# High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity

T J. Heger[1,2,3*], Ian J.W. Giesbrecht[3,4], Julia Gustavsen[5], Javier del Campo[3,6], Colleen T.E Kellogg[3,7], Kira M. Hoffman[3,6], Ken Lertzman [3,4], William W. Mohn[3,7], and Patrick J. Keeling[2,3]

[1] *Soil Science Group, CHANGINS, University of Applied Sciences and Arts Western Switzerland*

[2] *Botany Department, University of British Columbia, Canada*

[3] *Hakai Institute, Heriot Bay, British Columbia, Canada*

[4] *School of Resource and Environmental Management, Simon Fraser University, Canada*

[5] *Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Canada*

[6] *School of Environmental Studies, University of Victoria, Canada*

[7] *Department of Microbiology and Immunology, Life Sciences Institute, University of British Columbia, Canada*

[*] *For correspondence: Thierry Heger, Soil Science Group, CHANGINS, University of Applied Sciences and Arts Western Switzerland, CH-1260 Nyon, Switzerland E-mail: thierry.heger@changins.ch; Tel, +41 22 363 40 73*

**Key words:** high-throughput environmental sequencing, litter, moss, protist, *Sphagnum*, unicellular opisthokonts

**Running title:** Moss and litter-associated protists

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.14061

## Originality-Significance Statement

Our results provide the first detailed insights into the diversity and community structure of a broad range of litter and moss-associated protist groups, and expand our understanding of how edaphic properties, substrate composition, and vegetation cover shape moss- and litter-associated protist communities. A detailed phylogenetic analysis of unicellular opisthokonts also significantly increased our knowledge of the diversity of this group of protists in terrestrial ecosystems. The protocol developed in this study to extract DNA from several grams of litter and moss samples also enabled the study of a broad range of protist groups from habitats which are usually overlooked in molecular surveys of terrestrial protists.

## Abstract

Although previous studies, mostly based on microscopy analyses of a few groups of protists, have suggested that protists are abundant and diverse in litter and moss habitats, the overall diversity of moss and litter associated protists remains elusive. Here, high-throughput environmental sequencing was used to characterize the diversity and community structure of litter and moss-associated protists along a gradient of soil drainage and forest primary productivity in a temperate rainforest in British Columbia. We identified 3,262 distinct protist OTUs from 36 sites. Protists were strongly structured along the landscape gradient, with a significant increase in alpha diversity from the blanket bog ecosystem to the zonal forest ecosystem. Among all investigated environmental variables, calcium content was the most strongly associated with the community composition of protists, but substrate composition, plant cover and other edaphic factors were also significantly correlated with these communities. Furthermore, a detailed phylogenetic analysis of unicellular opisthokonts

identified OTUs covering most lineages, including novel OTUs branching with Discicristoidea, the sister group of Fungi, and with Filasterea, one of the closest unicellular relatives to animals. Altogether, this study provides unprecedented insight into the community composition of moss- and litter-associated protists.

## Introduction

Mosses and litter are major sources of organic matter in peatland and forest soils. They also harbor diverse communities of bacteria, fungi, and protists (unicellular eukaryotes). These moss and litter-associated microorganisms play key role in terrestrial ecosystem function by, for example, degrading soil organic matter and remineralizing nitrogen (Lindo and Gonzalez, 2010; Koller et al., 2013). Among protists, free-living heterotrophic protists play a central role in the decomposition food web by feeding on bacteria, fungi, and other microorganisms. Phototrophic and mixotrophic protists contribute to primary production at the soil surface (Gremmen et al., 2007; Jassey et al., 2015) and the broad diversity of soil protist parasites could play important roles in soil biotic interactions (Ramirez et al., 2014; Dupont et al., 2016; Geisen, 2016; Singer et al., 2016; Mahé et al., 2017). Protists are not only functionally important, but are also diverse, particularly in moss and litter habitats where morphospecies diversity is often higher than in the underlying soil horizons (Coûteaux, 1972; Bamforth, 2010). Despite the importance and diversity of protists from moss and litter habitats, research has been generally restricted to a few groups of protists, such as ciliates or testate amoebae, which are relatively easily identified by microscopy because of their characteristic morphological features. However, even within these groups of protists, it has become clear that cryptic species (which cannot be discriminated by morphology alone) frequently occur (Heger et al., 2013).

To overcome the limitations associated with morphology and culturing, molecular sequencing of 18S rDNA from environmental samples has been widely used to reveal protist diversity from different environments (López-García et al., 2001; Lawley et al., 2004; Behnke et al., 2006), including soil ecosystems (Ramirez et al., 2014; Dupont et al., 2016; Grossmann et al., 2016; Harder et al., 2016; Mahé et al., 2017; Seppey et al., 2017). However, high-throughput environmental sequencing (HTES) of protist diversity from terrestrial habitats is under represented compared with marine habitats in general, and particularly data from moss and litter habitats. Based on the short 18S V9 fragment, Parfrey et *al*. (2014) showed that micro-eukaryote communities differ greatly between leaf litter and soil samples. Singer and colleagues (2016) reported a great diversity of oomycetes in *Sphagnum* mosses, and Geisen et *al*. (2015b) used metatranscriptomic sequences (mostly short sequences) to characterize the protists from soil samples (including a few leaf litter samples). But no comprehensive study from moss and litter habitats has analysed large assemblages of protists and assessed which environmental drivers predominantly shape their spatial patterning across ecological gradients.

Our lack of knowledge of the diversity of protists is particularly striking for unicellular opisthokonts from moss/litter habitats and soils. Unicellular opisthokonts belong to two different clades of the Opisthonkont supergroup: the Holomycota and the Holozoa. Holozoa comprise animals and several unicellular lineages such as the choanoflagellates, the ichthyosporeans and filastereans. Holomycota comprise fungi and the unicellular lineages fonticulids and nuclearids (del Campo et al., 2015). Unicellular opisthokonts were previously believed to occur almost only in aquatic ecosystems, but recent studies have revealed that unicellular opisthokont representatives of the choanoflagellate lineage occur in litter and other soil habitats (Geisen et al., 2015b). It remains unclear whether representatives of the

other unicellular opisthokont lineages also occur in soils and particularly in moss/litter habitats (Geisen et al., 2015b).

One of the reasons for the limited amount of data available on protist communities from moss and litter habitats is likely due to DNA extraction protocols which are not optimized for extracting protists from extremely heterogeneous moss and litter material. In this study, we developed a protocol to extract protists from the soil surface (i.e. mosses and large litter debris) and used Illumina MiSeq HTES to characterize protist communities along a gradient of drainage and forest primary productivity ranging from blanket bogs to zonal temperate rainforests on Calvert and Hecate Islands (Central Coast of British Columbia, Canada; Figure 1). On these islands, though climatic conditions are relatively constant over the whole area, distinct ecosystem types occur at small spatial scales (Thompson et al., 2016). Across fine-scale environmental gradients there is substantial variation in plant cover and composition, substrate composition at the soil surface, and several edaphic variables. Among these last variables, pH is often a primary driver of soil protist communities (Shen et al., 2014; Dupont et al., 2016), varies strongly between ecosystem types. Thus, the mosaic of landscape ecosystem types on these islands presented an excellent opportunity to investigate protist diversity and community structure associated with moss and litter across a landscape gradient.

The main objectives of this study were to: 1) examine with a high-resolution molecular approach protist diversity from different ecosystem types along a gradient of drainage and forest primary productivity; 2) assess the diversity and relative abundance of the unicellular opisthokonts by using a curated reference database and phylogenetic analysis; 3) assess if distinct plant communities harbor correspondingly different protist communities; 4) evaluate how the richness and the composition of protists, putative phototrophic protists (PP) and free-living heterotrophic protists (FHP) change along the drainage and forest productivity

gradient; and 5) determine the primary environmental factors driving protist community structure. Since autotrophs (PP), and predators/decomposers (i.e. FHP) play different ecological roles in the soil surface, these two functional groups of protists might respond to different biotic and abiotic parameters along the ecological gradient.

# Results

### *Ecosystem types*

Litter and moss associated protist communities were investigated along a gradient of increasing tree productivity and soil drainage from four dominant ecosystem types: blanket bogs, bog woodlands, bog forests, and zonal forests. Each ecosystem type corresponds to a classified plant community (Supporting Information Table S1). Substrate composition and several edaphic variables such as pH and calcium differed between ecosystem types ($P < 0.05$; Supporting Information Table S2).

### *Overall protist community composition*

Using general eukaryotic primers to amplify the V4 fragment of the 18S (Stoeck et al., 2010), we obtained a total of 3,007,330 curated 18S reads from 36 soil surface samples across four ecosystem types. OTUs clustering of the sequences at 97% sequence similarity resulted in 8,332 non-singleton eukaryotic OTUs. The eukaryotic communities were characterized by a relative abundance of 15% of protists (39% of OTUs, 3262 OTUs), 10% of fungi (21% of OTUs), 15% of chloroplastida (10% of OTUs, 794 OTUs), and 60% metazoa (30% of OTUs, 2507 OTUs). Within the part of the protist community targetable with our PCR approach, alveolates (36.5% of reads, 26.8% of OTUs), rhizarians (34.5% of reads, 35.2% of OTUs), chlorophytes (17.1% of reads, 18.6% of OTUs) and stramenopiles (8.5% of reads, 11.2% of OTUs) were dominant while amoebozoans, unicellular opisthokonts, centrohelids,

cryptophyceans and other eucaryotes only accounted for 3.4 % of the abundance and 8.1 % of the diversity (Figure 2A). Within these major groups, seven dominant lineages (i.e. lineages including > 3% of reads of the major group) were identified within the amoebozoans (Dictyostelia, Discosea which include mycamoebids (i.e. LKM74 and LEMD255) (Blandenier et al., 2017)), Myxogastria, Schizoplasmodiida, and Tubulinea), two lineages within the chlorophytes (chlorophyceans and trebouxiophyceans), six lineages within stramenopiles (bicosoecids, chrysophyceans, diatoms, eustigmatales, MAST-12, and peronosporomycetes) and four lineages within the rhizarians (Cercomonadida, Glissomonadida, Imbricatea, and Thecofilosea) (Figure 2B). The diversity of the unicellular opisthokonts is described below and in Figure 3. The main lineages within the other major groups represented by less than 1% of the total number of sequences (i.e. centrohelids, cryptophyceans, and incertae sedis) were not detailed in the figure. The relative abundance of protist OTUs varied between $4.5*10^{-6}$ to 5.7%. Protist abundance was dominated by a few OTUs: the six most abundant OTUs constituted more than 25% of the total number of sequences and the 25 most abundant OTUs represented almost 50% of the total number of sequences. Out of the 25 most abundant OTUs, ten were assigned to rhizarians, seven to alveolates, six to chlorophytes, and two to stramenopiles (Figure 2C). The profiles of the rarefaction curves at the single sample level did not reach saturation indicating that not all possible OTUs were sequenced, but when considering the samples at the ecosystem type level, the rarefaction curves revealed a near saturation, suggesting that the majority of the diversity of the four different ecosystems was captured (Supporting Information Fig. S1).

