

# Soil protist diversity in the Swiss western Alps is better predicted by topo-climatic than by edaphic variables

Christophe V.W. Seppely<sup>1,2\*</sup>, Olivier Broennimann<sup>3,4</sup>, Aline Buri<sup>4</sup>, Erika Yashiro<sup>3,5</sup>, Eric Pinto-Figueroa<sup>3,5</sup>, David Singer<sup>1,6</sup>, Quentin Blandenier<sup>1,7</sup>, Edward A.D. Mitchell<sup>1,8</sup>, H el ene Niculita Hirzel<sup>9</sup>, Antoine Guisan<sup>3,4+</sup>, and Enrique Lara<sup>1,6+</sup>

<sup>1</sup>Laboratory of Soil Biodiversity, University of Neuch atel,  mile-Argand 11, 2000 Neuch atel, Switzerland

<sup>2</sup>Department of Arctic and Marine Biology, University of Troms , Framstredet 39, 9019 Troms , Norway

<sup>3</sup>Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

<sup>4</sup>Institute of Earth Surface Dynamics, University of Lausanne, G opolis, 1015 Lausanne, Switzerland

<sup>5</sup>Department of Fundamental Microbiology, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

<sup>6</sup>Department of Zoology, Institute of Biosciences, University of S o Paulo, Mat o 14, 05508-090 S o Paulo, Brazil

<sup>7</sup>Real Jard n Bot nico, CSIC, Plaza de Murillo 2, 28014 Madrid

<sup>8</sup>Botanical Garden of Neuch atel, Perthuis-du-Sault 58, 2000 Neuch atel, Switzerland

<sup>9</sup>Institute for Work and Health, Universities of Lausanne and Geneva, 1066 Epalinges-Lausanne, Switzerland

\*Corresponding author: [christophe.seppely@uit.no](mailto:christophe.seppely@uit.no)

+Co-last authors

## Abstract

**Aim:** General trends in spatial patterns of macroscopic organisms diversity can be reasonably well predicted from correlative models, using for instance topo-climatic variables for plants and animals allowing inference over large scales. By contrast, soil microorganisms diversity is generally considered as mostly driven by edaphic variables and, therefore, difficult to extrapolate on a large spatial scale based on predictive models. Here, we compared the power of topo-climatic vs. edaphic variables for predicting the diversity of various soil protist groups at the regional scale.

**Location:** Swiss western Alps.

**Taxa:** Full protist community and nine clades belonging to three functional groups: parasites (Apicomplexa, Oomycota, Phytomyxea), phagotrophs (Sarcomonadea, Tubulinea, Spirotrichea) and phototrophs (Chlorophyta, Trebouxiophyceae, Bacillariophyta).

**Methods:** We extracted soil environmental DNA from 178 sites along a wide range of elevations with a random-stratified sampling design. We defined protist Operational Taxonomic Units assemblages by metabarcoding of the V4 region of the ribosomal RNA small sub-unit gene. We

15 assessed and modelled the diversity (Shannon index) patterns of all selected groups as a function  
16 of topo-climatic and edaphic variables using Generalized Additive Models.

17 **Results:** The respective significance of topo-climatic and edaphic variables varied among taxo-  
18 nomic and – to a certain extent – functional groups: while many variables explained significantly  
19 the diversity of phototrophs this was less the case for parasites. Generally, topo-climatic vari-  
20 ables had a better predictive power than edaphic variables, yet predictive power varied among  
21 taxonomic and functional groups.

22 **Main conclusions:** Topo-climatic variables are, on average, better predictors of protist diversity at  
23 the landscape scale than edaphic variables, which opens the way to wide-scale sampling designs  
24 avoiding costly and time-consuming laboratory protocols. However, predictors of diversity differ  
25 considerably among taxonomic and functional groups; such relationships may be due to direct  
26 and/or indirect, e.g. biotic influences. Future prospects include using such spatial models to  
27 predict hotspots of diversity or pathogens outbreaks.

28 **KEYWORDS:** edaphic variables, predictive modelling, protist diversity, soils, spatial modelling,  
29 Swiss western Alps, topo-climatic variables.

## 30 Introduction

31 Protists, i.e. all eukaryotes with the exception of fungi, plants and animals are hyper-diverse in soil  
32 systems (Geisen et al., 2018; Mahé et al., 2017), where they play many ecological roles as primary  
33 producers, saprotrophs, predators, or parasites (Adl & Gupta, 2006; Geisen et al., 2016). Photo-  
34 synthetic groups are essential components of cryptogamic crusts (Elbert et al., 2012; Pushkareva,  
35 Johansen, & Elste, 2016) and constitute a significant source of organic carbon for soil organisms  
36 (Schmidt, Dyckmans, & Schrader, 2016; Seppey et al., 2017). Predatory protists occupy differ-  
37 ent levels of the microbial food web, as primary consumers of algae (cyanobacteria or eukaryotic),  
38 fungi and bacteria (Bonkowski & Clarholm, 2012; Dumack, Mueller, & Bonkowski, 2016; Hess &  
39 Melkonian, 2014), but also occupy higher trophic levels by predating on phagotrophic protists or  
40 even micro-Metazoa (e.g. nematodes) (Geisen et al., 2015; Gilbert, Amblard, Bourdier, Francez, &  
41 Mitchell, 2000). Parasites are thought to regulate natural populations, notably of animals (Mahé et  
42 al., 2017) and can be either very specific such as between the parasitic Gregarines and their animal  
43 hosts (Clopton, 2009), or generalist as for Phytomyxea species which can infect hosts from different  
44 eukaryotic kingdoms (Neuhauser, Kirchmair, Bulman, & Bass, 2014). Characterizing such complex  
45 communities is essential to understand the main on-going ecological processes in soil. This task has  
46 been rendered possible only recently with the development of high throughput sequencing, allowing  
47 to assess the taxonomic diversity of soil protists, to infer functional diversity and to determine how  
48 the patterns and drivers of this diversity compare to the better-known plants and animals.

