping of picoeucaryotes in marine ecosystems with quantitative pcr of the 18s rrna gene.* FEMS microbiology ecology, 52(1):79–92, 2005. (Cited on page 104.)

Zubkov M.V., and Tarran G.A. *High bacterivory by the smallest phytoplankton in the north atlantic ocean.* Nature, 455(7210): 224–226, 2008. (Cited on page 34.)

Škaloud P., and Rindi F. *Ecological differentiation of cryptic species within an asexual protist morphospecies: a case study of filamentous green alga klebsormidium (streptophyta).* Journal of Eukaryotic Microbiology, 60(4):350–362, 2013. (Cited on page 4.)

Šlapeta J., López-García P., and Moreira D. *Global dispersal and ancient cryptic species in the smallest marine eukaryotes.* Molecular Biology and Evolution, 23(1):23–29, 2006. (Cited on pages 7 and 53.)

Appendix I

List of figures

Appendix J

List of tables