***Detailed phylogenetic analysis of unicellular opisthokonts***

A total of 6,211 unicellular opisthokont reads belonging to 103 distinct OTUs (at 97% similarity) were identified based on the Silva115 eukaryotic database. These 103 distinct

OTUs comprised 3.14% of the total protist richness and 1.39% of the total abundance (Figure 2B). The curated opisthokont reference database and phylogenetic tree from del Campo et *al.* (2015) were then used to determine the phylogenetic placement of these 103 OTUs within unicellular opisthokont lineages. Among the 103 OTUs, 88 OTUs (4,157 reads) were successfully placed on the reference tree. The other 15 OTUs were discarded since the PyNAST algorithm (Caporaso et al., 2010a) failed to align them on the reference alignment. The 88 OTUs covered most major groups of unicellular opisthokonts: 33 OTUs (969 reads) were affiliated to the Craspedida subclade 1 (choanoflagellates), 22 OTUs to the Fonticulida (2259 reads), 17 OTUs to the Ichtyophonida (323 reads), four OTUs to the Filasterea (489 reads), two OTUs to the Nucleariida group (52 reads), two OTUs to the Dermocystida (7 reads), and eight OTUs (58 reads) belonged to the choanoflagellates but could not been classified. In contrast, none of the OTUs were placed within the Craspedida subclade 2, the Acanthoecida, or the Corallochytrea (Figure 3). Most highly and relatively abundant unicellular opisthokont OTUs (i.e. >100 reads) were present in all four ecosystem types, except OTU21699 (Filasterea) which did not occur in the blanket bog ecosystem (Figure 3 and Supporting Information Table S3). The majority of the unicellular opisthokont OTUs from this study were novel. Only 16 OTUs among the 88 OTUs placed within the phylogenetic tree with RAxML-EPA were more than 97% similar to any 18S rDNA sequence of the entire unicellular opisthokont reference database. Thirteen OTUs out of these 16 were highly similar to freshwater sequences in the reference database, two OTUs to soil sequences, and one to a salt lake sequence (i.e. brackish lake with a salinity of c.a. 8 psu). No unicellular opisthokont retrieved from this study was highly similar (≥97%) to any exclusively marine OTU in the reference database. The other unicellular opisthokont taxa identified in this study (72 OTUs) were less than 97% similar to any soil, freshwater and marine sequences of the reference database (Supporting Information Table. S3).

***Patterns of protist richness and diversity along the gradient of ecosystem types***

The OTU richness after rarefaction ranged from 289 to 546 per sample for protists, 202 to 415 for free-living heterotrophic protists and from 34 to 81 for phototrophic protists. Shannon's diversity index varied from 3.9 to 5.3 for protists, 3.3 to 5.0 for free-living heterotrophic protists and 1.9 to 3.7 for phototrophic protists (Figure 4). Linear mixed effect models revealed an overall increase of both OTU richness and diversity of protists over the gradient of drainage and forest primary productivity ($P$<0.05) with the highest OTU richness and diversity values found in the zonal forest and the lowest OTU richness and diversity in the blanket bog (Figure 4A, Supporting Information Table. S4). Similar patterns of diversity were found for functional groups of free-living heterotrophic and phototrophic protists (Figures 4B-C), as well as for most dominant lineages of protists (Supporting Information Table S4). The diversity of most major lineages of protists also tended to increase along the gradient, but not all relationships were significant. Dinoflagellates were the only group of protists which decreased along the gradient, but the relationship was not significant (Supporting Information Table S4). However, the relative abundance of the main protist groups stays fairly stable across the gradient (Supporting Information Fig. S2). The multiple-site framework proposed by (Baselga and Orme, 2012) was used to quantify protist beta-diversity among samples and evaluate whether the variation is due to OTUs turnover and/or nestedness. Protist communities exhibited high beta-diversity values ($\beta_{SOR}$=0.92) and most of the variation in OTU composition along the investigated ecological gradient was explained by OTU turnover ($\beta_{SIM}$=0.91). In contrast, the nestedness component was negligible ($\beta_{NES}$=0.01).

***Relationships between environmental factors and protist communities***

The unconstrained NMDS ordination based on relative OTUs abundance revealed clear community structure of protists and of the two functional groups along the drainage and forest primary productivity gradient. Community structure differences were particularly striking between the blanket bog and the three forested ecosystem types (i.e. bog woodland, bog forest, and zonal forest). Within ecosystem types, samples from the same watershed tend to be more closely related together than the ones from distinct watershed (Figure 5). This observation is not surprising since samples from the same ecosystem type and watershed were collected from the same Ecosystem Comparison Plot (see sampling procedure and Figure S4). Permutational Multivariate Analysis of Variance (PERMANOVA) indicated that protist, FHP, and PP community composition differed significantly across the gradient (Supporting Information Table S5).

We then analysed the correlations between plant cover, substrate composition of the samples (non-*Sphagnum* mosses, *Sphagnum* mosses or litter) and edaphic variables with the community composition of protist, free-living heterotrophic protist, and phototrophic protist community structures. RELATE tests revealed an overall significant relationship between protist and the recorded environmental variables ($P$ <0.05). Using distance-based linear modelling (*DistLM*), we then sought to identify the variables which best explained the community structure of the three datasets. For the protist dataset, sequential tests identified calcium as the strongest predictor of protist communities, followed by the relative abundance of *Sphagnum*, herb cover, sodium, tree cover, C/N ratio, total sulphur, pH, and the relative abundance of litter. These variables explained 48.2 % of the variability observed in the protist communities. The same characteristics were significant predictors of FHP community structure and explained 49.4 % of the overall variability. For PPs, the model also selected calcium as the best predictor of the community structure, followed by potassium, tree cover, herb cover, sulphur, and the relative abundance of non-*Sphagnum* moss (Table 1). These

results suggest that edaphic variables, but also substrate composition and aboveground vegetation are strongly correlated with protist, FHP, and PP communities. Water content was not included among the best explanatory factors affecting protist community structure datasets. Distance-based Redundancy Analysis (db-RDA) of protists, FHP and PP communities were performed using the variables selected with *DistLM.* The total variation in the protist, FHP and PP data explained by dbRDA1 and dbRDA2 were respectively 27.3, 28.7, and 27.3%. The three separate db-RDA plots showed that the strongest environmental variable, calcium, together with pH and tree cover variables were highly correlated with axis 1 which explained most of the variability in microbial communities and tended to separate the samples from blanket bog to zonal forest (Figure 6). The relative abundance of litter and especially C/N ratios contributed most strongly to the axis 2 of the protist and FHP db-RDA plot while the axis 2 of the PP db-RDA plot was mainly defined by potassium (Figure 6C).

Moving-window redundancy analysis further confirmed that the association of protists with edaphic variables was generally stronger than between protists and the other selected variables. This analysis also revealed two abrupt transitions along the ecological gradient. The first one occurred at the boundary between blanket bogs-bog woodlands (Figure 7, sequence 9) where dependence of protists to selected variables dropped to a very low level. The second transition occurred at the boundary between bog forests-zonal forests (Figure 7, sequence 18) where the dependence of protists to selected variables increased again.

Indicator species analysis identified 154 abundant protist OTUs that were either significantly associated to one ecosystem or a combination of ecosystems. Among the 31 OTUs associated to only one ecosystem, 19 OTUs were associated to blanket bogs, one OTU to bog woodlands and 11 to zonal forests. Amongst the ones associated with a combination of ecosystems, the majority were characteristic of forested ecosystems and in particular to zonal forests-bog forests-bog woodlands (78 OTUs) and bog forests-zonal forests (20 OTUs)

(Supporting Information Table S6).

# Discussion

Recent molecular surveys have suggested that soil protist diversity is extremely diverse, but only a handful of studies based on high-throughput environmental sequencing approach reported microeukaryote community data from moss and/or the litter surface layer (Parfrey et al., 2014; Geisen et al., 2015b; Singer et al., 2016). We developed a protocol to extract protist DNA from several grams of litter and moss samples and assessed the protist diversity using Illumina sequencing based on the V4 fragment of the 18S with general eukaryotic primers. This study allowed us to gain information on a broad range of protist groups, including many taxa which are usually overlooked with traditional morphology-based approaches. With the detailed analysis of the diversity of unicellular opisthokonts, we also significantly increased our knowledge of the diversity of this group of protists in terrestrial ecosystems. To the best of our knowledge, this is the first detailed analysis of litter- and moss-associated protists, and their relationships to edaphic variables and plant cover.

### *Moss and litter harbor a great diversity of unknown protists*

Our results showed that associated moss and litter protist communities from four ecosystem types of the coastal temperate rainforest are highly diverse. After stringent sequence filtration, we retained a total of 3,263 protist OTUs. Among them, only 408 (8.8%) were more than 97% similar to any sequences from the Silva115 database. The percentage of unknown sequences might be overestimated, since the Silva reference database does not include all known protist sequences, but the proportion is still very high. These findings not only indicated that protist communities from litter and moss samples comprised a high proportion of novel sequences, but also support the general idea that protists from terrestrial

ecosystems remain poorly described (Geisen et al., 2017). The large percentage of novel sequences from this study is likely due to two combined factors. First, we explored protist communities from poorly investigated compartments of the soil profile and second, we explored protists from a remote and poorly investigated biome. Indeed, almost nothing is known on protist diversity from temperate rainforests and especially the North American Pacific Coast rainforest which covers an extensive region comparable to the area of the United Kingdom (Carpenter et al., 2014) and where year-round moist and mild conditions (Alaback, 1996) offer suitable conditions for protist community development.

Taxa reported in this study were characterized by distinct ecological functions. The tentative assignments of protists to different functional groups at a high taxonomic level revealed a great majority of free-living heterotrophic protists, followed by phototrophic protists (including symbionts) which belonged to several groups such as trebouxiophyceans, chlorophyceans and diatoms. Furthermore, several lineages of putative parasitic lineages such as apicomplexans, perkinsids and ichthyosporeans were identified. These last results support recent findings indicating that protist parasites and in particular apicomplexans are abundant and diverse in soils (Ramirez et al., 2014; Geisen et al., 2015a; Dupont et al., 2016; Mahé et al., 2017).

***Extensive unicellular opisthokont diversity in moss and litter habitats***

One of the most striking results of our study is the extensive diversity of unicellular opisthokonts retrieved from litter and moss samples from the soil surface. The entire unicellular opisthokont reference database from del Campo *et al.* (2015) comprised only a few terrestrial sequences (22 out of 828; 3%). The aforementioned database and an associated phylogenetic 18S rDNA tree allowed us to identify 88 terrestrial OTUs which covered almost all major phylogenetic lineages of unicellular opisthokonts, except for the Corallochytrea

lineage which comprised only one taxon putatively endemic to coral reefs, *Corallochytrium limacisporum* (Raghukumar, 1987).