49 As a whole, soil protist communities have been shown to respond to edaphic condition such as  
50 gradients of pH (Dupont, Griffiths, Bell, & Bass, 2016), nutrients and moisture (Singer et al., 2018)  
51 as well as pesticide amounts (Ekelund, 1999; Foissner, 1999; Nesbitt & Adl, 2014) and other pertur-  
52 bations (Foissner, 1997). These variables are rarely integrated in spatial modelling of biodiversity in  
53 general Mod, Scherrer, Luoto, & Guisan, (2016) for plant communities), especially at broad spatial  
54 scales, because they are most often not available at the sites of species observations and not easily  
55 generalizable in a spatially-explicit way (Buri et al., 2017; Dubuis et al., 2013). On the other hand,  
56 topo-climatic variables (such as slope steepness or air temperature) can be more easily modelled at  
57 large spatial scales using digital elevation models based on interpolations of weather stations and/or  
58 remote sensing methods. These variables already proved useful to model the spatial distribution of  
59 plants and animals (Franklin, 2010; Guisan, Thuiller, & Zimmermann, 2017; Peterson et al., 2011)  
60 but, to our knowledge, very rarely on micro-organisms. As a consequence, spatial modelling of the

61 distribution of microorganisms has been restricted to small areas or aquatic environments (Bulit, 2014;  
62 Fraile, Schulz, Mulitza, & Kucera, 2008; King et al., 2010; Langer, Weinmann, Loetters, Bernhard,  
63 & Roedder, 2013; Mitchell et al., 2000; Zaric, Schulz, & Mulitza, 2006; Zinger, Shahnavaz, Baptist,  
64 Geremia, & Choler, 2009). The development of such models at the landscape scale would allow  
65 assessing at a much broader scale the processes driven by micro-organisms, such as nutrient cycling,  
66 predicting zones at risk of pathogenic outbreaks or simply identifying protist diversity hotspots.

67 Here, we built spatial predictive models of protist diversity, focussing on general communities as  
68 well as on nine broad protist taxa chosen within three functional groups - phototrophs, phagotrophs  
69 and parasites - along a wide elevation gradient in the western Swiss Alps. We assessed the diversity  
70 of protists in 178 meadow soil samples, resulting from a robust random-stratified field survey by  
71 metabarcoding of the V4 regions of the small sub-unit rRNA gene. This study assessed the extent of  
72 protist diversity in mountainous meadows and determined to what extent two sets of environmental  
73 variables (edaphic and topo-climatic) can predict this diversity over the whole Swiss western Alps  
74 of the Vaud state. In addition, we brought an interpretation of the patterns observed based on  
75 knowledge of the lifestyles of the different groups surveyed.

## 76 **Material and Methods**

### 77 **Sampling**

78 Meadow soils were sampled from 194 plots distributed across the Swiss western Alps; of these plots,  
79 178 samples successfully yielded sequencing data and were used in the current study. Sampling was  
80 performed from July 4<sup>th</sup> to September 1<sup>st</sup> 2013 according to a random stratified sampling design.  
81 From each plot, five soil cores (100 grams per core between the depths of 0-5 cm after removing  
82 plants, mosses, and insects) were taken from the four corners and the centre of a 2 m<sup>2</sup> plot. The five  
83 cores, were then pooled in a sterile plastic bag and kept in an icebox or at 4 °C until DNA extraction  
84 and soil analyses were done. A subsample of the pooled soil was also flash frozen at each sampling  
85 site and kept frozen until further soil analyses. For more details, see Yashiro et al. (2016).

### 86 **Edaphic variables**

87 We selected eight edaphic variables that were measured directly in the field or on the soil samples.  
88 Soil temperature was measured in the field. The relative humidity (rh) was assessed by weighing the  
89 mass of the soil sample before and after drying at 105 °C during 2 days. Soil organic Carbon content  
90 was determined by loss of ignition (LOI) at 1050 °C. The percentage of shale was determined by laser  
91 granulometry. The pH and electro-conductivity (EC) were measured from a soil and Milli-Q water  
92 slurry in a 1:2.5 and 1:5 (wt/vol) ratio respectively. Total phosphorus amount was determined by  
93 colorimetric analysis after a mineralisation at 550 °C with Mg(NO<sub>3</sub>)<sub>2</sub>. The C/N ratio was calculated  
94 from the total organic carbon and nitrogen percentages measured by ROCK EVAL pyrolysis (Vinci  
95 Technologies, Ruell-Malmaison, France) and combustion infrared spectroscopy (Carlo Erba CNS2500  
96 CHN), respectively. See Yashiro et al. (2016) and Buri et al. (2017) for more details.

