



Contents lists available at ScienceDirect

Environment International

journal homepage: [www.elsevier.com/locate/envint](http://www.elsevier.com/locate/envint)

## Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems

David Singer<sup>a,b,c,\*</sup>, Christophe V.W. Seppey<sup>a,d,\*</sup>, Guillaume Lentendu<sup>a</sup>, Micah Dunthorn<sup>e,f</sup>, David Bass<sup>g</sup>, Lassâad Belbahri<sup>a</sup>, Quentin Blandenier<sup>a,h</sup>, Didier Debroas<sup>i</sup>, G. Arjen de Groot<sup>j</sup>, Colombar de Vargas<sup>k,l</sup>, Isabelle Domaizon<sup>m</sup>, Clément Duckert<sup>a</sup>, Irina Izaguirre<sup>n</sup>, Isabelle Koenig<sup>a</sup>, Gabriela Mataloni<sup>o</sup>, M. Romina Schiaffino<sup>p</sup>, Edward A.D. Mitchell<sup>a,q</sup>, Stefan Geisen<sup>r,1,\*</sup>, Enrique Lara<sup>a,h,\*</sup>

<sup>a</sup> Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

<sup>b</sup> Department of Zoology, Institute of Biosciences, University of São Paulo, 05508-090, Brazil

<sup>c</sup> UMR CNRS 6112 LPG-BIAF, Université d'Angers, Angers Cedex 1, France

<sup>d</sup> Department of Arctic and Marine Biology, University of Tromsø, Framstredet 39, 9019 Tromsø, Norway

<sup>e</sup> Department of Eukaryotic Microbiology, University of Duisburg-Essen, D-45141 Essen, Germany

<sup>f</sup> Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, D-45141 Essen, Germany

<sup>g</sup> Department of Life Sciences, The Natural History Museum, London, UK

<sup>h</sup> Real Jardín Botánico de Madrid, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain

<sup>i</sup> Université Clermont Auvergne, CNRS, Laboratoire Microorganismes: Genome et Environnement, F-63000 Clermont-Ferrand, France

<sup>j</sup> Wageningen Environmental Research, Wageningen UR, P.O. Box 47, 6700 AA Wageningen, the Netherlands

<sup>k</sup> Sorbonne Université, CNRS, Station Biologique de Roscoff, UMR 7144, ECOMAP, 29680 Roscoff, France

<sup>l</sup> Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/GOSEE, 3 rue Michel-Ange, 75016 Paris, France

<sup>m</sup> Université Savoie Mont Blanc, INRAE, CARRTEL, 74200 Thonon-les-Bains, France

<sup>n</sup> Departamento de Ecología, Genética y Evolución, IEGEBA (UBA-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

<sup>o</sup> Instituto de Investigación e Ingeniería Ambiental (IIIA), CONICET, Universidad Nacional de San Martín, Provincia de Buenos Aires 1650, Argentina

<sup>p</sup> Universidad Nacional del Noroeste de la Provincia de Buenos Aires, Centro de Investigaciones y transferencia del Noroeste de la Provincia de Buenos Aires (CITNOBA) – UNNOBA-UNSA-CONICET, Junín 6000, Argentina

<sup>q</sup> Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

<sup>r</sup> Laboratory of Nematology, Wageningen University and Research, Droevendaalsesteeg 1, 6708PB Wageningen, the Netherlands

### ARTICLE INFO

Handling Editor: Frederic Coulon

#### Keywords:

Terrestrial systems  
Metabarcoding  
Microbial eukaryotes  
Taxonomic and functional diversity  
Ocean

### ABSTRACT

Protists dominate eukaryotic diversity and play key functional roles in all ecosystems, particularly by catalyzing carbon and nutrient cycling. To date, however, a comparative analysis of their taxonomic and functional diversity that compares the major ecosystems on Earth (soil, freshwater and marine systems) is missing. Here, we present a comparison of protist diversity based on standardized high throughput 18S rRNA gene sequencing of soil, freshwater and marine environmental DNA. Soil and freshwater protist communities were more similar to each other than to marine protist communities, with virtually no overlap of Operational Taxonomic Units (OTUs) between terrestrial and marine habitats. Soil protists showed higher  $\gamma$  diversity than aquatic samples. Differences in taxonomic composition of the communities led to changes in a functional diversity among ecosystems, as expressed in relative abundance of consumers, phototrophs and parasites. Phototrophs (eukaryotic algae) dominated freshwater systems (49% of the sequences) and consumers soil and marine ecosystems (59% and 48%, respectively). The individual functional groups were composed of ecosystem-specific taxonomic groups.

\* Corresponding authors at: UMR CNRS 6112 LPG-BIAF, Université d'Angers, Angers Cedex 1, France (D. Singer). Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland (C.V.W. Seppey). Laboratory of Nematology, Wageningen University and Research, Droevendaalsesteeg 1, 6708PB Wageningen, the Netherlands (S. Geisen). Real Jardín Botánico de Madrid, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain (E. Lara).

E-mail addresses: [david.singer@univ-angers.fr](mailto:david.singer@univ-angers.fr), [david.singer.bio@outlook.com](mailto:david.singer.bio@outlook.com) (D. Singer), [christophe.seppey@uit.no](mailto:christophe.seppey@uit.no) (C.V.W. Seppey), [stefan.geisen@wur.nl](mailto:stefan.geisen@wur.nl) (S. Geisen), [enrique.lara@rjb.csic.es](mailto:enrique.lara@rjb.csic.es) (E. Lara).

<sup>1</sup> These authors contributed equally to the paper.

<https://doi.org/10.1016/j.envint.2020.106262>

Received 7 May 2020; Received in revised form 31 October 2020; Accepted 2 November 2020

0160-4120/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Parasites were equally common in all ecosystems, yet, terrestrial systems hosted more OTUs assigned to parasites of macro-organisms while aquatic systems contained mostly microbial parasitoids. Together, we show biogeographic patterns of protist diversity across major ecosystems on Earth, preparing the way for more focused studies that will help understanding the multiple roles of protists in the biosphere.