The four new filasterean OTUs (489 reads) retrieved from this study are of great interest. The current known filasterean diversity is limited: in the 18S reference tree generated by del Campo et *al.* (2015) there are only two filasterean taxa, *Capsaspora owczarzaki* (Hertel et al., 2002) and *Ministeria vibrans* (Tong, 1997). Recently two new species have been added to the filastereans, *Pigoraptor vietnamica* and *Pigoraptor chileana* (Hehenberger et al., 2017). Additionally in the same publication the cluster of environmental sequences named MAOP1 have been placed within the filastereans with support for the first time, increasing the number of environmental OTUs within the group. Filastereans are one of the closest unicellular relatives to the metazoa and are therefore, key organisms for assessing the transition from unicellular eukaryotes to metazoan (Suga et al., 2013).

Another key finding of this study is the great diversity of OTUs affiliated to the Discicristoidea lineage the sister group of Fungi (Torruella et al., 2015). Despite the evolutionary importance of this lineage, little is known about the diversity of this group comprising two groups of unicellular opisthokonts, the *Fonticula* and Nucleariida clades. In the reference database from del Campo et al. (2015), the *Fonticula* clade contained only one characterized taxon, *Fonticula alba,* initially isolated from dog dung (Worley et al., 1979; Brown et al., 2009) and several freshwater and marine environmental sequences. However, a new taxon has been recently described and it branched within our *Fonticula* clade (López-Escardo et al., 2018). These findings suggest that the fonticulids (*sensu lato*) might not be monophyletic. The second clade, the Nucleariida clade contained 19 freshwater and salt lake sequences, which were clustered into four Nucleariida OTUs at 97% identity. In this study, we retrieved 22 new *Fonticula* OTUs corresponding to a total of 2,259 reads. Our BLAST analysis revealed that the genetic similarity between the OTUs from this study and the closest

match of the entire reference database was always lower than 96%. This illustrates how the diversity of *Fonticula* taxa from soil is poorly described. Similarly, the HTES analysis from del Campo et al. (2015) revealed an unexpected diversity and abundance of *Fonticula* in oxic sediments. In this study, two OTUs belonging to the Nuclearian group were found in the blanket bog samples. One of them is likely a new taxon since the similarity to any sequence of the reference database is only 91% while the second one was highly similar to the environmental sequence GQ330607 retrieved from a Swiss peatbog (Lara et al., 2011).

Furthermore, our results demonstrate the presence of choanoflagellates and ichthyosporeans (i.e., Dermocystida and Ichtyophonida) from moss and litter samples, as recently reported in other soil HTES and metatranscriptomic studies (Geisen et al., 2015b; Dupont et al., 2016). Interestingly, all choanoflagellate OTUs retrieved from this study branched within the freshwater and soil clade Craspedida subclade 1 or have unclear affiliations within the Choanoflagellate group. The absence of reads in the Craspedida subclade 2 and Acanthoecida groups is likely not due to primer mismatches since del Campo *et al.* (2015) retrieved a huge diversity of Craspedida and Acanthoecida reads with the same set of primers. Altogether this suggests that Acanthoecida is composed exclusively of marine unicellular opisthokonts and supports the general idea that transitions between marine and freshwater or soil habitats are rare, at least within some lineages of protists (Logares et al., 2009; Heger et al., 2010; Parfrey et al., 2014). In contrast to this, a few soil environmental sequences have been assigned to Acanthoecida by Geisen *et al.* (2015b), but these conclusions have to be interpreted with caution since their sequences were short and not integrated in a phylogenetic context.

***Protist communities are dominated by a small number of OTUs***

We also investigated the 25 most abundant OTUs in the dataset which represent almost 50

percent of all sequences. Contrary to the percentage of unknown sequences in the overall protist dataset, the majority of these OTUs matched known sequences in the Silva database. Here also, our results indicated that phototrophic protists were relatively abundant in litter and moss samples. Indeed, seven out of the 25 most abundant protist OTUs belonged to phototrophic taxa (chlorophytes and diatoms). Putative heterotrophic taxa comprised three ciliates, two apicomplexans, one alveolate of unclear affiliation, and ten Cercozoa (Figure 3C). We also identified an abundant chrysomonad related to the genus *Ochromonas.* Chrysomonads, which are frequently heterotrophic or mixotrophic, and *Ochromonas* sequences in particular have previously been reported in peat bogs (Lara et al., 2011). Surprisingly, we also identified one dinoflagellate among these most abundant OTUs. Dinoflagellates were believed to occur almost only in aquatic ecosystems but recent HTES studies have also reported dinoflagellate taxa from soils (Lentendu et al., 2014).

***Patterns of diversity and relationships between protists and environmental variables along the landscape gradient***

Although many studies have documented how plants, animals, bacteria and fungi change along ecological gradients, very little is known about how a broad range of taxonomic and functional groups of protists differ across ecosystem types, and which environmental variables drive their communities in the surface soil horizons. Our results showed that protists are strongly structured along the landscape gradient from blanket bog to zonal forest (increasing tree productivity and soil drainage), with a significant increase in alpha diversity from the non-forested blanket bog to the zonal forest (Figure 4). This pattern was not only observed for the overall protist dataset, but also for almost all dominant taxonomic groups of protists individually. Altogether, these results support the idea that non-forested bogs harbor specific communities of metazoans, bacteria and protists, with relatively low diversity,

compared to adjacent forested ecosystems (Page et al., 2006; Lamentowicz et al., 2010).

Protist communities along the investigated ecological gradient were significantly driven by a combination of covarying edaphic, substrate composition, and plant cover variables. Among all individual variables investigated, calcium content had by far the highest relationship with protist community composition. These results contrast with recent HTES studies where pH and soil water content were generally identified as the main environmental factors shaping protist and microeukaryote communities in different soils types and from various ecosystems (Bates et al., 2013; Ramirez et al., 2014; Shen et al., 2014; Shi et al., 2015; Dupont et al., 2016). The potential influence of calcium on protist community structure is, however, not a new concept. Several studies have identified calcium as the predominant factor explaining community patterns of some specific groups of protists in mosses along a poor to rich bog/fen gradient (Opravilova and Hajek, 2006; Jassey et al., 2014), suggesting the significance of calcium content might be specific to bog environments, perhaps because it is limited or mobilization is somehow impaired in bogs. It remains unclear whether or not calcium content has a direct or an indirect effect on protist community composition, and the physiological effect of calcium content on protist communities has not yet been investigated in detail (Jassey et al., 2014). Beside calcium, four other edaphic factors also affect the distribution patterns of protists: the sodium content, sulphur content, C/N ratio, and pH. In addition to the four edaphic factors selected, substrate composition also emerged as a significant variable explaining protist data (Figure 6). Therefore, these results support previous findings suggesting that substrate composition, especially *Sphagnum* moss acts as a strong environmental filter on protist and other microbial communities (Heger et al., 2013). Our results also reveal that protist community composition structure was correlated to vascular vegetation such as herb and tree covers. These results agree with conclusions from recent studies showing significant relationships between vegetation and microeukaryotes and

protist communities from bulk soil samples (Ramirez et al., 2014; Tedersoo et al., 2016).We also evaluated whether or not the functional groups, FHP and PP, responded to similar drivers to the protist communities. The FHP communities responded to the same environmental variables as the protists as a whole. Phototrophic protists were also affected by edaphic factors such as calcium content and, to a lesser extent, plant cover and substrate composition. However, the individual factors affecting PP and FHP were not all identical. For example, *Sphagnum* was of lesser importance in explaining PP composition.

However, the moving-window analysis revealed that the association of protists with selected variables is not constant throughout the ecological gradient. Stronger association of protists with selected environmental variables at the extremes of the gradient suggests that protist communities from blanket bogs and zonal forests tend to have narrower ecological tolerances than species from the bog woodlands and the bog forests. This interpretation is in agreement with results from the NMDS and indicator species analysis which revealed very distinct protist communities in the blanket bogs and to a lesser extend in zonal forests. Altogether, these results were consistent with the recent findings from Payne *et al.* (2016) who document an important shift in testate amoeba communities at the boundary between forested and open bog ecosystems.

Linkages between aboveground and belowground communities and ecosystem processes have received a lot of attention in recent years (Wardle et al., 2004 ; van der Putten et al., 2013), but protist communities have usually not been incorporated in these works. Here, we examined protists communities at the interface between these two systems: we found that soil surface protists form assemblages – with repeating patterns of diversity and composition – that correspond with variation in aboveground plant assemblages along an environmental gradient. Despite some overlap among ecosystem types, this association suggests a method for scaling plot-level microbial community observations to the landscape

scale, since the extent of these same ecosystem types have been mapped over extensive areas of the rainforests of British Columbia, via the BC Government's Terrestrial Ecosystem Mapping initiative (Green (2014) for this study area). This suggests that the aboveground ecosystem properties that can be readily mapped through remote sensing (Supporting information Fig. S3 and Thompson *et al.* (2016)) could be used to predict and characterize landscape scale spatial mosaics of protist diversity and composition at the aboveground-belowground interface.

## Conclusions

In this study, we used a modified DNA extraction protocol and a HTES approach to characterize a more comprehensive protist community than previously investigated, from moss and litter samples across a landscape gradient. Our work provides detailed insights into the diversity and community structure of a broad range of protist groups, and expands our understanding of how edaphic properties, substrate composition, and vegetation cover shape terrestrial protist communities. Also, the strong linkages between the aboveground plant assemblages and soil surface protist communities suggest that ecosystem types might be useful for scaling plot-level protist community observations to the landscape scale. Furthermore, a detailed analysis of unicellular opisthokonts supports the idea that a great proportion of protist diversity from moss/litter habitats and soils in general, is still unknown.

# Experimental Procedures

### *Study site*

Calvert and Hecate are remote islands located on the central coast of British Columbia, Canada (Figure 1). The climate of the study area is characterized by high precipitation throughout the year (mean annual precipitation of 3,063 mm), cool summers, and mild

winters (mean annual temperature of 8.6 °C) (Pojar et *al.* (1991)1981-2010 climate norms interpolated for a representative plot (ECP02) using http://www.climatewna.com/ (Wang et al., 2016)). According to the BC Biogeoclimatic Ecosystem Classification (BEC) system, the investigated sites belong to the central variant of the very wet hypermaritime coastal western hemlock subzone (Banner et al., 1993). Four dominant ecosystem types characterized by an increase of tree productivity and soil drainage were investigated in this area: blanket bogs, bog woodlands, bog forests, and zonal forests (Supporting Information Fig. S3). This gradient of increasing tree productivity is thought to be driven by a gradient of soil drainage quality, when controlling for geology and disturbance (Banner et al., 2005). Each ecosystem type corresponds with (and is partially defined by) a classified plant community (vascular plants and bryophytes) previously defined in BC's province–wide BEC zones (Banner et al., 1993). Blanket bogs are non-forested sites with wetland plants, shrub-sized coniferous trees and shallow, nutrient-poor soils, and poor soil drainage. Bog woodlands are treed wetlands with very open canopies and sparse cover of shore pine (*Pinus contorta* var. *contorta*), yellow-cedar (*Xanthocyparis nootkatensis*), and western redcedar (*Thuja plicata*). Bog forests have better drained soils and more tree biomass, but also support wetland plants and often show wetland-like soil properties. Zonal forests are upland sites with greater tree productivity and soil drainage (Banner et al., 2005).