### 97 **Topo-climatic variables**

98 Values for seven topo-climatic variables were retrieved from maps of 25 square meter resolution for  
99 each sample location. We used the number of growing degree-days above 0 °C (gdd), potential evapo-  
100 transpiration (etp), topography (topo), slope southness (asp) and slope steepness (slp)(Zimmermann  
101 & Kienast, 1999; Zimmermann, Edwards, Moisen, Frescino, & Blackard, 2007). In addition, we  
102 calculated the summer temperature average (tmean678) and precipitation sum (psum678) for the

103 months of June to August with values of monthly temperature means and precipitation sums from  
104 1981 to 2010. See Buri et al. (2017) for more details.

## 105 Molecular analysis

106 DNA was extracted from the soil samples using the MoBio PowerSoil DNA extraction kit (Calsbad,  
107 CA, USA) following the manufacturer instructions. The V4 region of the 18S rRNA gene was then am-  
108 plified using the general eukaryotic primers TAREuk454FWD1 and TAREukREV3 (CCAGCASCYCGGTAATTC  
109 / TYRATCAAGAACGAAAGT) (Stoeck et al., 2010). The PCR mix was composed of 3  $\mu$ L DNA extract,  
110 0.4  $\mu$ L of 10 mg/mL BSA, 4  $\mu$ L of PCR buffer (Promega GoTaq M7845), 0.2  $\mu$ L of Taq polymerase  
111 (Promega GoTaq M7845), 0.6  $\mu$ L of dNTPs (Promega kit U1420), 0.6  $\mu$ L of each primer (MicroSynth,  
112 Balgach, Switzerland), and 10.6  $\mu$ L of ultra-pure water. The PCR reactions started with a denat-  
113 uration step at 95 °C for 5 min followed by 45 cycles of 94 °C for 30 s, 47 °C for 45 s and 72  
114 °C for 1 min, and terminated with an elongation step of 72 °C for 10 min. For each DNA sample,  
115 the amplifications were performed in triplicate with a PTC-200 Peltier Thermo Cycler (BioConcept,  
116 Allschwil, Switzerland). DNA was then quantified with a Qubit<sup>®</sup> 2.0 Fluorometer (Invitrogen) and 20  
117 ng of each triplicate were pooled. A DNA library was prepared from the pools using the TruSeq Nano  
118 PCR-free Library Preparation kit and the paired-end 2x300 bp sequencing was done on an Illumina<sup>®</sup>  
119 MiSeq at the University of Geneva (Molecular Systematics & Environmental Genomics Laboratory).

## 120 Bioinformatics pipeline

121 Good quality sequences were selected based on their nucleotides phred scores. Every sequence with  
122 a phred score average below 20 for a 50 nucleotides window was discarded. The chimeras were then  
123 removed using the program VSEARCH 1.11.1 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) by  
124 comparing the environmental sequences 1) with each other for each replicate and 2) against the PR<sup>2</sup>  
125 database trimmed according to the V4 primers (downloaded on the 12 September 2016; Guillou et  
126 al. (2013)). To reduce the noise caused by very rare sequences, we then discarded every singleton.  
127 Triplicates were then pooled according to their respective samples and OTUs were built with the  
128 program SWARM 2.1.8 (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2015). The dominant  
129 sequence of each OTU was taxonomically assigned by aligning it to the trimmed PR<sup>2</sup> database using  
130 the global pairwise alignment program GGSEARCH 36.3.6 (Pearson, 2000).

131 We removed every OTU that did not belong to protists, namely Metazoa, Embryophyceae and  
132 Fungi. We also discarded OTUs with a percentage of identity (PID) below 65% with the database  
133 PR<sup>2</sup> as sequences with such low PID are usually of prokaryotic origin (threshold verified manually  
134 by aligning low PID environmental sequences on GenBank database). From the 178 plots, 4 were  
135 sampled twice and 13 were sampled three times during the sampling period. For each of these 17  
136 plots we took the average (2 samples) or median (3 samples) sequence abundance of each OTU for  
137 the samples from the same plot. In addition of the total protist community matrix, we also selected  
138 nine broad taxonomic groups (i.e. clades, low taxonomic resolution) from three functional groups  
139 (1) parasites: Apicomplexa, Oomycota, Phytomyxea; (2) phagotrophs: Sarcomonadea, Tubulinea,  
140 Spirotrichea and (3) phototrophs: Chlorophyceae, Trebouxiophyceae, Bacillariophyta). These taxa  
141 were selected because they are abundant and diverse in soils and are functionally homogeneous. For  
142 each of these taxa, we established a PID threshold verified manually on GenBank to discarded potential  
143 misidentification (Apicomplexa: 80%, Oomycota: 80%, Phytomyxea: 75%, Sarcomonadea: 80%,  
144 Tubulinea: 75%, Spirotrichea: 90%, Chlorophyceae: 90%, Trebouxiophyceae: 85%, Bacillariophyta:  
145 77%).