## 1. Introduction

Assessing patterns of biodiversity — how they vary among taxonomic groups, biomes and ecosystems — is an old (Wallace, 1876) but still relevant question in biology (Gaston, 2000; Mora et al., 2011; Tittensor et al., 2010; Wall et al., 2001) that provides a starting point to understand ecosystem functioning and their associated services provided by the organisms living there (Cotterill et al., 2007). Most diversity on Earth is microbial, and protists seem to form a substantial part of this diversity (Adl et al., 2019). Estimating protist diversity is, however, challenging because protists are mostly small and inconspicuous. Environmental DNA-based methods have revealed a high and largely unknown taxonomic diversity of both prokaryotic (Delgado-Baquerizo et al., 2018) and eukaryotic microorganisms (de Vargas et al., 2015; Tedersoo et al., 2014).

Protists play critical roles in Earth's systems; for instance marine photosynthetic protists (the eukaryotic part of phytoplankton) fix as much carbon as all terrestrial plants together (Falkowski et al., 1998). Still heterotrophic protists, including many uncharacterized parasitic and mutualistic symbionts, have recently been shown to be even more diverse and abundant than their phototrophic counterparts (de Vargas et al., 2015; Lima-Mendez et al., 2015; Seeleuthner et al., 2018). Freshwater systems, with their wide variation in physical and chemical properties may host drastically different communities (Boenigk et al., 2018; Debroas et al., 2017), and soil protists may champion biodiversity due to the complex and highly dynamic distribution of water in soil pores that creates extremely heterogeneous environments in time and space. This constant change of conditions promotes the activity of only a part of the microbial community, while the rest remains inactive. This mechanism has been suggested to be key in promoting high microbial diversity in prokaryotes (Tecon and Or, 2017), and similar mechanisms have also been suggested for protists (Velasco-González et al., 2020).

Pioneering studies of soil protists diversity encompassing limited geographical areas and soil types, have revealed an extreme level of an unknown diversity. Indeed as shown in three neotropical forests, only 8.1% of all sequences had  $\geq 95\%$  similarity with known references sequences (Mahé et al., 2017). How these organisms relate to those encountered in the better-known aquatic systems can help in better understanding the diversity and functioning of soil protists. Thus, there is a need for studies integrating as many ecosystems as possible. Of particular relevance to environmental microbiology is knowing which groups contribute to the taxonomic and functional diversity in each ecosystem, how do overall diversity compare and how variable it is in each ecosystem (Geisen et al., 2017).

This study compares the diversity of protists across the marine and freshwater sunlit and top soil ecosystems. Using both published and new data from the same genetic marker and sequencing technology (Illumina high-throughput sequencing of the v9 region of the 18S rRNA gene), our first aim was to compare the  $\alpha$ ,  $\beta$  and  $\gamma$  diversity among the three ecosystems for the overall protists community and for 55 dominant taxa. We expected the entire protist diversity ( $\gamma$ ) to be highest in soil due to the highest habitat niche space. Furthermore, we expected the higher temporal and spatial heterogeneity in freshwater and soil environments to increase protist  $\beta$  diversity compared to marine protists. Our second major hypothesis was that the taxonomic composition of protist communities was most dissimilar in marine samples due to the salinity barrier (Logares et al., 2009). In contrast, we tested the hypothesis that the functional composition of protists is most dissimilar in soils due to the dominance of lesser known consumers, whereas we expected

phototrophs to dominate aquatic environments.

## 2. Material and methods

### 2.1. Environmental samples

In order to compare the protist diversity in the main ecosystems that compose the earth biosphere, we investigated the diversity of marine plankton, freshwater plankton and soil. Our meta-analysis took advantage of published data as well as new samples to complete a dataset composed of 122 sampling sites: 28 marine plankton, 21 freshwater plankton and 73 soil samples (Table G.1–2). Marine plankton samples were retrieved from the Tara Oceans project (<http://taraoceans.sb-roscoff.fr/EukDiv/>) and covered two depths (surface and Deep-Chlorophyll Maximum: DCM) from the world tropical to temperate oceans. For each sample, 80 to 100 L (until filters were clogged with biomass) of seawater were collected and the plankton was deposited on 0.8  $\mu\text{m}$  mesh size filters. The filter membranes were preserved into liquid nitrogen until DNA extraction (de Vargas et al., 2015). In freshwater samples (ca. 5 L of surface water, until filters were clogged), plankton was deposited on 0.2  $\mu\text{m}$  polycarbonate filter. The filter-membranes with concentrated cells were cut into ca. 1  $\text{mm}^2$  pieces, preserved in a DNA preservation buffer (Lifeguard, MoBio Carlsbad CA, USA) and kept at a temperature below  $-20^\circ\text{C}$  prior to DNA extraction. Soil samples consisted of ca. 2 g of the upper organic horizon (0–5 cm). Samples were taken in sterile conditions and kept in DNA preservation buffer (Lifeguard, MoBio Carlsbad CA, USA). No sample was kept over one month before DNA extraction, which ensures the stability of the microbial communities.

### 2.2. DNA extraction, PCR assays, and high-throughput sequencing

Soil and freshwater DNA was extracted with the MoBio PowerSoil extraction kit (Carlsbad, CA, USA) according to the manufacturer instructions. This approach as already been used in several publications in freshwater (Lara et al., 2015; Schiaffino et al., 2016), soils (Singer et al., 2020; Seppey et al., 2017) and marine samples (de Vargas et al., 2015; Lara et al., 2009). Marine plankton DNA was extracted with the DNA Elution buffer kit (Macherey-Nagel, KG, USA) (<http://taraoceans.sb-roscoff.fr/EukDiv/>). The SSU rRNA V9 fragments were PCR amplified using the eukaryotic primers 1380f/1510r (Amaral-Zettler et al., 2009), according to the protocols described in (Singer et al., 2016) for soil and freshwater samples, and the universal primers 1389f/1510r (Amaral-Zettler et al., 2009; de Vargas et al., 2015) for marine plankton. The minor difference in primer sets is not expected to strongly impact the taxonomic distribution between marine and freshwater/soils (Geisen et al., 2019). The amplicons were sequenced with a HiSeq Illumina 2000 and 2500 sequencer for both soils and freshwater, and with an Illumina Genome Analyzer for marine samples. The sequencing reads are available through the projects PRJEB6609 (marine samples) and PRJEB41211 for the other samples on the European Nucleotide Archive of the European Molecular Biology Laboratory.