***Experimental design and sampling***

Twelve Ecosystem Comparison Plots (ECPs) were established along gradients of drainage and tree productivity from two watersheds on Calvert Island and from one watershed on Hecate Island (Figure 1). Each watershed was represented by four ECPs, representing four dominant ecosystem types: blanket bog, bog woodland, bog forest, and zonal forest. Plot locations were selected to be representative of these four ecosystem types (Supporting

Information Table S1). Each ECP is 20 m x 20 m, oriented with two sides following the slope. Each ECP was represented by a symmetrical 3-by-3 grid of nine points, spaced at 7.5 m and oriented according to the slope (Supporting Information Fig. S4). If a tree or a rock prevented sampling at a specified point, alternative points were sampled 0.5 m to the left when facing up-slope, or 0.5 m to the right, or 1.0 m to the left. At each sampling point, a soil surface sample (i.e. moss and/or litter) of 10 x10 cm was collected using plastic gloves and an ethanol-cleaned knife. When the thickness of the soil surface layer exceeded >15 cm, only the first 15 cm were taken. The three subsamples from each transect were pooled in a plastic bag, resulting in a total of three samples (i.e. three composite samples) per ECP. Samples were stored in soft coolers containing ice packs and transported to the Hakai laboratory within 6 hours. From each composite sample, a representative subsample of about 9 g was randomly taken and placed into a 50 ml Falcon tube and kept at -80°C until DNA extraction.

***Environmental data collection***

We used the remaining material of the samples for physico-chemical and substrate composition analyses. To characterize the substrate composition of the samples, we quantified the relative abundance of *Sphagnum* mosses (living and dead tissues), non-*Sphagnum* mosses (living and dead tissues), and litter from each sample. The relative abundance of these three substrate types was visually estimated on the basis of a three level scale: 0: <5%, 1: 5-25%, 2: 25-75%, and 3: >75%) after distributing the material homogenously on a 50 x 50 cm tray. For the physicochemical measurement of the samples, around 20 g of air-dried litter subsamples were sent for analysis to the Analytical Chemistry Laboratory of the Ministry of Environment in Victoria, British Columbia. Contents of Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, P, S and Zn were measured after extraction by Mehlich III solution. Total organic matter was measured by loss on ignition; total N, C and S were

measured using combustion analysis and available P was measured with Bray P1 method. Soil pH was measured in water. Water content of each subsample was determined by oven drying at 65°C for 72h. Finally, to evaluate the impact of vegetation on protist community; vascular plant cover of each ECP was visually estimated in summer 2013 during the optimum season of the vegetation. Since the vegetation was fairly homogenous within the plots, the percent cover of trees, shrubs, and herbs was surveyed at the ECP level.

***DNA extraction***

To enhance the recovery of protist (mostly free-living protists) from a relatively large amount of litter and moss material, we developed a protocol to mechanically separate the small size fraction which includes protists, other small organisms and small particles from litter and moss before DNA extraction. Comparable protocols were successfully used for microscopy-based and molecular assessments of protist communities and other microbial groups from soil surface samples (Jassey et al., 2011; Jassey et al., 2013; Singer et al., 2016). Approximately 9 g of previously homogenized material was transferred into a 500-ml glass bottle with 90 ml of distilled water and gently shaken intermittently. Once the sample was thawed, the solution was thoroughly shaken 30 times, then passed through a 150 µm mesh. The retained litter and moss particles were pressed with a spatula washed with 10 ml of distilled water to extract the maximum of filtrates. The retained litter and moss particles were then transferred to a 500-ml glass bottle containing 90 ml of distilled water and the procedure was repeated twice. The three different filtrates were combined to a final volume of 300 ml. A subsample of 40 ml was then filtered through a 0.8-µm pore-size Supor-membrane (Pall, Ann Arbor, MI, USA). Half of the filter was immediately cut into small pieces (*ca*. 1 $mm^2$) with a sterile razor blade and DNA extraction was performed with the FastDNA SPIN Kit for soil (MP Biomedicals, Solon, OH, USA) following the recommendations of the manufacturer. For mechanical lysis

of the cells, bead beating was performed twice on a FastPrep FP120 Instrument (MP Biomedicals, USA) for 30 s at speed 5.0 m/s. DNA extracts were quantified using a Nanodrop ND-1000 spectrophotometer, adjusted to 5ng DNA $\mu L^{-1}$ in ultra-pure water and stored at -80°C. All non-disposable tools used for the DNA extraction procedure were cleaned with 10% bleach, rinsed with distilled water and autoclaved.

***18S rRNA amplification and Illumina sequencing***

Protist communities were investigated using high-throughput Illumina sequencing. The hypervariable V4 region of the 18S rRNA gene (~380 bp) was amplified with the general eukaryotic primer pair TAReuk454FWD1 and TAReukREV3 (Stoeck et al., 2010) combined with Illumina adapter sequences, a pad, a linker as well as 12-bp Golay barcodes on the reverse primers (Supplementary Table S7) according to the procedure of Caporaso et al. (2012). PCR were conducted in a total reaction volume of 20 µl, using 10 µl Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Fisher Scientific Inc., Waltham, MA, USA), 0.6 µL of DMSO, 1 µl of each primer (10µM), 6.4 µl of ultra-pure water and 1 µl of template DNA (i.e. 5ng). PCRs consisted of an initial denaturation step at 98°C for 30 s, followed by 10 cycles of 10 s at 98°C, 30 s at 53°C and 30 s at 72°C, and then by 19 cycles of 10 s at 98°C, 30 s at 48°C and 30 s at 72°C, and ending with a final elongation step of 7 min at 72°C. For each extracted DNA sample, amplicons from two PCRs were pooled together and checked for successful amplification by running 3µL of the PCR product in a 2% agarose gel with GelRed nucleic acid stain (Biotium Inc., Hayward, CA, USA). Blank controls were used in all amplification steps and remained negative throughout the experiments. Amplicons were quantified by fluorometry with the QuBit HS dsDNA kit (Life Technologies, Carlsbad, CA, USA). Approximately 70 ng of amplicons for each sample were pooled and then purified using the QIAquick PCR purification kit (Qiagen, Hilden,

Germany), following the recommendations of the manufacturer. Amplicons were quantified by fluorometry with the QuBit HS dsDNA kit (Life Technologies, Carlsbad, CA, USA). Approximately 70 ng of amplicons for each sample were pooled and then purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany), according to the manufacturercturering to the manufacturerfacturer for each sample were pooled and then purified using the  for each sample Aquick P QIAquick Png and Genotyping Core Facility of the University California Los Angeles. The library quality was verified using a Bioanalyzer Expert 2100 High Sensitivity chip (Agilent Technologies, California) and qPCR was performed to determine cluster density. Paired-end sequencing of the library was performed with the Illumina MiSeq platform using the MiSeq Reagent v3 chemistry (Illumina, San Diego, CA, USA) that enables 300-bp paired-end reads. Amplicon data are available on NCBI Sequence Read Archive (SRA) under project number: PRJNA396681.

***Sequence data processing and protist taxonomic assignment***

Sequence quality was initially examined using FastQC (Andrews et al., 2014). Paired-end reads were then merged with PEAR v0.9.0 (Zhang et al., 2014) using default parameters. Stringent quality filtering was conducted with Usearch v.8 (Edgar, 2010). Combined sequences shorter than 360-bp or with expected errors >0.5 were discarded. The absence of sequencing adapters in the dataset was verified using Trimmomatic version 0.32 (Bolger et al., 2014). Chimeric sequences were detected with Vsearch, v1.0.16 (Rognes et al., 2016) using the Silva 18S database as the reference (release 115) and chimeric sequences were discarded. We then used the QIIME v.1.7 (Caporaso et al., 2010b) sequence preprocessing pipeline for OTU picking and taxonomic annotation. Clustering of OTUs at 97% similarity was performed with the subsampled open reference protocol (Rideout et al., 2014). The SILVA reference dataset (release 115) was used as the reference for OTU picking with

UCLUST v.1.2.22q (Edgar, 2010) and for taxonomy assignments (Quast et al., 2013). Sequences that failed to be aligned to the Silva eukaryotic reference database were discarded from the dataset. We then used the functions split_otu_table_by_taxonomy.py, filter_samples_from_otu_table.py and merge_otu_table.py within QIIME to extract protist sequences from the eukaryotic dataset. Protist dataset included sequences from the following groups: amoebozoans, chlorophytes, centrohelids, cryptophyceans, incertae sedis, SAR, and unicellular opisthokontons. Within the Archaeoplastida supergroup, we filtered out Streptophyta sequences. The unicellular opisthokonts comprised sequences assigned as filasterean, choanomonads, ichthyosporeans, discicristoideans and unclassified unicellular choanoflagellates which did not belong to fungi and metazoa. To test whether protist diversity patterns along the gradient differed between two main functional groups, we classified protist sequences into phototrophic protists (including symbionts) and free-living heterotrophic protists according to del Campo et *al.* (2014). Protist lineages (level D3, SILVA) which might comprise both phototrophic and heterotrophic protists were excluded from this analysis. The phototrophic functional groups comprised chlorophytes, diatoms, eustigmatales, xanthophyceans while the free-living heterotrophic functional group comprised heterotrophic protists as defined above excluding the parasitic protist lineages apicomplexans, ichthyosporeans, and perkinsids. Because our HTES approach is based on extracted DNA, our data might include OTUs derived from extracellular DNA or encysted cells.

***Phylogenetic analyses of unicellular opisthokonts***

To further investigate the phylogenetic affiliation of OTUs assigned as unicellular opisthokonts, OTUs were aligned to the unicellular opisthokont reference alignment using PyNAST with default parameters (Caporaso et al., 2010a) and placed on a reference tree

from del Campo et al. (2015) using the Evolutionary Placement Algorithm (EPA) of RAxML (Berger et al., 2011). The manually-curated unicellular opisthokont reference database used in this study was established by del Campo *et al.* (2013; 2015). It consisted of 828 unicellular opisthokont sequences obtained from all published environmental studies based on clone libraries and the reference tree consisted of 164 unicellular opisthokont sequences clustered at a threshold of 97% similarity. The final tree, including reference sequences and the successfully placed OTUs of this study, was displayed with iTOL (Letunic and Bork, 2007). OTUs successfully placed on the reference tree were subjected to a BLASTN against the whole unicellular opisthokont database from del Campo *et al.* (2015) to evaluate their similarity to published sequences.