## 146 Richness and diversity analyses

147 For each of the ten taxonomic data sets (all protists plus nine broad groups), OTU richness and  
148 Shannon diversity (H) were calculated, and the differences between their statistical distributions tested  
149 by a multiple comparisons of mean rank sums test (Nemenyi test; Hollander, Wolfe, & Chicken, 2015,  
150 `posthoc.kruskal.nemenyi.test` function, 'PMCMR' package 4.1; Pohlert, 2014). Computation  
151 of H indices includes quantitative data, classically proportion of a given species in a given sample,  
152 which can be reasonably inferred by numbers of reads in High Throughput Sequencing data. Indeed,  
153 there is a correspondence between this number of reads and the biovolume of individual organisms  
154 that has been showed for many groups of protists (Giner et al., 2016; Kosakyan, Mulot, Mitchell, &  
155 Lara, 2015; Weber & Pawlowski, 2013). H indices provide thus a reasonable estimation of the OTU  
156 diversity. We then assessed the relationships between Shannon diversity (H), and topo-climatic and  
157 edaphic variables. For this, we firstly assessed pairwise correlations between all predictors and, in  
158 all pairs with correlation  $>0.7$ , we only kept the expectedly most causal one for further analyses to  
159 avoid collinearity issues (Dormann et al., 2013) (see Appendix Fig. S1.1 in Supporting Information).  
160 Then, for each of the ten data sets, H was modelled as a function of the environmental variables  
161 using a Generalized Additive Model (GAM; assuming Gaussian residuals and identity link function).  
162 For each data set, three models were calibrated; the first with topo-climatic variables only, the second  
163 with edaphic variables only, and the third with both sets of variables. All models were iterated  
164 100 times based on bootstraps composed of 80% of the 178 original samples. In total  $10 \times 3 \times 100$   
165 models were fitted. For each model, the predictive power was estimated as the Root Mean Square  
166 Error (RMSE) calculated on the independent samples not included to build the model (20% left-out  
167 samples). The effect of taxonomic group and the set of predictors on predictive power (RMSE) was  
168 tested by a Nemenyi test. Finally, the diversities of the nine broad taxa and total protist diversity  
169 were extrapolated on the full area of the western Swiss Alps with a GAM including the topo-climatic  
170 variables (i.e. the only spatially-explicit variables).

## 171 Results

### 172 Observed diversity patterns

173 We retrieved a total of 24'322'487 good quality sequences of which 97% were not chimeric and  
174 71% were not singletons. The 17'110'114 remaining sequences were clustered into 41'048 OTUs of  
175 which 19'260 were assigned to protists (Table 1). Protist diversity was dominated (proportion of  
176 sequences) by Cercozoa, (principally Sarcomonadea and Thecofilosea), and Alveolata of which more  
177 than half were assigned to Apicomplexa and ca. 45% to Ciliophora (mostly from classes Spirotrichea,  
178 Oligohymenophorea, Litostomatea and Colpodea) (see Appendix Fig. S1.2). The three other domi-  
179 nant groups were the Stramenopiles (including Oomycota and Bacillaryophyta), Amoebozoa (includ-  
180 ing Tubulinea) and Archaeplastida (with Chlorophyceae and Trebouxiophyceae) (see Appendix Fig.  
181 S1.2).

182 The nine chosen taxa jointly contributed to over half (54%) of all retained sequences and rep-  
183 resented over 35% of the total OTU richness (Table 1). The average richness per sample of these  
184 clades varied from 7 (Phytomyxea) to 249 (Sarcomonadea). Richness was in average lowest for pho-  
185 tototrophs (15 OTUs / sample) and highest for phagotrophs (122 OTUs / sample; Fig. 1). Shannon  
186 diversity indices followed the same trend, varying from an average value of 1.1 (Phytomyxea) to 4.3  
187 (Sarcomonadea).

## 188 Environmental models of diversity

189 The percentage of variance (adjusted-R<sup>2</sup>) of the Shannon diversity in the total and broad taxonomic  
190 groups explained by the combination of both topo-climatic and edaphic variables ranged from 6%  
191 (Bacillariophyta) to 33% (Chlorophyceae) (Table 2). The environmental variables explaining a sig-  
192 nificant ( $p < 0.05$ ) portion of the protist diversity in these models with combined datasets were: slope  
193 steepness (in 4 taxa), pH (3 taxa), mean summer temperature (2 taxa), Soil Organic C (2 taxa),  
194 shale percentage (1 taxon), C/N (1 taxon), phosphorus (1 taxon) and EC (1 taxon) (Table 2). The  
195 predictive power showed lower RMSE values (i.e. a better power) for the topo-climatic than for the  
196 edaphic variables for all taxa except the Chlorophyceae, Trebouxiophyceae and Sarcomonadea where  
197 the values were higher or similar (Fig. 2). In addition, the RMSE of the models calculated on the  
198 edaphic and topo-climatic variables together were never significantly lower than the RMSE calculated  
199 for the topo-climatic variable alone. The RMSE also varied among taxonomic groups and the diversity  
200 of certain taxa were significantly better predicted (e.g. Oomycota) than others (e.g. Apicomplexa)  
201 (Table 3). This RMSE variation was also observed at the functional level (see Appendix Table S1.1):  
202 the RMSE of functional group were always as good as or better than the RMSE calculated from the  
203 total community.