### 2.3. Generation of protist operational taxonomic units (OTUs)

Environmental rRNA SSU V9 reads from all samples were first merged with the program flash v1.2.11 (Magoc and Salzberg, 2011) and trimmed to remove the eukaryotic primers with a custom script ([https://github.com/cseppey/bin\\_src\\_my\\_prog/blob/master/perl/trim\\_prime](https://github.com/cseppey/bin_src_my_prog/blob/master/perl/trim_prime)

r.pl). Sequences from all samples were then dereplicated with a custom script ([https://github.com/csepppey/bin\\_src\\_my\\_prog/blob/master/cpp/derep.cpp](https://github.com/csepppey/bin_src_my_prog/blob/master/cpp/derep.cpp)) prior to OTU clustering performed with swarm v2.1.13 with the fastidious option (Mahé et al., 2015). OTUs were finally assigned using the global alignment option on vsearch v2.4.4 (Rognes et al., 2016) against the PR<sup>2</sup> database v4.12.0 (Guillou et al., 2012). Chimerical sequences were removed by comparing the reads within the dataset with the program vsearch v2.4.4 (Rognes et al., 2016).

Certain OTUs were then sorted out according to three criteria, as originating from putative non-protists and macroscopic organisms. 1) OTUs belonging to Metazoa, Embryophyceae, Fungi and Rhodophyta were removed, as well as 2) the ones having a sequence with a length below 100 nucleotides, and 3) the ones having a percentage of identity below 60% with the database as putative prokaryotes.

#### 2.4. Functional assignment of OTUs

We assigned OTUs to three major functional groups, based on their taxonomic affiliation according to three main trophic modes: Phototrophic (Archaeplastida, Ochrophyta), Parasitic (Apicomplexa, Ichtyosporidia, MALV, Peronosporomycetes, Syndiniales), Consumers (Ciliophora, Rhizaria, non-Fungi Opisthokonta, Amoebozoa, non-Ochrophyta and non-Peronosporomycetes Stramenopiles) (Table F.1).

We considered as phototrophic any organism that used photosynthesis, independently of the capacity of the organism to also retrieve carbon through heterotrophy (mixotrophy). The organisms that acquired food through phagocytosis were labelled consumers. We decided to consider as parasite any symbiont reducing the fitness of its host, independently if the host is killed by its symbiont (parasitoid) or not. We also consider all Peronosporomycetes as parasites because parasitism is considered as the ancestral condition of the group (Misner et al., 2014), even though many of them often live as saprotrophs and behave as facultative parasites.

We verified manually the precise phylogenetic affiliation of all Dinophyceae OTUs to infer their function. However, 55% of dinokaryotic OTUs were still assigned to “unknown Dinophyceae” which prevented functional assignment. These organisms may range functionally from phototrophic to strictly heterotrophic consumers. In marine systems, it has been estimated that 58% of species are strictly consumers while 42% perform photosynthesis. In freshwater systems, only 12% are consumers while 88% are phototrophic (Gómez, 2012). As phototrophy is a trait that changes rapidly in dinokaryote evolution (Saldarriaga et al., 2004), we considered that these numbers represented both the percentage of OTUs and of reads in both marine and freshwater systems. Soil dinoflagellates have been poorly described, but our results show more similarities with freshwater communities than with marine samples; therefore, we considered also that 12% of soil dinoflagellates were consumers (Fig. A.1).

Chrysophyceae present different nutrient acquisition strategies ranging from phototrophic to consumers (Boenigk et al., 2005). Recent works suggest that Chrysophyceae are ancestrally phototrophic, some groups or species having lost photosynthetic ability later during their evolutionary history (Dorrell et al., 2019). We considered as consumers those OTUs assigned to *Chromulinospumella*, *Segregatospumella*, *Oikomonas* (Clade B2), *Acrispumella*, *Apoikia*, *Edaphospumella*, *Kephyrion*, *Spumella*, and *Poteriospumella* (Clade C), as well as the Clade F (*Paraphysomonas* and allied genera). The rest were considered as phototrophic. As for the Dinophyceae, many Chrysophyceae OTUs were not functionally assignable (44% of the OTUs of the first bootstrap). To our knowledge, there is no information on the consumer/phototroph ratio for this clade as it is for Dinophyceae in marine and freshwaters. In order to include the functionally unknown Chrysophyceae, we calculated the ratio consumer/phototroph for the functionally assigned OTUs in each sample. Then, we pooled reads and OTUs number of functionally unclassified Chrysophyceae according to the ratio consumer/phototroph.

Our sampling design did not allow us to assess temporal patterns

such as algal blooms, increasing the proportion of phototroph, parasites developing on the algae or consumers feeding on the soluble organic matter left after the parasitic infection. In addition, certain groups of species can switch rapidly between different trophic lifestyles. Nevertheless, the overall signal can provide an insight of the importance of each function in the different ecosystems.

#### 2.5. Statistical analyses

In order to avoid bias related to 1) the number of samples per ecosystem and 2) the sequencing depth, we ran the analyses on 100 bootstraps composed of 60 samples (20 samples per ecosystems) of 10'000 reads each. The figures were selected from a bootstrap best representing the overall trend among the 100 bootstraps (Annex H1-5). We acknowledge that our conservative sequence cutoff to 10'000 reads per sample is far from complete taxon sampling. Yet, we here did not aim to obtain complete species inventories but rather to enable reliable community comparisons of the numerically dominant taxa across ecosystems.