***Statistical analyses***

Statistical analysis was performed in R version 3.2.2 (R Development Core Team, 2015), unless otherwise indicated. To evaluate taxon accumulation of protists per 1) sample and 2) ecosystem type, rarefaction curves were constructed using the function rarefy from the R package vegan (Oksanen et al., 2015). Protists, two protist functional groups (phototrophic protists and free-living heterotrophic protists), and the most abundant groups of protists were normalized in vegan to an equal sampling depth with the function rarefy before diversity and community structures analyses (Supporting Information Table S8). To describe the alpha diversity, the species richness (number of OTUs) and Shannon's diversity index (H) were calculated on the rarefied datasets. To test whether the species richness and diversity of protists and other subgroups of protists differed between ecosystems, we used linear mixed effects models, with "ecosystem types" as a fixed factor and sites as a random factor to take into account the non-independency of the replicates within the ecosystem comparison plots. These analyses were conducted using the nlme package in R (Pinheiro and Bates, 2000). To

evaluate differences between ecosystem types, Post hoc tests were performed using the 'glht' function of R package 'multcomp' and corrected for the false discovery rate (Benjamini and Hochberg, 1995). The overall OTUs composition dissimilarity (beta-diversity) was measured using the Sørensen dissimilarity index. Beta-diversity was then partitioned to quantify the nestedness (i.e. OTUs loss) and OTUs turnover (i.e. OTUs replacement) along the ecological gradient using the function beta.multi of the R package betapart (Baselga et al., 2017). A permutational multivariate analysis of variance (PERMANOVA) using Bray–Curtis distance matrices based on square root transformed OTU abundances was performed with 999 permutations using the 'vegan' R package to test if the community composition of each protist dataset differed significantly across the gradient. Pairwise tests were then used to compare differences between ecosystem types with the adjusted *P* values method (Benjamini and Hochberg, 1995), after verifying that there were equal variances between treatment groups using the betadisper tests in vegan. To visualize differences in the composition of protists, free-living heterotrophic protists, and phototrophic protists among ecosystem types, nonmetric multidimensional scaling (NMDS) was conducted with the Bray-Curtis distance as above using the 'metaMDS' function (Oksanen et al., 2015). To evaluate the relationships between selected edaphic and vegetation cover factors on the OTU-based Bray-Curtis similarity matrices of protists, phototrophic protist (PP), and free-living heterotrophic protist (FHP) communities, we used distance-based linear modelling (*DistLM*) (McArdle and Anderson, 2001) after verifying that the relationships between protists, PP or FHP community structure and a broad set of environmental variables were significant with the non-parametric Mantel-type test RELATE. Forward selection along with the adjusted $R^2$ selection criterion were used to identify variables which best explained the changes in community structure. The RELATE test and *DistLM* were performed using PRIMER v7 (Clarke and Gorley, 2006). The selected variables were subsequently used to build a

constrained ordination plot using the best-fitted model in a distance-based redundancy analysis (db-RDA) (Legendre and Anderson, 1999) in vegan. ANOVA permutation tests were used to assess the significance of the individual axes and the overall models. Furthermore, we performed a moving-window analysis (Kent et al., 1997; Carlson et al., 2010) to quantify and test whether the relationships between protist communities and selected variables change along the gradient of drainage and tree productivity. A window width of 12 samples was advanced one sample at the time from the less productive towards the most productive end of the gradient. The first window included 9 samples from the blanket bogs and 3 from bog woodlands while the last window included 3 samples from the bog forests and 9 the zonal forests. Db-RDA, adjusted $R^2$ and permutation tests were used to assess the relationships between protist communities and selected variables (see above) from 25 distinct windows. The number of samples included in each window (12) represented a trade-off between investigating short sections of the gradient and incorporating sufficient samples for the calculations. Finally, we performed an indicator species analysis using the multipatt algorithm in the Indicspecies package (De Caceres and Legendre, 2009) to identify common and abundant protist OTUs (>0.01% of sample) that were significantly associated with one or a combination of ecosystem types.

## Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to PJK (2014-03994), and funding and logistical support from the Tula Foundation and Hakai Institute. TH was supported by the SNSF grant (PA00P3 145374), and JdC by a Marie Curie IOF grant (FP7-PEOPLE-2012-IOF - 331450 CAARL) and a grant to the CMDE from the Tula Foundation. We thank Eric Peterson, Christina Munck and the staff of the Hakai Institute for support for sampling and field work, Allen Banner, Andy

MacKinnon and Sari Saunders for plot establishment and characterization, members of the Keeling lab and Mohn lab, Laura Parfrey and her lab members for thoughtful discussions, Keith Holmes for graphical assistance, and three anonymous reviewers for their constructive criticism and corrections. The authors declare no conflict of interest.

## Supporting information

Supporting Information Fig. S1-S4

Supporting Information Table S1-S8

# References

Alaback, P.B. (1996) Biodiversity Patterns in Relation to Climate: The Coastal Temperate Rainforests of North America. In *Ecological Studies*: Springer pp. 105-133.

Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F., and Wingett, S. (2014) FastQC. A quality control tool for high throughput sequence data. *Babraham Bioinformatics, http://wwwbioinformaticsbabrahamacuk/projects/fastqc/*.

Bamforth, S.S. (2010) Distribution of and insights from soil protozoa of the Olympic coniferous rain forest. *Pedobiologia* **53**: 361-367.

Banner, A., LePage, P., Moran, J., and de Groot, A. (2005) The HyP3 project: Pattern, process and productivity in hypermaritime forests of coastal British Columbia - a synthesis of 7-year results. In, pp. 1-142.

Banner, A., MacKenzie, W., Haeussler, S., Thomson, S., Pojar, J., and Trowbridge, R. (1993) *A field guide to site identification and interpretation for the Prince Rupert forest region*: British Columbia Ministery of Forest, Victoria, BC.

Baselga, A., and Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution* **3**: 808-812.

Baselga, A., Orme, C.D.L., Villeger, S., De Bortoli, J., and Leprieur, F. (2017) Betapart: partitioning beta diversity into turnover and nestedness components, R package version 1.4.

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

Behnke, A., Bunge, J., Barger, K., Breiner, H.W., Alla, V., and Stoeck, T. (2006) Microeukaryote community patterns along an O-2/H2S gradient in a supersulfidic anoxic Fjord (Framvaren, Norway). *Applied and Environmental Microbiology* **72**: 3626-3636.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289-300.

Berger, S.A., Krompass, D., and Stamatakis, A. (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology* **60**: 291-302.

Blandenier, Q., Seppey, C.V.W., Singer, D., Vlimant, M., Simon, A., Duckert, C., and Lara, E. (2017) *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* **64**: 257-265.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Brown, M.W., Spiegel, F.W., and Silberman, J.D. (2009) Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Molecular Biology and Evolution* **26**: 2699-2709.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266-267.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N. et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal* **6**: 1621-1624.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.

Carlson, M.L., Flagstad, L.A., Gillet, F., and Mitchell, E.A.D. (2010) Community development along a proglacial chronosequence: are above-ground and below-ground community structure controlled more by biotic than abiotic factors? *Journal of Ecology* **98**: 1084-1095.

Carpenter, D.N., Bockheim, J.G., and Reich, P.F. (2014) Soils of temperate rainforests of the North American Pacific Coast. *Geoderma* **230**: 250-264.

Clarke, K.R., and Gorley, R.N. (2006) *PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth*.

Coûteaux, M.-M. (1972) Distribution des Thécamoebiens de la litière et de l'humus de deux sols forestiers d'humus brut. *Pedobiologia* **12**: 237-243.

De Caceres, M., and Legendre, P. (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566-3574.

del Campo, J., and Ruiz-Trillo, I. (2013) Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Molecular Biology and Evolution* **30**: 802-805.

del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. (2014) The others: our biased perspective of eukaryotic genomes. *Trends in Ecology & Evolution* **29**: 252-259.

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

Dupont, A.O.C., Griffiths, R.I., Bell, T., and Bass, D. (2016) Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. *Environmental Microbiology* **18**: 2010-2024.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

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

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

Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M.M., Bonkowski, M., and Urich, T. (2015b) Metatranscriptomic census of active protists in soils. *ISME Journal* **9**: 2178-2190.

Geisen, S., Mitchell, E.A.D., Wilkinson, D.M., Adl, S., Bonkowski, M., Brown, M.W. et al. (2017) Soil protistology rebooted: 30 fundamental questions to start with. *Soil Biology & Biochemistry* **111**: 94-103.

Green, R.N. (2014) Reconnaissance level terrestrial ecosystem mapping of priority landscape units of 309 the coast EBM planning area: Phase 3. In. Vancouver, Canada: Blackwell and Associates.

Gremmen, N.J.M., Van De Vijver, B., Frenot, Y., and Lebouvier, M. (2007) Distribution of moss-inhabiting diatoms along an altitudinal gradient at sub-Antarctic Iles Kerguelen. *Antarctic Science* **19**: 17-24.

Grossmann, L., Jensen, M., Heider, D., Jost, S., Glucksman, E., Hartikainen, H. et al. (2016) Protistan community analysis: key findings of a large-scale molecular sampling. *ISME Journal* **10**: 2269-2279.

Harder, C.B., Ronn, R., Brejnrod, A., Bass, D., Abu Al-Soud, W., and Ekelund, F. (2016) Local diversity of heathland Cercozoa explored by in-depth sequencing. *ISME Journal* **10**: 2488-2497.

Heger, T.J., Mitchell, E.A.D., and Leander, B.S. (2013) Holarctic phylogeography of the testate amoeba *Hyalosphenia papilio* (Amoebozoa: Arcellinida) reveals extensive genetic diversity explained more by environment than dispersal limitation. *Molecular Ecology* **22**: 5172-5184.

Heger, T.J., Mitchell, E.A.D., Golemansky, V., Todorov, M., Lara, E., Leander, B.S., and Pawlowski, J. (2010) Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. *Molecular Phylogenetics and Evolution*: 113-122.

Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P., and Keeling, P.J. (2017) Novel freshwater predators reshape holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of animals. *Current Biology*: 2043–2050.

Hertel, L.A., Bayne, C.J., and Loker, E.S. (2002) The symbiont *Capsaspora owczarzaki,* nov gen. nov sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoea. *International Journal for Parasitology* **32**: 1183-1191.

Jassey, V.E.J., Gilbert, D., Binet, P., Toussaint, M.L., and Chiapusio, G. (2011) Effect of a temperature gradient on *Sphagnum fallax* and its associated living microbial communities: a study under controlled conditions. *Canadian Journal of Microbiology* **57**: 226-235.