## 204 Discussion

### 205 General patterns of micro-eukaryotic diversity in soils

206 Our study revealed several important findings on patterns of protist diversity across temperate moun-  
207 tain landscapes. Phagotrophs (e.g. Sarcomonadea & Tubulinea) and parasites (Apicomplexa) were  
208 the most abundant functional groups in terms of read abundance. Apicomplexan sequences, albeit  
209 numerous, were much less abundant and diversified than in Neotropical soils: as arthropods are less  
210 abundant and diversified in temperate regions, this brings further support to the hypothesis that soil  
211 apicomplexan diversity mirrors that of arthropods in the ecosystem (Mahé et al., 2017). Another  
212 abundant parasitic group is the Oomycetes (Stramenopiles), which contains many plant parasites,  
213 but also animal pathogens and a few free-living, saprotrophic forms (Beakes, Glockling, & Sekimoto,  
214 2012; Lara & Belbahri, 2011). Oomycetes were shown to be common and diverse in temperate soil  
215 systems (Seppey et al., 2017; Singer et al., 2016). By contrast they are less abundant and diverse in  
216 neotropical forest soil ecosystems, where they comprise mostly animal parasites (Mahé et al., 2017).

217 Within phagotrophs, the high proportion of sequences from Cercozoa (mostly to Sarcomonadea)  
218 was in line with previous soil eukaryotic DNA surveys (Bates et al., 2013; Harder et al., 2016; Seppey  
219 et al., 2017). Earlier studies based on microscopical observations showed the prevalence of these  
220 groups in soils (Adl & Gupta, 2006). Ciliates were also a well-represented phagotrophic group, and  
221 were dominated by Spirotrichea, which corroborates also other findings on soil protist molecular  
222 diversity (Lara, Berney, Ekelund, Harms, & Chatzinotas, 2007). In summary, the protist communities  
223 found in the Swiss western Alps are typical for average soil ecosystems and the findings can probably  
224 be extrapolated to other regions.

### 225 Model fit and predictive power of topo-climatic and edaphic variables

226 Slope steepness and pH were the two variables most often found to significantly contribute to the fit of  
227 our different protist diversity models. Slope steepness affects drainage and leaching of nutrients and is  
228 generally inversely correlated to soils depth. Nevertheless, a enhanced drainage reduces the likelihood  
229 of water-logging which would select for very specialized protists tolerating anoxia and generally would  
230 lead to lower diversity. Soil pH is well known as a major driver of microbial diversity, including bacteria



231 (Santoyo, Hernandez-Pacheco, Hernandez-Salmeron, & Hernandez-Leon, 2017; Yashiro et al., 2016),  
232 fungi (Noyce et al., 2016; Pellissier et al., 2014; Zhang, Jia, & Yu, 2016) and protists (Bates et al.,  
233 2013; Dupont et al., 2016). The relationship between pH and protist diversity was significant only for  
234 three groups, being negative for two groups of phagotrophs (Spirotrichea and Sarcomonadida) and  
235 positive for Chlorophyceae. It is unclear if these relationships reflect a direct effect of pH or rather  
236 indirect effects such as biotic effects (e.g. impact on bacterial or fungal food sources), the availability  
237 of nutrients for the growth of autotrophs, or other drivers.

238 Predictability varies also to a large extent between functional groups. Indeed, while many variables  
239 explained significantly the diversity of phototrophs and phagotrophs, it was less so for parasites (see  
240 Appendix Table S1.1). The latter depend only indirectly on environmental values, and mainly on their  
241 hosts, which brings logically supplementary noise in analyses. For nine out of the ten taxonomic group  
242 tested, the predictive power of the topo-climatic variables was either significantly better, or at least  
243 not different than the ones including the edaphic variables. Moreover, it was never lower than the  
244 predictive power of the models including both sets of variables. This suggests that, within the levels  
245 of predictability achieved, predictive models built solely on topo-climatic variables are as accurate, or  
246 possibly even better than the models built with the addition of edaphic variables. These variables are  
247 available at large scales and are already largely used for modelling the spatial distribution of macro-  
248 organisms (Guisan & Zimmermann, 2000), to the contrary of local edaphic values that are always  
249 tedious and costly to measure in the landscape across large regions and environmental gradients.  
250 These findings open the way to larger sampling designs that could further increase the performance  
251 of models.

252 The correspondence between OTUs and biological species has always been a hot topic in eukaryotic  
253 environmental microbiology. Indeed, a single SSU rRNA gene sequence may include, in certain groups,  
254 a wide diversity of species with different lifestyles and ecological preferences. This has been shown for  
255 different soil protists such as ciliates (Lara, & Acosta-Mercado, 2012); in contrast, in Myxomycetes  
256 (Amoebozoa), SSU sequences are truly hypervariable and discriminate relatively accurately between  
257 species (Dahl et al., 2018). There is, therefore, no general rule that applies to all eukaryotes. However,  
258 SSU sequences are generally considered good proxy for eukaryotic diversity, as they were the first  
259 benchmark for protist barcoding (Pawlowski et al., 2012), and therefore, predictions based this proxy  
260 can as well be considered a good estimation of actual protists' diversity.