We estimated the  $\gamma$  diversity of each ecosystem with the bootstrap estimator of species richness (BES) and a species accumulation curve (SAC) (functions `specpool` and `specaccum`, package `vegan` v. 2.5–2 (Oksanen et al., 2018)). We also estimated the  $\alpha$  diversity (Shannon index, function `diversity`, package `vegan` v. 2.5–2, (Oksanen et al., 2018)) of each of the 55 major lineages that represent the bulk of the known protist diversity in each type of ecosystem, respectively, following the same procedure as described above.

We assessed the similarity patterns among protist communities ( $\beta$ -diversity) by non-metric multidimensional scaling (NMDS). NMDS was calculated on Bray-Curtis dissimilarities retrieved from the sequence relative abundance. The significance of differences between pairs of ecosystems were measured by permutation tests (10'000 permutations; functions `vegdist`, `metaMDS` and `envfit`, package `vegan` v. 2.5–2 (Oksanen et al., 2018)). To estimate which ecosystems were showing the highest  $\beta$  diversity, we compared the distributions of Bray-Curtis distances within ecosystem.

We finally tested for differences among ecosystems for  $\alpha$  and  $\beta$  diversity indices by pairwise tests for multiple comparison of mean rank sums (Nemenyi test,  $P < 0.05$ ; function `posthoc.kruskal.nemenyi.test`, package `PMCMR` v. 4.3 (Pohlert, 2014)) in each bootstrap. We also use this approach to test the differences between ecosystems for the log of the number of reads, the log of the richness and the Shannon diversity within each of the 55 dominant taxonomic groups, as well as for the functional groups relative abundance.

#### 2.6. Dataset consistency

This study is, to a certain extent, a meta-analysis because part of the data (concretely, the marine data) was derived from other studies. Thus, most samples were obtained with distinct methods. Thus, despite targeting the same rRNA region (V9) and using the same sequencing technology (Illumina sequencing) the sampling and laboratory procedures used were not homogeneous, thus potentially producing biases. We therefore tested how robust our data was to methodological biases, including (1) different filtration and sample preservation protocols (2) use of different extraction kits and (3) use of different primer sets.

- (1). Pore filter sizes varied between freshwater and marine samples. In both cases, filter mesh sizes are below 1  $\mu\text{m}$ , which corresponds to the size of the smallest eukaryotes known, such as for instance *Ostreococcus tauri* (Chrétiennot-Dinet et al., 1995). We are confident, therefore, that we retrieved all eukaryotes in the water column, and therefore are comparable to soil in the range of sizes that were considered.
- (2). Extraction kits differed between freshwater/soil and marine samples. While it has been shown that slightly different protist

community profiles were retrieved by using commercial and “homemade” protocols, the differences concerned only particular groups (Santos et al., 2017). Given the differences that have been shown between marine and freshwater/soil communities, it is difficult to evaluate how far this bias affects the big picture presented here as retrieved with two commercial kits, but differences are most likely minor.

- (3). Primer sets differed between freshwater/soil and marine samples. However, these two sets amplify the same spectrum of the eukaryotic diversity (Geisen et al., 2019). We tested this assumption *in silico* from data that used the primers on the same samples (Amaral-Zettler et al., 2009; Stoeck et al., 2009) to demonstrate that biases produced by different primer sets were negligible (Appendix I).

### 3. Results and discussion

#### 3.1. Communities richness: $\alpha$ and $\gamma$ diversity

As expected, protist  $\gamma$  diversity predicted by BES was higher for soils (16'337 OTUs  $\pm$  892 SE) than for the two aquatic ecosystems (freshwater: 11'490 OTUs  $\pm$  756, marine water: 12'540 OTUs  $\pm$  689) (Fig. 1A, Table H.1). Gamma diversity was highest in soils in 26.3  $\pm$  0.1 of the taxonomic groups, 11.0  $\pm$  0.1 in marine samples and 11.7  $\pm$  0.1 in freshwater (Fig. 2, B.1). Even though OTU richness did not reach saturation, most of the SAC trends clearly discriminated the gamma diversity of the three ecosystems; trends that would likely stand with a deeper sampling effort (Fig. B.1).

Freshwater had the lowest  $\alpha$  diversity (3.68  $\pm$  0.18 SE), while soils and marine waters did show significantly difference, soil being slightly

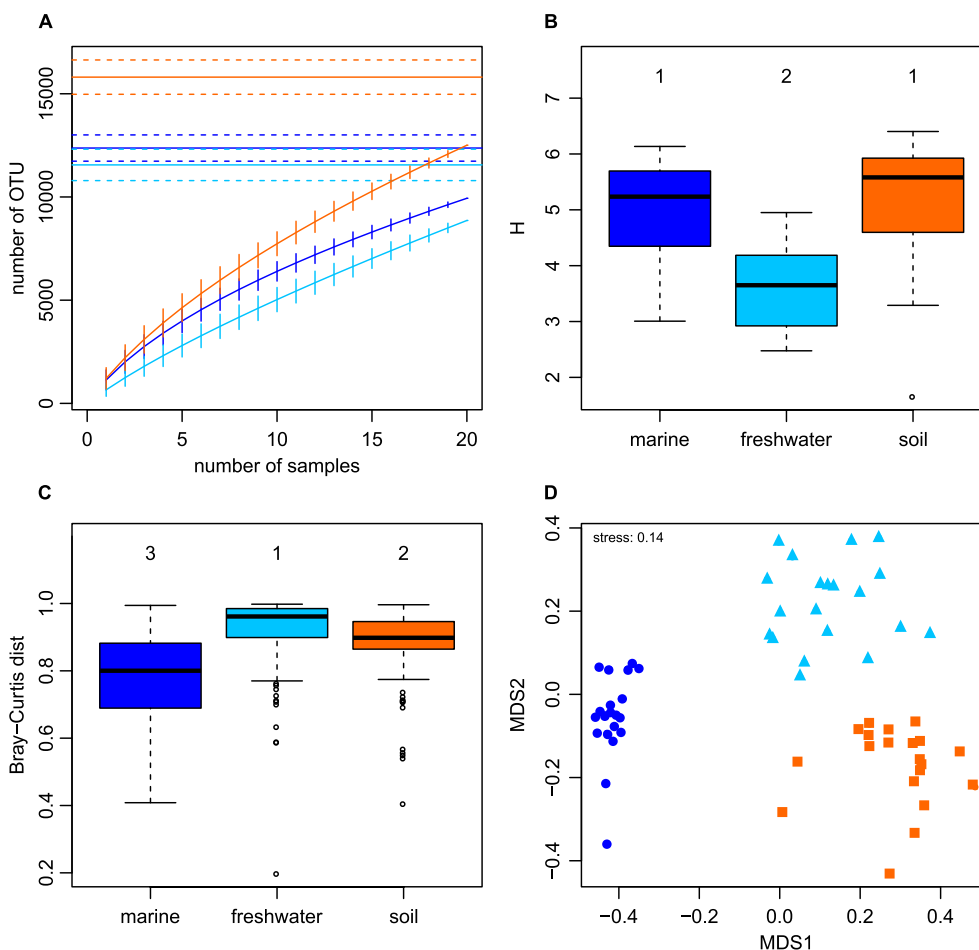
more diverse (respectively, 5.25  $\pm$  0.22 SE and 5.05  $\pm$  0.08 SE) (Fig. 1B, Table H.1).