Jassey, V.E.J., Lamentowicz, L., Robroek, B.J.M., Gazbka, M., Rusinska, A., and Lamentowicz, M. (2014) Plant functional diversity drives niche-size-structure of dominant microbial consumers along a poor to extremely rich fen gradient. *Journal of Ecology* **102**: 1150-1162.

Jassey, V.E.J., Chiapusio, G., Binet, P., Buttler, A., Laggoun-Defarge, F., Delarue, F. et al. (2013) Above- and belowground linkages in *Sphagnum* peatland: climate warming affects plant-microbial interactions. *Global Change Biology* **19**: 811-823.

Jassey, V.E.J., Signarbieux, C., Hattenschwiler, S., Bragazza, L., Buttler, A., Delarue, F. et al. (2015) An unexpected role for mixotrophs in the response of peatland carbon cycling to climate warming. *Scientific Reports* **5**: 10.

Kent, M., Gill, W.J., Weaver, R.E., and Armitage, R.P. (1997) Landscape and plant community boundaries in biogeography. *Progress in Physical Geography* **21**: 315-353.
Koller, R., Robin, C., Bonkowski, M., Ruess, L., and Scheu, S. (2013) Litter quality as driving factor for plant nutrition via grazing of protozoa on soil microorganisms. *FEMS Microbiology Ecology* **85**: 241-250.

Lamentowicz, M., Lamentowicz, L., van der Knaap, W.O., Gabka, M., and Mitchell, E.A.D. (2010) Contrasting species-environment relationships in communities of testate amoebae, bryophytes and vascular plants along the fen-bog gradient. *Microbial Ecology* **59**: 499-510.

Lara, E., Mitchell, E.A.D., Moreira, D., and Garcia, P.L. (2011) Highly diverse and seasonally dynamic protist community in a pristine peat bog. *Protist* **162**: 14-32.

Lawley, B., Ripley, S., Bridge, P., and Convey, P. (2004) Molecular analysis of geographic patterns of eukaryotic diversity in Antarctic soils. *Applied and Environmental Microbiology* **70**: 5963-5972.

Legendre, P., and Anderson, M.J. (1999) Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**: 1-24.

Lentendu, G., Wubet, T., Chatzinotas, A., Wilhelm, C., Buscot, F., and Schlegel, M. (2014) Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Molecular Ecology* **23**: 3341-3355.

Letunic, I., and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127-128.

Lindo, Z., and Gonzalez, A. (2010) The bryosphere: an integral and influential component of the Earth's biosphere. *Ecosystems* **13**: 612-627.

Logares, R., Brate, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K., and Rengefors, K. (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends in Microbiology* **17**: 414-422.

López-Escardo, D., López-García, P., Moreira, D., Ruiz-Trilló, I., and Torruella, G. (2018) *Parvularia atlantis* gen. et sp. nov., a Nucleariid filose amoeba (Holomycota, Opisthokonta). *Journal of Eukaryotic Microbiology*.

López-García, P., Rodriguez-Valera, F., Pedros-Alio, C., and Moreira, D. (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**: 603-607.

Mahé, F., Vargas, C.d., Bass, D., Czech, L., Stamatakis, A., Lara, E. et al. (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nature Ecology and Evolution* **1**: 1-8.

McArdle, B.H., and Anderson, M.J. (2001) Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* **82**: 290-297.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. et al. (2015) Vegan: community ecology package.

Opravilova, V., and Hajek, M. (2006) The variation of testacean assemblages (Rhizopoda) along the complete base-richness gradient in fens: A case study from the Western Carpathians. *Acta Protozoologica* **45**: 191-204.

Page, S.E., Rieley, J.O., and Wüst, R. (2006) Lowland tropical peatlands of Southeast Asia. In: Elsevier, pp. 145-172.

Parfrey, L.W., Walters, W.A., Lauber, C.L., Clemente, J.C., Berg-Lyons, D., Teiling, C. et al. (2014) Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Frontiers in Microbiology* **5**.

Payne, R.J., Creevy, A., Malysheva, E., Ratcliffe, J., Andersen, R., Tsyganov, A.N. et al. (2016) Tree encroachment may lead to functionally-significant changes in peatland testate amoeba communities. *Soil Biology & Biochemistry* **98**: 18-21.

Pinheiro, J.C., and Bates, D.M. (2000) Linear mixed-effects models: basic concepts and examples. In *Mixed-effects models in S and S-PLUS*, pp. 3-56.

Pojar, J., Klinka, K., and Demarchi, D.A. (1991) *Coastal Western Hemlock Zone*.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**: D590-D596.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Version 3.2.2. *R Development Core Team, Vienna, Austria http://wwwR-projectorg*.

Raghukumar, S. (1987) Occurrence of the thraustochytrid, *Corallochytrium- imacisporum* gen. et sp. nov in the coral reef lagoons of the Lakshadweep islands in the Arabian sea. *Botanica Marina* **30**: 83-89.

Ramirez, K.S., Leff, J.W., Barberan, A., Bates, S.T., Betley, J., Crowther, T.W. et al. (2014) Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proceedings of the Royal Society B-Biological Sciences* **281**.

Rideout, J.R., He, Y., Navas-Molina, J.A., Walters, W.A., Ursell, L.K., Gibbons, S.M. et al. (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2**: 25.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahe, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: 22.

Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E.A.D., and Lara, E. (2017) Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. *Soil Biology and Biochemistry* **112**: 68-76.

Shen, C., Liang, W., Shi, Y., Lin, X., Zhang, H., Wu, X. et al. (2014) Contrasting elevational diversity patterns between eukaryotic soil microbes and plants. *Ecology* **95**: 3190-3202.
Shi, Y., Xiang, X.J., Shen, C.C., Chu, H.Y., Neufeld, J.D., Walker, V.K., and Grogan, P. (2015) Vegetation-associated impacts on arctic tundra bacterial and microeukaryotic communities. *Applied and Environmental Microbiology* **81**: 492-501.

Singer, D., Lara, E., Steciow, M.M., Seppey, C.V.W., Paredes, N., Pillonel, A. et al. (2016) High-throughput sequencing reveals diverse oomycete communities in oligotrophic peat bog micro-habitat. *Fungal Ecology* **23**: 42-47.

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

Suga, H., Chen, Z.H., de Mendoza, A., Sebe-Pedros, A., Brown, M.W., Kramer, E. et al. (2013) The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nature Communications* **4**: 1-9.

Tedersoo, L., Bahram, M., Cajthaml, T., Polme, S., Hiiesalu, I., Anslan, S. et al. (2016) Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *ISME Journal* **10**: 346-362.

Thompson, S.D., Nelson, T.A., Giesbrecht, I., Frazer, G., and Saunders, S.C. (2016) Data-driven regionalization of forested and non-forested ecosystems in coastal British Columbia with LiDAR and RapidEye imagery. *Applied Geography* **69**: 35-50.

Tong, S.M. (1997) Heterotrophic flagellates and other protists from Southampton Water, UK. *Ophelia* **47**: 71-131.

Torruella, G., de Mendoza, A., Grau-Bove, X., Anto, M., Chaplin, M.A., del Campo, J. et al. (2015) Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Current Biology* **25**: 2404-2410.

van der Putten, W.H., Bardgett, R.D., Bever, J.D., Bezemer, T.M., Casper, B.B., Fukami, T. et al. (2013) Plant-soil feedbacks: the past, the present and future challenges. *Journal of Ecology* **101**: 265-276.

Wang, T., Hamann, A., Spittlehouse, D., and Carroll, C. (2016) Locally cownscaled and spatially customizable climate data for historical and future periods for North America. *PLOS ONE* **11**.

Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setala, H., van der Putten, W.H., and Wall, D.H. (2004) Ecological linkages between aboveground and belowground biota. *Science* **304**: 1629-1633.

Worley, A.C., Raper, K.B., and Hohl, M. (1979) *Fonticula alba* - new cellular slime-mold (acrasiomycetes). *Mycologia* **71**: 746-760.

Zhang, J.J., Kobert, K., Flouri, T., and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614-620.

## Table and Figure Legends

**Table 1**: Sequential tests from *DistLM* model of environmental predictors of protist, FHP, PP community structures. Significance indicates the addition of the variable significantly

increases the proportion of explained variance in the model. Significance levels: $P<0.001$ '***', $P<0.1$ '**', $P<0.05$ '*'. Abbreviations: see legend figure 6.

**Figure 1**: Sketch maps of the west coast of North America (A) showing the study area on Calvert and Hecate Islands (B), followed by detailed map displaying the locations of the twelve studied ecosystem comparison plots (ECPs). In the figure C, contours of the three investigated watersheds are displayed in red. Geographical coordinates of the ECPs are given in the Supporting Information Table S1. The complex mosaic of ecosystem types of the study area is displayed in the Supporting Information Fig. S1.

**Figure 2:** Taxonomic composition of the eukaryotic and protist OTUs (97) retrieved from soil surface samples from Calvert and Hecate Islands. A) Pie charts display the relative abundance and the number of OTUs of eukaryotes and protists at a high level of taxonomic assignment. B) Relative abundance (in blue) and number of OTUs (in white) of high rank dominant lineages of protists. Within each major group, high rank lineages that are represented by < 3% of sequences are grouped into the category "Other". The taxonomic composition of Centrohelida, Cryptophyceae and unicellular opisthokonts which individually accounts for < 1 % of the total protist sequences is not detailed. C) Rank abundance of the 25 most abundant protist OTUs that contributed to more than 25% of the total sequencing reads and the 0.3 % of the total protist species richness. The names of the taxa were obtained from the lowest taxonomical rank available after taxonomical assignment with the Silva reference database (similarity > 97%).

**Figure 3:** Phylogenetic 18S rDNA tree including unicellular opisthokont sequences from a reference database (del Campo et al. 2015) and this present study. The major unicellular

taxonomic groups found in this study are displayed with different colors. The histograms represent the number (square root transformed) of high-throughput environmental sequencing reads of the 88 unicellular opisthokont OTUs retrieved in this study.

**Figure 4:** OTU richness and diversity of protists, free-living heterotrophic protists and phototrophic protists along a gradient of ecosystem types. Different letters indicate significant differences between ecosystem types ($P < 0.05$).

**Figure 5:** Non-metric multidimensional scaling (NMDS) ordination plots for protists (A), free-living heterotrophic protists (B) and phototrophic protists (C) from four ecosystem types. Different colors and symbols highlight respectively different ecosystem types and watersheds.