## 261 **Interpretation of the spatial patterns of protist diversity modelled with topo-** 262 **climatic variables**

263 As for macro-organisms (D'Amen, Pradervand, & Guisan, 2015; Dubuis et al., 2011; McCain, 2005;  
264 Reymond, Purcell, Cherix, Guisan, & Pellissier, 2013), but unlike other soil micro-organisms (Bryant  
265 et al., 2008; Fierer et al., 2010; Pellissier et al., 2014), protists diversity show clear spatial and  
266 elevational patterns when only topo-climatic variables are taken to build the model (Fig. 3). This  
267 patterns seems to be driven by summer temperature in most cases (see Appendix Table S1.1), either  
268 in a positive (Bacillariophyta, Phytomyxea and Tubulinea), unimodal (Apicomplexa, Sarcomonadea  
269 and Spirotrichea) or negative way (Chlorophyceae, Oomycota). A positive correlation of diversity with  
270 temperature (and, thus, productivity) is a typical pattern in macroecology that can be related to the  
271 species-energy hypothesis as long as moisture is not a limiting factor (Fernández et al., 2016), or other  
272 models for diversity patterns (Huston, 1994; see Spehn & Körner, (2009) for elevation gradients). On  
273 the other hand, if moisture is limiting, unimodal patterns are to be expected, and diversity peaks where  
274 both moisture and energy are optimal (water energy model: Fernández et al. (2016)) intermediate  
275 disturbance hypothesis or mid-domain effect (discussed for the same area in Dubuis et al. (2011)).  
276 Finally, Chlorophyceae and Oomycota are typically sensitive to high temperatures and desiccation,  
277 both including often flagellated life stages for dispersal that needs at least a thin water film to disperse

278 (Jeger & Pautasso, 2008). In addition, Chlorophyta high diversity in the lowest temperature zone  
279 (Fig. 3) could be explained by the fact that micro-eukaryotic algae have a higher growth rate at  
280 low temperatures, favouring diversification in cold environments (Rose & Caron, 2007) or possibly  
281 reduced competition from vascular plants.

## 282 Conclusion

283 We showed that the diversity of some taxa and functional groups, is explained up to >30% by  
284 topo-climatic and edaphic conditions. A somewhat surprising result is that topography and climate  
285 predicted protist diversity as well or better than the edaphic variables more commonly used in soil  
286 microbial studies. This implies that soil protist diversity patterns could be at least partly inferred, for  
287 some groups (e.g. Chlorophyceae) and to some extent (22%), based on topo-climatic spatial models  
288 only.

289 Such an approach could be applied at finer taxonomic levels to predict the distribution of individual  
290 species, which would be of high socio-economic relevance in the case of invasive agricultural or  
291 forestry pests of economic importance such as certain oomycetes. The models could be improved by  
292 refining the taxonomic groups, as taxa responding more homogeneously to environmental conditions  
293 may show stronger correlation with abiotic variables than the broad group classification we used.  
294 For instance, the Oomycota contain organisms belonging to other functional groups than parasites  
295 (e.g. saprotroph; Beakes et al., 2012; Lara & Belbahri, 2011) or able to target a wide range of  
296 hosts (e.g. *Phytophthora cinnamomi*; Hardham, 2005). These improvements would pave the way  
297 toward extrapolation of protists diversity across large spatial scales and provide useful tools to identify  
298 biodiversity hotspots, predict spatially the risk of pathogen infection or model soil protist diversity  
299 according to future environmental change scenarios.

## 300 Acknowledgements

301 The authors would like to thank all fields and laboratory assistants and technicians who participate in  
302 the alpine soil project, with a particular thank to Amandine Pillonel, Laura Desponds and Dessislava  
303 Savova Bianchi for the laboratory work. Special thanks to the Transports Publics du Chablais and  
304 Glacier 3000 (<https://www.glacier3000.ch/en>) who provide their installations for free to carry  
305 the soils from the sampling sites. We would also like to thanks all anonymous farmers who allowed  
306 sampling on their lands.

307 The study was funded by the Swiss National Science Foundation under the projects 310003A  
308 143960 to E.L. and 31003A-152866 (SESAM'ALP), PDFMP3-135129 (MICROBIAL BIOGEOGRA-  
309 PHY) and CR23I2-162754 (INTEGRALP) to A.G., as well as the internal funding of the Universities  
310 of Neuchâtel and Lausanne. E.L. would also like to thanks the program "Atracción de talentos" from  
311 the Community of Madrid project 2017-T1/AMB-5210. E.Y. also thanks the European Community  
312 FP7-PEOPLE-2010-IIF program (MP-Alps, grant agreement 273965), the Agassiz Foundation, and  
313 the Pro-Femmes Fellowship program from the Faculty of Biology and Medicine of the University of  
314 Lausanne.

315 A significant part of the computations was also performed on resources provided by the Calcula-  
316 tions Center of the Faculty of Science of the University of Neuchâtel and by UNINETT Sigma2 - the  
317 National Infrastructure for High Performance Computing and Data Storage in Norway.