In line, 20.67  $\pm$  0.28 out of 55 broad taxonomic groups targeted showed significantly higher  $\alpha$  diversity in soils while 9.15  $\pm$  0.8 were more diverse in marine samples and 0.90  $\pm$  0.05 were more diverse in freshwater (Fig. B.2). This probably higher  $\alpha$  diversity in soils followed our prediction based on the co-existence of several dormant and active organisms that differ in their ecological requirements but reactivate under suitable conditions.

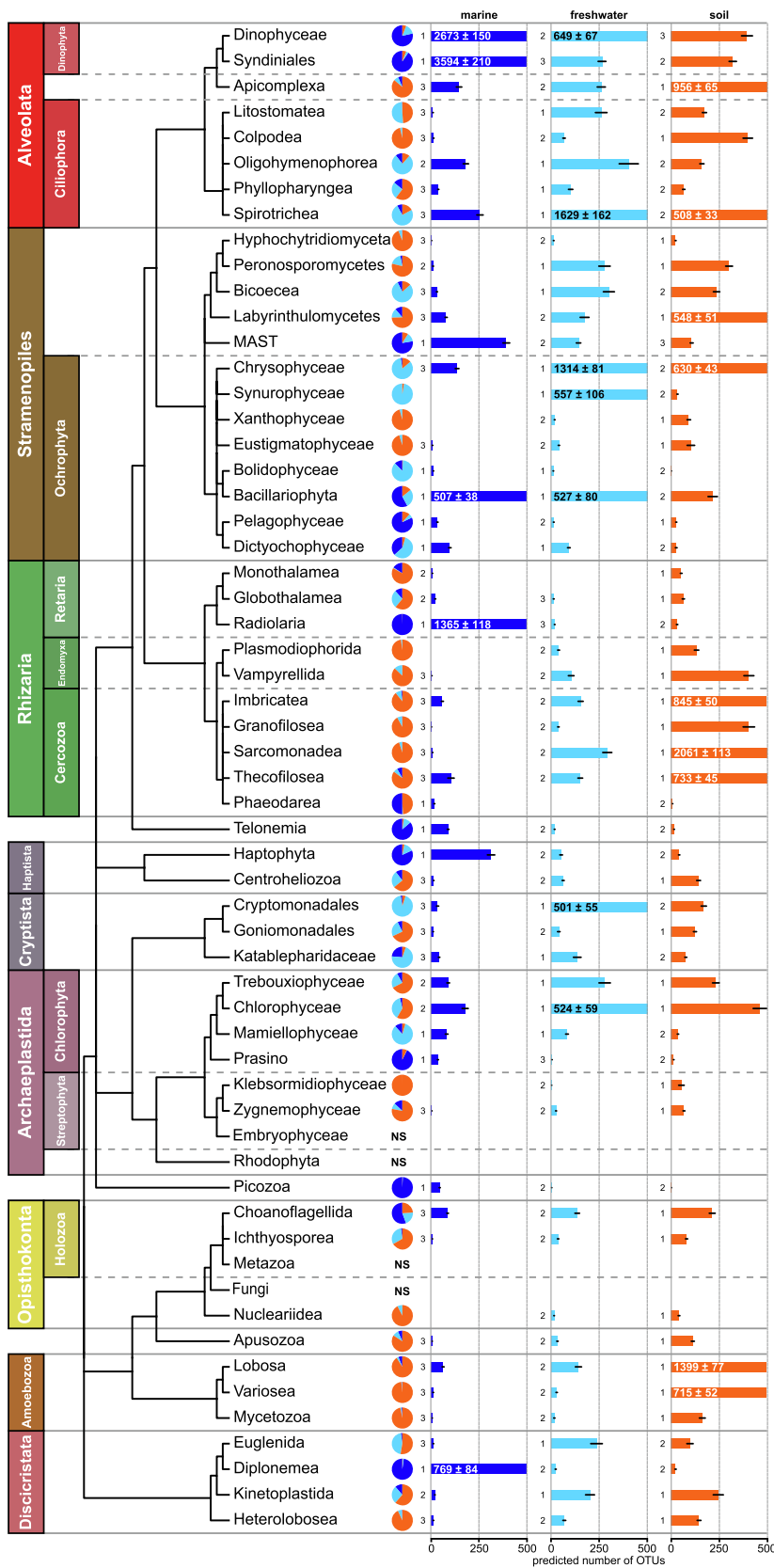
The high  $\alpha$  and  $\gamma$  diversity in soils contrast with the low number of described soil protist species (ca. 21'000 species (Mora et al., 2011)), while estimations suggest the existence of millions (Geisen et al., 2018). We compared the level of knowledge for all three ecosystems based on the percentage of identity between the OTU reads and the best match with the PR<sup>2</sup> eukaryotic ribosomal database (Guillou et al., 2012) (Fig. C.1, Table H.2). Soil OTUs had, on average, a significantly lower percentage of similarity with known sequences (87.28%  $\pm$  0.08 SE) than freshwater systems OTUs (90.56%  $\pm$  0.08 SE), with the best percentage of identity with the database for the OTUs found in marine ecosystems (91.39%  $\pm$  0.09 SE, Nemenyi test:  $P < 0.05$ ).