**Figure 6:** Distance based redundancy analysis (db-RDAs) with selected edaphic and vegetation variables that explained most of the variability in the communities of protists (a), free-living heterotrophic protists (b), phototrophic protists (C). Samples from blanket bogs are displayed in yellow, bog woodlands in orange, bog forests in brown and zonal forests in red. Abbreviations of the environmental variables: Ca, Calcium (mg/Kg); C/N (Carbon to Nitrogen ratio); Herb, Herb cover (%); K, Potassium (mg/Kg); Litter, Relative abundance of litter content; Na, Sodium (mg/Kg); Non-*Sphagnum* moss, Relative abundance of non-*Sphagnum* moss; S, Sulphur (mg/Kg); *Spha*, Relative abundance of *Sphagnum*; S, Sulphur (mg/Kg); TS, Total Sulphur (%) and Tree, Tree cover (%).

**Figure 7:** Moving-window redundancy analysis illustrating the relationships between protists, edaphic variables (i.e Ca, C/N ratio, total sulphur, pH and sodium), substrate

composition (i.e. relative abundance of *Sphagnum*, relative abundance of litter) and vegetation cover (i.e. tree cover and herb cover) along the gradient of tree productivity and soil drainage. The black, dark grey and light grey lines display respectively edaphic variables, vegetation cover and substrate composition. Solid symbols display significant relationships ($P < 0.05$).The x-axis designates the position of moving window from which the explained variation was calculated along the gradient (see experimental procedures for further details regarding the selection of samples for each window).

# Table

**Table 1**

| | | Variance (%) | *Pseudo-F* values |
|---|---|---|---|
| ***Protists*** | Ca | 19.36 | 8.16*** |
| | Relative abundance of *Sphagnum* | 5.18 | 2.26*** |
| | Herb cover | 5.08 | 2.31*** |
| | Na | 3.44 | 1.59** |
| | Tree cover | 3.31 | 1.56** |
| | C/N | 2.92 | 1.39* |
| | Total S | 3.14 | 1.53** |
| | pH | 3.08 | 1.52** |
| | Relative abundance of litter | 2.72 | 1.36* |
| ***Free-living heterotrophic protists*** | Ca | 20.57 | 8.80*** |
| | Relative abundance of *Sphagnum* | 4.70 | 2.07*** |
| | Herb cover | 4.66 | 2.12*** |
| | Na | 3.54 | 1.65*** |
| | Tree cover | 3.49 | 1.66** |
| | pH | 2.92 | 1.41* |
| | C/N | 2.93 | 1.43* |
| | Total S | 3.33 | 1.65** |
| | Relative abundance of litter | 3.30 | 1.29* |
| ***Phototrophic protists*** | Ca | 19.42 | 8.19*** |
| | K | 5.84 | 2.58*** |
| | Tree cover | 4.35 | 1.97*** |
| | Herb cover | 3.43 | 1.59** |
| | S | 3.34 | 1.57* |
| | Relative abundance of non-*Sphagnum* mosses | 3.04 | 1.45* |

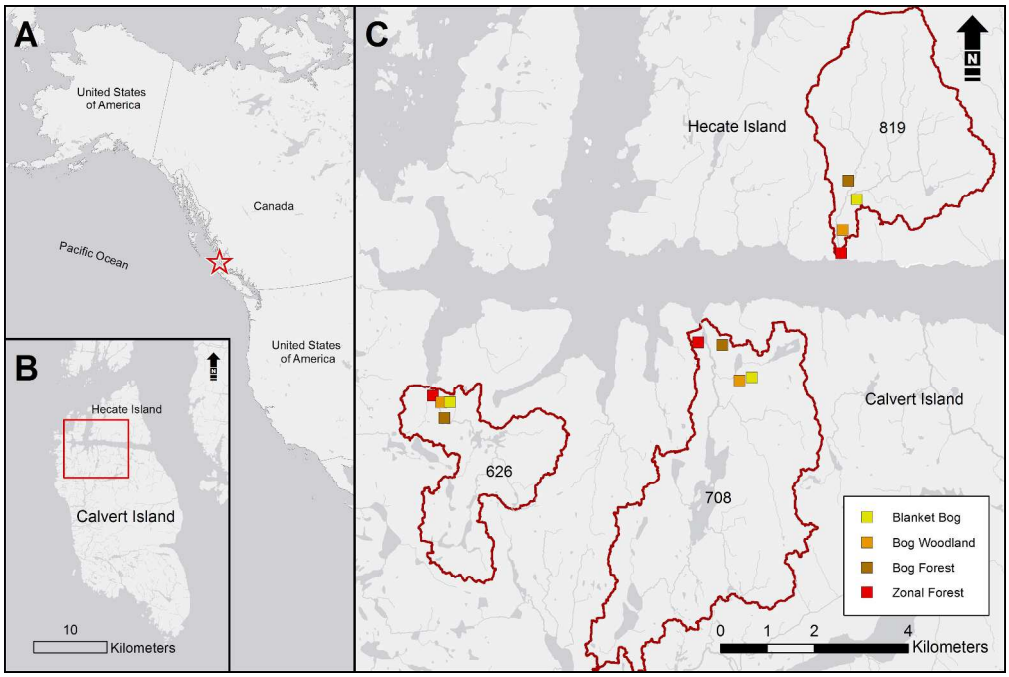


Figure 1: Sketch maps of the west coast of North America (A) showing the study area on Calvert and Hecate Islands (B), followed by detailed map displaying the locations of the twelve studied ecosystem comparison plots (ECPs). In the figure C, contours of the three investigated watersheds are displayed in red. Geographical coordinates of the ECPs are given in the Supporting Information Table S1. The complex mosaic of ecosystem types of the study area is displayed in the Supporting Information Fig. S1.

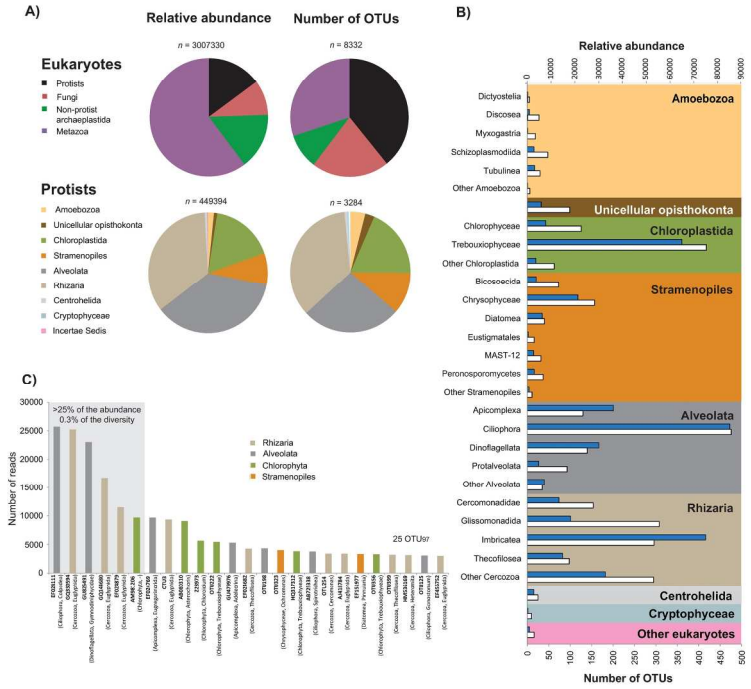


Figure 2: Taxonomic composition of the eukaryotic and protist OTUs (97) retrieved from soil surface samples from Calvert and Hecate Islands. A) Pie charts display the relative abundance and the number of OTUs of eukaryotes and protists at a high level of taxonomic assignment. B) Relative abundance (in blue) and number of OTUs (in white) of high rank dominant lineages of protists. Within each major group, high rank lineages that are represented by < 3% of sequences are grouped into the category "Other". The taxonomic composition of Centrohelida, Cryptophyceae and unicellular opisthokonts which individually accounts for < 1 % of the total protist sequences is not detailed. C) Rank abundance of the 25 most abundant protist OTUs that contributed to more than 25% of the total sequencing reads and the 0.3 % of the total protist species richness. The names of the taxa were obtained from the lowest taxonomical rank available after taxonomical assignment with the Silva reference database (similarity > 97%).

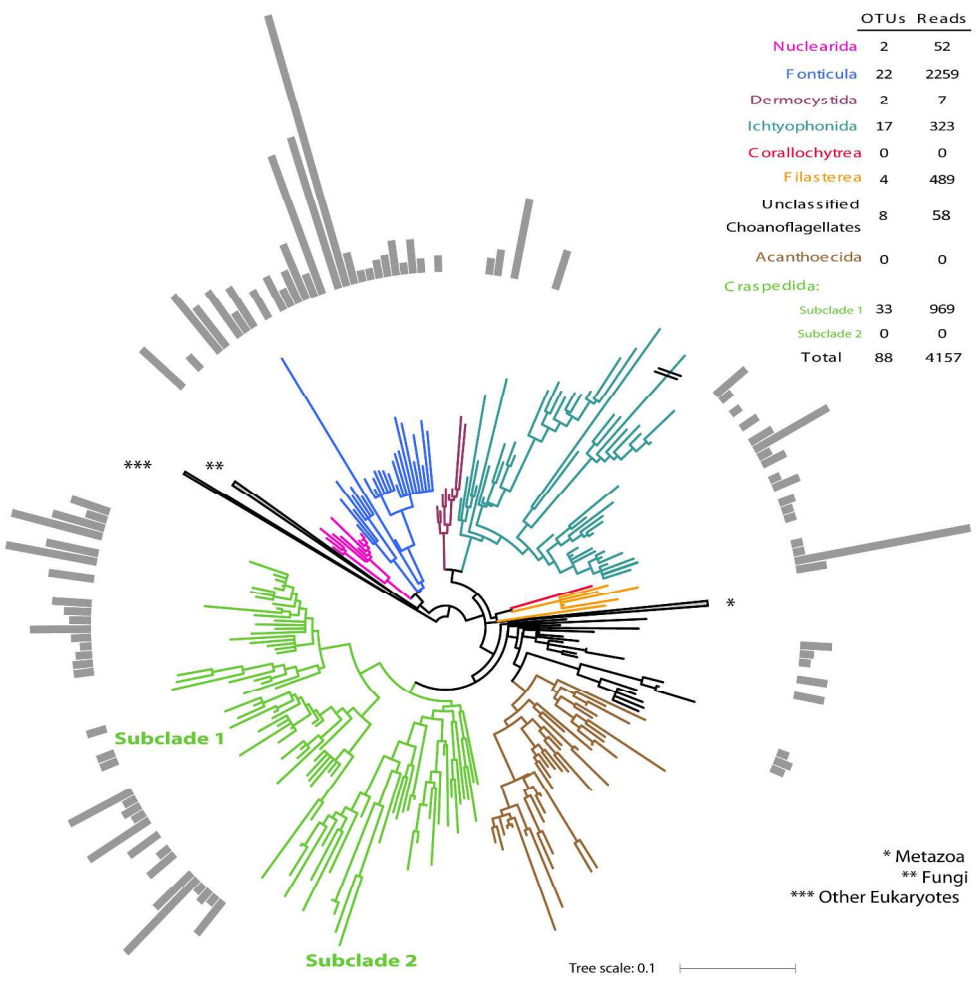


Figure 3: Phylogenetic 18S rDNA tree including unicellular opisthokont sequences from a reference database (del Campo et al. 2015) and this present study. The major unicellular taxonomic groups found in this study are displayed with different colors. The histograms represent the number (square root transformed) of high-throughput environmental sequencing reads of the 88 unicellular opisthokont OTUs retrieved in this study.