## 318 References

- 319 Adl, M. S., & Gupta, V. V. S. R. (2006). Protists in soil ecology and forest nutrient cycling. *Canadian Journal of Forest Research*,  
320 36(7), 1805–1817.
- 321 Bates, S. T., Clemente, J. C., Flores, G. E., Walters, W. A., Parfrey, L. W., Knight, R., & Fierer, N. (2013). Global biogeography  
322 of highly diverse protistan communities in soil. *Isme Journal*, 7(3), 652–659.
- 323 Beakes, G. W., Glockling, S. L., & Sekimoto, S. (2012). The evolutionary phylogeny of the oomycete "fungi". *Protoplasma*, 249  
324 (1), 3–19.
- 325 Bonkowski, M., & Clarholm, M. (2012). Stimulation of Plant Growth through Interactions of Bacteria and Protozoa: Testing the  
326 Auxiliary Microbial Loop Hypothesis. *Acta Protozoologica*, 51(3), 237–247.
- 327 Bryant, J. A., Lamanna, C., Morlon, H., Kerkhoff, A. J., Enquist, B. J., & Green, J. L. (2008). Microbes on mountainsides:  
328 Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of The National Academy of Sciences of The  
329 United States of America*, 105(1), 11505–11511.
- 330 Bulit, C. (2014). Good Reasons and Guidance for Mapping Planktonic Protist Distributions. *Acta Protozoologica*, 53(1), 13–27.
- 331 Buri, A., Cianfrani, C., Pinto-Figueroa, E., Yashiro, E., Spangenberg, J. E., Adatte, T., ... Pradervand, J. N. (2017). Soil factors  
332 improve predictions of plant species distribution in a mountain environment. *Progress in Physical Geography*, 41(6), 703–722.
- 333 Clopton, R. E. (2009). Phylogenetic relationships, evolution, and systematic revision of the septate Gregarines (Apicomplexa:  
334 Eugregarinorida: Septatorina). *Comparative Parasitology*, 76(2), 167–190.
- 335 Dahl, M. B., Brejnrod, A. D., Unterseher, M., Hoppe, T., Feng, Y., Novozhilov, Y., ... Schnittler, M. (2018). Genetic barcoding of  
336 dark-spored myxomycetes (Amoebozoa)-Identification, evaluation and application of a sequence similarity threshold for species  
337 differentiation in NGS studies. *Molecular Ecology Resources*, 18(2), 306–318.
- 338 D'Amen, M., Pradervand, J. N., & Guisan, A. (2015). Predicting richness and composition in mountain insect communities at  
339 high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, 24(12), 1443–1453.
- 340 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., ... Lautenbach, S. (2013). Collinearity: a review of  
341 methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46.
- 342 Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J. P., & Guisan, A. (2011). Predicting spatial patterns of plant species  
343 richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, 17  
344 (6), 1122–1131.
- 345 Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., & Guisan, A. (2013). Improving the prediction of plant species  
346 distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science*, 24(4),  
347 593–606.
- 348 Dumack, K., Mueller, M. E. H., & Bonkowski, M. (2016). Description of *Lecythium terrestris* sp nov (Chlamydomphryidae,  
349 Cercozoa), a Soil Dwelling Protist Feeding on Fungi and Algae. *Protist*, 167(2), 93–105.
- 350 Dupont, A. O. C., Griffiths, R. I., Bell, T., & Bass, D. (2016). Differences in soil micro-eukaryotic communities over soil pH  
351 gradients are strongly driven by parasites and saprotrophs. *Environmental Microbiology*, 18(6), 2010–2024.
- 352 Ekelund, F. (1999). The impact of the fungicide fenpropimorph (Corbel (R)) on bacterivorous and fungivorous protozoa in soil.  
353 *Journal of Applied Ecology*, 36(2), 233–243.
- 354 Elbert, W., Weber, B., Burrows, S., Steinkamp, J., Buedel, B., Andreae, M. O., & Poeschl, U. (2012). Contribution of cryptogamic  
355 covers to the global cycles of carbon and nitrogen. *Nature Geoscience*, 5(7), 459–462.
- 356 Fernández, L. D., Fournier, B., Rivera, R., Lara, E., Mitchell, E. A. D., & Hernandez, C. E. (2016). Water-energy balance,  
357 past ecological perturbations and evolutionary constraints shape the latitudinal diversity gradient of soil testate amoebae in  
358 south-western South America. *Global Ecology and Biogeography*, 25(10), 1216–1227.
- 359 Fierer, N., McCain, C. M., Meir, P., Zimmermann, M., Rapp, J. M., Silman, M. R., & Knight, R. (2010). Microbes do not follow  
360 the elevational diversity patterns of plants and animals. *Ecology*, 92(4), 797–804.
- 361 Foissner, W. (1997). Protozoa as bioindicators in agroecosystems, with emphasis on farming practices, biocides, and biodiversity.  
362 *Agriculture Ecosystems & Environment*, 62(2-3), 93–103.
- 363 Foissner, W. (1999). Soil protozoa as bioindicators: pros and cons, methods, diversity, representative examples. *Agriculture  
364 Ecosystems & Environment*, 74(1), 95–112.
- 365 Fraile, I., Schulz, M., Mulitza, S., & Kucera, M. (2008). Predicting the global distribution of planktonic foraminifera using a  
366 dynamic ecosystem model. *Biogeosciences*, 5(3), 891–911.
- 367 Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge: Cambridge University Press.