#### 3.2. Community heterogeneity; $\beta$ diversity

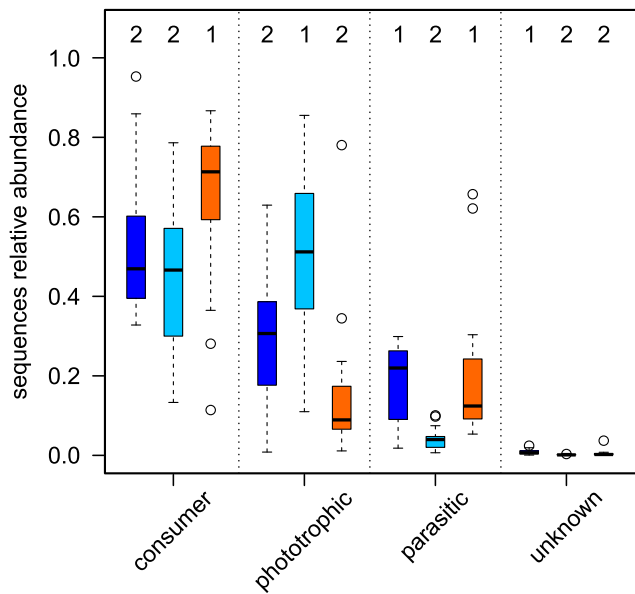
$\beta$  diversity was highest in freshwater (0.924  $\pm$  0.006), slightly but significantly lower in soils (0.894  $\pm$  0.006) and lowest in the marine plankton (0.777  $\pm$  0.009) (Fig. 1C, Table H.1). This higher biological heterogeneity in freshwater and soil systems reflects their respective heterogeneous abiotic conditions compared to the relatively buffered conditions in the ocean. Indeed, soil and freshwater samples vary considerably in their pH values which has a known impact on microbial



**Fig. 1.** General patterns of diversity of protists retrieved from a sampling throughout marine waters, freshwaters and soils. (A) Species accumulation curves calculated from protist communities retrieved from marine (dark blue), freshwater (cyan) and soil (orange) samples. The horizontal plain and dashed lines above the species accumulation curves indicate the predicted number of OTUs (bootstrap estimator of species richness) and standard error associated for each ecosystem respectively. (B) Shannon diversity indices per ecosystem. (C) Bray-Curtis distances within each ecosystems. (D) Ordination plot (non-metric multidimensional scaling) of OTU protists communities from marine water (dark blue circle), freshwater (cyan triangle) and soil (orange square). Numbers above the distributions of sub-figure B and C indicate significantly different distributions (Nemenyi test  $P < 0.05$ ), “1” representing the highest distribution and “3” the lowest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** Schematic phylogenetic tree of the main protist taxa and their relative abundance in three ecosystems (marine water: dark blue, freshwater: cyan, soil: orange). Barplots represent the OTU richness of each taxon predicted by bootstrap estimator of species richness in each ecosystem. Predictions higher than 500 OTUs are written numerically. Numbers in front of bars indicate significantly different groups (Nemenyi test  $P < 0.05$ ), "1" representing the highest distribution and "3" the lowest. Pie-chart represent the relative abundance of reads in function of each ecosystem (marine water: dark blue, freshwater: cyan, soil: orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Reads relative abundance from OTUs assigned to heterotrophs, phototrophs, parasites or unknown functional group in marine ecosystems (dark blue), freshwater (cyan) and soil (orange). Numbers at the top of the boxplot indicate significantly different distributions within a functional groups (Nemenyi test  $P < 0.05$ ), “1” representing the highest distribution and “2” the lowest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

life and diversity, as well as nutrient amounts (Delgado-Baquerizo et al., 2018). The passive dispersal of soil protists, which may tend to homogenize diversity is likely lower than for aquatic protists, as only some protists spread by wind and animals, many others relying on slow active dispersal (Wilkinson et al., 2011). Likewise, freshwater systems are fragmented in the landscape which may cause barriers to dispersal (Dodson, 1992; Reche et al., 2005). By contrast, protists are likely to have high dispersal in marine waters due to buffered condition of seawater pH, ion concentrations only varying to a limited extent, and oceans being highly connected through the global marine circulation (Richter et al., 2019). The homogeneity in community composition between marine communities was also corroborated by the smaller ordination space covered in the NMDS analysis, in comparison to soil and freshwater samples (Fig. 1D). The NMDS also showed that the three communities were different from each other (permutation test  $P < 0.001$  in the 100 bootstraps)

Our results illustrated the importance of the salinity barrier as a driver for protist diversity (Balzano et al., 2015; Logares et al., 2009), as soil and freshwater communities differ less from each-other than they do from marine communities (Fig. 1D). This is also shown by the high number of OTUs shared between freshwater and soil samples ( $1152 \pm 9$ ), while only  $79 \pm 1$  OTUs are shared between marine and freshwater samples, and  $14 \pm 0$  between soil and marine samples (Fig. D.1, H.5). Only  $25 \pm 0$  OTUs were recorded in all ecosystems, including small heterotrophic flagellates i.e. the kinetoplastid Neobodonid, and the Raphid-Pennate diatoms from which genera *Nitzschia* and *Pseudo-nitzschia*, which are indeed known to tolerate a wide range of salinity (Scholz and Liebezeit, 2012). However, considering the rather coarse taxonomic discrimination allowed by the 18S rRNA gene (Alverson, 2008) and the small length of the markers used in this study, it is possible that these ubiquitous OTUs correspond to different but relatively closely-related species.