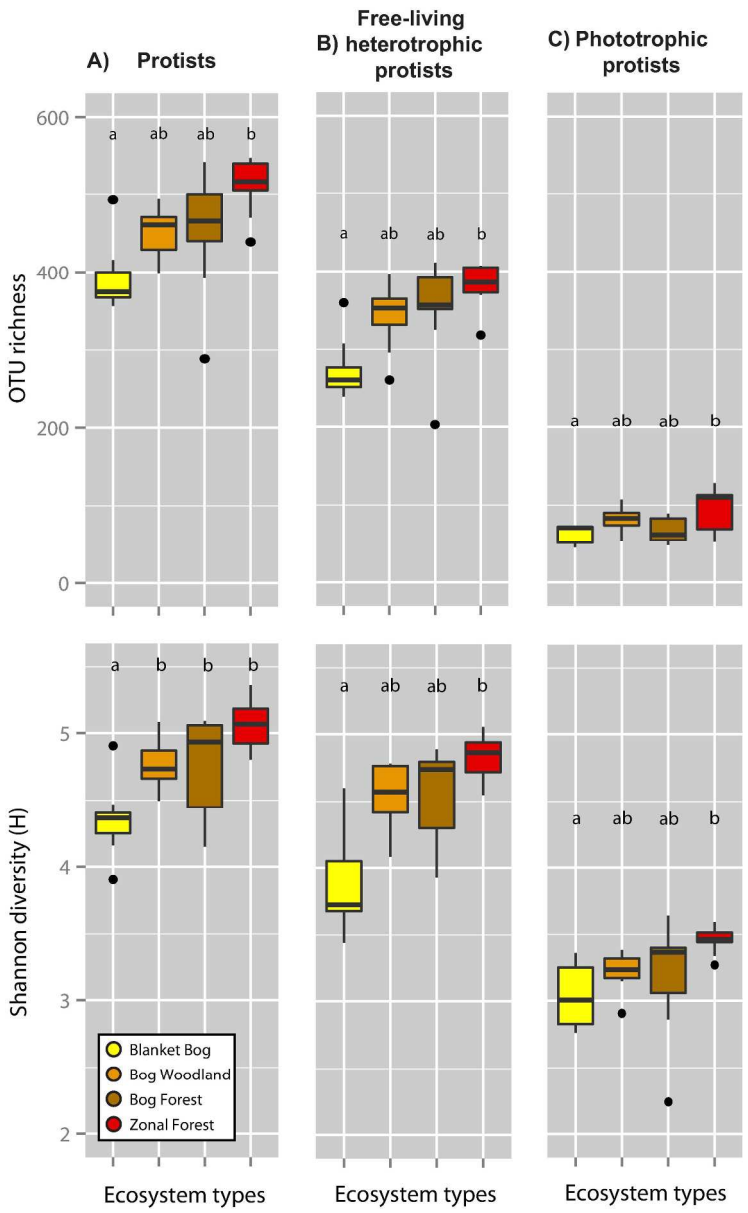


Figure 4: OTU richness and diversity of protists, free-living heterotrophic protists and phototrophic protists along a gradient of ecosystem types. Different letters indicate significant differences between ecosystem types ($P < 0.05$).

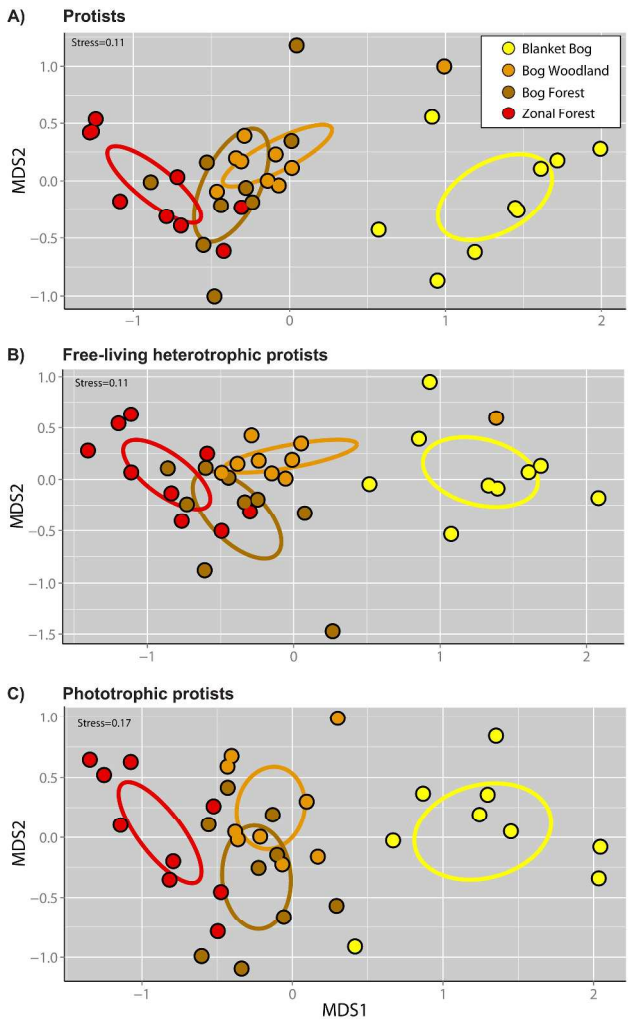


Figure 5: Non-metric multidimensional scaling (NMDS) ordination plots for protists (A), free-living heterotrophic protists (B) and phototrophic protists (C) from four ecosystem types. Different colors and symbols highlight respectively different ecosystem types and watersheds.

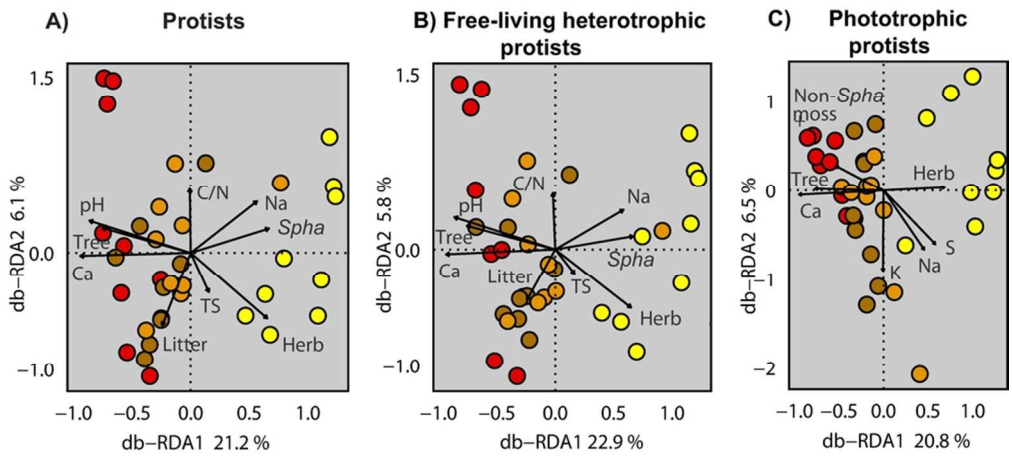


Figure 6: Distance based redundancy analysis (db-RDAs) with selected edaphic and vegetation variables that explained most of the variability in the communities of protists (a), free-living heterotrophic protists (b), phototrophic protists (C). Samples from blanket bogs are displayed in yellow, bog woodlands in orange, bog forests in brown and zonal forests in red. Abbreviations of the environmental variables: Ca, Calcium (mg/Kg); C/N (Carbon to Nitrogen ratio); Herb, Herb cover (%); K, Potassium (mg/Kg); Litter, Relative abundance of litter content; Na, Sodium (mg/Kg); Non-Sphagnum moss, Relative abundance of non-Sphagnum moss; S, Sulphur (mg/Kg); Spha, Relative abundance of Sphagnum; S, Sulphur (mg/Kg); TS, Total Sulphur (%) and Tree, Tree cover (%).

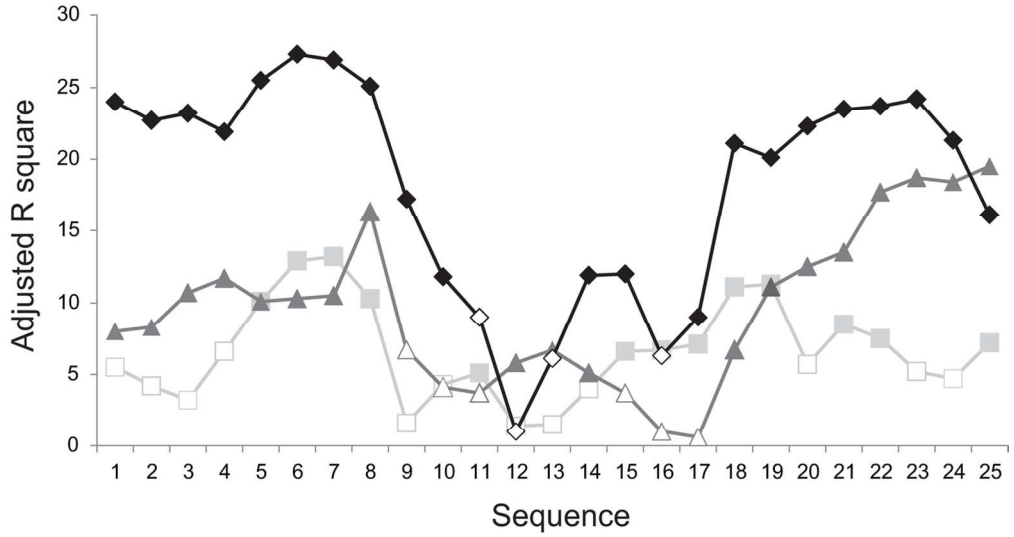


Figure 7: Moving-window redundancy analysis illustrating the relationships between protists, edaphic variables (i.e Ca, C/N ratio, total sulphur, pH and sodium), substrate composition (i.e. relative abundance of Sphagnum, relative abundance of litter) and vegetation cover (i.e. tree cover and herb cover) along the gradient of tree productivity and soil drainage. The black, dark grey and light grey lines display respectively edaphic variables, vegetation cover and substrate composition. Solid symbols display significant relationships ($P < 0.05$).The x-axis designates the position of moving window from which the explained variation was calculated along the gradient (see experimental procedures for further details regarding the selection of samples for each window).