### 3.3. The distribution of taxonomic and functional diversity

Consumers dominated sequence relative abundance in soil (Fig. 3,

Table H.3). While rhizarians in terrestrial systems (freshwater and soil) were dominated by Cercozoa, marine rhizarians mostly belonged to Radiolaria (Fig. E.2). Ciliates were overall, respectively, the richest and second richest group of consumers in freshwater and soil ecosystems and show the same pattern for relative abundance (Fig. E.2). Their broad range of lifestyles is key to their evolutionary success in many ecosystems (Lynn, 2008) although their highly polyploid macronucleus might somehow artificially inflate their estimated abundance in the environment (de Vargas et al., 2015). Stramenopiles constituted the third most represented major group of consumers overall (Fig. E.2). Non-photosynthetic Stramenopiles included organisms collectively known as MAST (Massana et al., 2004) (MARine STramenopiles, now known to also thrive in other ecosystems), Bicoecia and Labyrinthulomycetes. Labyrinthulomycetes were previously considered largely marine until new deep-branching phagotrophic groups were discovered in other ecosystems (Gomaa et al., 2013). Our data confirm their prevalence in soil, as they present even higher  $\alpha$  and  $\gamma$  diversity in these systems (Fig. B.1-2).

Phototrophs were overall the second most abundant functional group, and dominated freshwater systems (Fig. 3, Table H.3). Interestingly, their relative abundance in marine plankton was not significantly higher than in soil (Fig. 3, Table H.3, Nemenyi test:  $P > 0.05$ ). Given the importance of phototrophy in marine systems for global carbon fixation and nutrient cycling (Falkowski et al., 1998), this relative abundance suggests a remarkable and unexpected role of microbial photosynthesis in soils. Although the standing biomass of soil phototrophs (“subaerial algae”) is negligible as compared with plants, their turnover is likely much faster and accelerated by specific predation (Mann and Vanormelingen, 2013; Seppey et al., 2017), as is well documented in the ocean. Mostly photosynthetic dinoflagellates were extremely abundant in marine systems while Chrysophyceae dominated the freshwaters and Archaeplastida were widespread in soils (Fig. E.3). Interestingly, the main freshwater phototrophic groups (phototrophic Chrysophyceae, Cryptophyta, phototrophic Dinophyceae) behave also alternatively as consumers (mixotrophs). In soils, these groups are way less represented, with the groups being present often having lost the capacity for photosynthesis (such as Clade C Chrysophytes, collectively designated as *Spumella* spp. (Boenigk et al., 2005). Hence, in freshwater systems the number of strict consumers seems lower than in other systems, but this role may be assumed to some extent by mixotrophic phototrophs. Many dinoflagellates and Archaeplastida often form mutualistic associations with other protists, but also fungi and animals (LaJeunesse, 2001). Their high abundance agrees with the immense importance of photosymbiosis at a global level.

Parasitic taxa represented roughly 15 to 20% of all sequences in marine and soil systems, but only around 5% in freshwater systems (Fig. 3), where taxa of the fungal phylum Chytridiomycota likely fill the ample niche of eukaryotic parasites (Rasconi et al., 2012). Chytrids are indeed generally well represented in environmental DNA-based diversity surveys (Debroas et al., 2017; Lepère et al., 2019). In soils, genuine parasites (who did not necessarily kill their host) were dominated by Apicomplexans, an extremely diverse and abundant group known to infect Metazoans (Mahé et al., 2017; Singer et al., 2020), and plant parasites such as many terrestrial oomycetes and plasmodiophorids (Fig. E.4). In marine systems the parasitoid strategy (in which the host is killed to complete its life cycle) dominates with an overwhelming relative abundance of Syndiniales, which lyse host cells before spreading propagules (Fig. E.4). Freshwater systems comprised parasitoid groups such as Perkinsea, Syndiniales and most certainly Chytridiomycota (Sime-Ngando et al., 2015), but also “true” parasites like Peronosporomycetes, Apicomplexa and Ichtyosporia (Fig. E.4). This suggests a high connectivity of terrestrial systems with the macroscopic food web (plants and animals) whereas in aquatic systems (freshwaters and marine waters) parasitism would affect mostly other microbes that would be rapidly lysed, thus creating bloom dynamics which may accelerate the pace of nutrient turnover (Mangot et al., 2013). The

longer lifespan of macroscopic organisms and the fact that these might not be systematically killed by their parasites suggests different nutrient transfer through the systems.

In spite of all differences observed among the three ecosystems, our survey revealed a remarkable functional homogeneity. Heterotrophy dominates all processes in general, taking into account that even in freshwater where the proportion of phototrophs is higher, many pigmented organisms are actually mixotrophic (i.e. combining phagocytosis and photosynthesis (Sanders, 1991)). Phototrophs appear in lower proportions, their standing biomass probably constrained by intense predation enhancing nutrient turnover. The action of parasites also likely contributes to nutrient turnover, by lysing microbial cells through a parasitoid strategy in aquatic ecosystems and increasing the death rate of macro-organisms in terrestrial ecosystems.

### 3.4. Conclusions

A key result of our study is that soils host the richest protist communities on Earth. Thus, while all ecosystems contain a wealth of undescribed protist species, this is especially true in soils. In line with our predictions, spatial and ecological heterogeneity are correlated with diversity, as shown by the higher  $\beta$  diversity observed in freshwater and soil environments. This calls for an increased exploration of less prospected ecosystem types to find new diversity followed by the expansion of the existing curated and annotated databases such as UniEuk (Berney et al., 2017).

Our estimations for functional diversity were only partially fulfilled. On one hand, our study underlines the importance of phagocytosis through the relative abundance and diversity of consumers in all ecosystems, and especially in soils. On the other hand, the amount of photosynthesizing organisms were not significantly different in marine and soil samples. Given the recognized importance of marine phototrophs for global nutrient turnover, these results suggest that soil phototrophs might have a larger contribution to the global carbon cycle than previously thought.

### 4. Authors' contributions

D.S., C.V.W.S., E.A.D.M and E.L. designed the experiments; D.S., M.D., C.V., D.B., L.B., Q.B., D.D., C.D., I.D., I.L., G.M., M.R.S., E.A.D.M., A.G. and E.L. provided the samples; D.S., performed the laboratory work; C. V.W.S., D.S., G.L. and E.L. analyzed the data; D.S., C.V.W.S., S.G., E.A.D.M., and E.L. wrote the first version of the manuscript, which was then edited by all co-authors.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work was supported by the Swiss National Science Foundation 31003A\_143960 & 31003A\_163254 to E.L., P2NEP3-178543 to D.S. and 31003A\_182531 to E.A.D.M.; internal funding of the University of Neuchâtel to E.A.D.M.; the 'Atracción de talentos' from the Community of Madrid, project 2017-T1/AMB-5210 and the project MYXOTROPIC VI (PGC2018-094660-B-I00) from the Spanish Government to E.L.; NWO-VENI grant from the Netherlands Organisation for Scientific Research (016.Veni.181.078) to S.G.; the Norwegian Research Council projects 270252 (BiodivERSA-Climate change impacts on Arctic soil and lake microbiomes) and 256132 (ERANet-LAC, METHAnogenic Biodiversity and activity in Arctic and Sub-Antarctic Ecosystems affected by climate change) to C.V.W.S.; the French Government 'Investissements d'Avenir' program OCEANOMICS (ANR-11-BTBR-0008) to C.d.V., and the Gordon

and Betty Moore Foundation through grant GBMF5257 UniEuk. The authors would like to thank all the peoples involved in the sampling and in the laboratory and the Calculations Center of the Faculty of Science of the University of Neuchâtel as well as UNINETT Sigma2 – the National Infrastructure for High Performance Computing (project NN9549K) and Data Storage (project NS9593K) in Norway for the computing facilities.

### Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2020.106262>.

### References

- Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Cepička, I., Chistyakova, L., Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V., Zhang, Q., 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* 66 (1), 4–119.
- Alverson, A.J., 2008. Molecular systematics and the diatom species. *Protist* 159 (3), 339–353.
- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., Huse, S.M., 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS one* 4, e6372. <https://doi.org/10.1371/journal.pone.0006372>.
- Balzano, S., Abs, E., Leterme, S.C., 2015. Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.* 74 (3), 263–277.
- Berney, C., Ciuprina, A., Bender, S., Brodie, J., Edgcomb, V., Kim, E., Rajan, J., Parfrey, L. W., Adl, S., Audic, S., Bass, D., Caron, D.A., Cochrane, G., Czech, L., Dunthorn, M., Geisen, S., Glöckner, F.O., Mahé, F., Quast, C., Kaye, J.Z., Simpson, A.G.B., Stamatakis, A., del Campo, J., Yilmaz, P., de Vargas, C., 2017. UniEuk: Time to Speak a Common Language in Protistology! *J. Eukaryot. Microbiol.* 64 (3), 407–411.
- Boenigk, J., Pfandl, K., Stadler, P., Chatzinotas, A., 2005. High diversity of the 'Spumellalike' flagellates: an investigation based on the SSU rRNA gene sequences of isolates from habitats located in six different geographic regions. *Environ. Microbiol.* 7 (5), 685–697.
- Boenigk, J., Wodniok, S., Bock, C., Beisser, D., Hempel, C., Grossmann, L., Lange, A., Jensen, M., 2018. Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *MBMG* 2, e21519. <https://doi.org/10.3897/mbmg.2.21519>.
- Chrétiennot-Dinet, M.-J., Courties, C., Vaquer, A., Neveux, J., Claustre, H., Lautier, J., Machado, M.C., 1995. A new marine picococcal eukaryote: *Ostreococcus tauri* gen. et sp. nov. (Chlorophyta, Prasinophyceae). *Phycologia* 34 (4), 285–292.
- Cotterill, F.P.D., Al-Rasheid, K., Foissner, W., 2007. Conservation of protists: is it needed at all? In: *Protist Diversity and Geographical Distribution*. Springer, pp. 193–209.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Becot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandel-Lewis, S., Coordinators, T.O., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, 1261605. <https://doi.org/10.1126/science.1261605>.
- Debroas, D., Domaizon, I., Humbert, J.-F., Jardillier, L., Lepère, C., Oudart, A., Taïb, N., 2017. Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiol. Ecol.* 93, fix023. <https://doi.org/10.1093/femsec/fix023>.
- Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D. J., Bardgett, R.D., Maestre, F.T., Singh, B.K., Fierer, N., 2018. A global atlas of the dominant bacteria found in soil. *Science* 359 (6373), 320–325.
- Dodson, S., 1992. Predicting crustacean zooplankton species richness. *Limnol. Oceanogr.* 37 (4), 848–856.
- Dorrell, R.G., Azuma, T., Nomura, M., Audren de Kerdrel, G., Paoli, L., Yang, S., Bowler, C., Ishii, K.-I., Miyashita, H., Gile, G.H., Kamikawa, R., 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc. Natl. Acad. Sci. USA* 116 (14), 6914–6923.
- Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* 281, 200–206. <https://doi.org/10.1126/science.281.5374.200>.
- Gaston, K.J., 2000. Global patterns in biodiversity. *Nature* 405 (6783), 220–227.
- Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D., Jousset, A., Krashevska, V., Singer, D., Spiegel, F.W., Walochnik, J., Lara, E., 2018. Soil protists: a fertile frontier in soil biology research. *FEMS Microbiol. Rev.* 42, 293–323. <https://doi.org/10.1093/femsec/fuy006>.
- Geisen, S., Mitchell, E.A.D., Wilkinson, D.M., Adl, S., Bonkowski, M., Brown, M.W., Fiore-Donno, A.M., Heger, T.J., Jassey, V.E.J., Krashevska, V., Lahr, D.J.G., Marcisz, K.,