- 368 Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T., & Bonkowski, M. (2015). Pack hunting by a common soil amoeba  
369 on nematodes. *Environmental Microbiology*, 17(11), 4538–4546.
- 370 Geisen, S., Koller, R., Huenninghaus, M., Dumack, K., Urich, T., & Bonkowski, M. (2016). The soil food web revisited: Diverse  
371 and widespread mycophagous soil protists. *Soil Biology & Biochemistry*, 94, 10–18.
- 372 Geisen, S., Mitchell, E. A. D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., ... Lara, E. (2018). Soil protists: a fertile  
373 frontier in soil biology research. *Fems Microbiology Reviews*, 42(3), 293–323.
- 374 Gilbert, D., Amblard, C., Bourdier, G., Francez, A. J., & Mitchell, E. A. D. (2000). Le régime alimentaire des Thécamoebiens  
375 (Protista, Sarcodina). *L'Annee Biologique*, (39), 57–68.
- 376 Giner, C. R., Forn, I., Romac, S., Logares, R., de Vargas, C., & Massana, R. (2016). Environmental Sequencing Provides  
377 Reasonable Estimates of the Relative Abundance of Specific Picoeukaryotes. *Applied and Environmental Microbiology*, 82,  
378 4757–4766.
- 379 Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., ... Christen, R. (2013). The Protist Ribosomal Reference  
380 database (PR<sup>2</sup>): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids  
381 Research*, 41(1), D597–D604.
- 382 Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3),  
383 147–186.
- 384 Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models*. Cambridge: Cambridge,  
385 University Press.
- 386 Harder, C. B., Ronn, R., Brejnrod, A., Bass, D., Abu Al-Soud, W., & Ekelund, F. (2016). Local diversity of heathland Cercozoa  
387 explored by in-depth sequencing. *ISME Journal*, 10(10), 2488–2497.
- 388 Hardham, A. R. (2005). *Phytophthora cinnamomi*. *Molecular Plant Pathology*, 6(6), 589–604.
- 389 Hess, S., & Melkonian, M. (2014). Ultrastructure of the Algivorously Amoebophilic *Viridiraptor invadens* (Glissomonadida,  
390 Cercozoa). *Protist*, 165(5), 605–635.
- 391 Hollander, M., Wolfe, D. A., & Chicken, E. (2015). *Nonparametric statistical methods* (3rd ed.). Hoboken, NJ: Wiley.
- 392 Huston, M. A. (1994). *Biological diversity: the coexistence of species on changing landscapes*. Cambridge: Cambridge University  
393 Press.
- 394 Jeger, M. J., & Pautasso, M. (2008). Comparative epidemiology of zoospore plant pathogens. *European Journal of Plant  
395 Pathology*, 122(1), 111–126.
- 396 King, A. J., Freeman, K. R., McCormick, K. F., Lynch, R. C., Lozupone, C., Knight, R., & Schmidt, S. K. (2010). Biogeography  
397 and habitat modelling of high-alpine bacteria. *Nature Communications*, 1.
- 398 Kosakyan, A., Mulet, M., Mitchell, E. A. D., & Lara, E. (2015). Environmental DNA COI barcoding for quantitative analysis of  
399 protists communities: A test using the *Nebela collaris* complex (Amoebozoa; Arcellinida; Hyalospheniidae). *European Journal  
400 of Protistology*, 51(4), 311–320.
- 401 Langer, M. R., Weinmann, A. E., Loettters, S., Bernhard, J. M., & Roedder, D. (2013). Climate-Driven Range Extension of  
402 *Amphistegina* (Protista, Foraminiferida): Models of Current and Predicted Future Ranges. *Plos One*, 8(2).
- 403 Lara, E., & Acosta-Mercado, D. (2012). A molecular perspective on ciliates as soil bioindicators. *European Journal of Soil Biology*,  
404 49(S1), 107–111.
- 405 Lara, E., & Belbahri, L. (2011). SSU rRNA reveals major trends in oomycete evolution. *Fungal Diversity*, 49(1), 93–100.
- 406 Lara, E., Berney, C., Ekelund, F., Harms, H., & Chatzinotas, A. (2007). Molecular comparison of cultivable protozoa from a  
407 pristine and a polycyclic aromatic hydrocarbon polluted site. *Soil Biology & Biochemistry*, 39(1), 139–148.
- 408 Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon  
409 clustering. *PeerJ*, 3.
- 410 Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., ... Dunthorn, M. (2017). Parasites dominate hyperdiverse  
411 soil protist communities in Neotropical rainforests. *Nature Ecology and Evolution*, 1(4).
- 412 McCain, C. M. (2005). Elevational gradients in diversity of small mammals. *Ecology*, 86(2), 366–372.
- 413 Mitchell, E. A. D., Borcard, D., Buttler, A. J., Grosvernier, P., Gilbert, D., & Gobat, J. M. (2000). Horizontal distribution patterns  
414 of testate amoebae (Protozoa) in a *Sphagnum magellanicum* carpet. *Microbial Ecology*, 39(4), 290–300.
- 415 Mod, H. K., Scherrer, D., Luoto, M., & Guisan, A. (2016). What we use is not what we know: environmental predictors in plant  
416 distribution models. *Journal of Vegetation Science*, 27(6), 1308–1322.

- 417 Nesbitt, J. E., & Adl, S. M. (2014). Differences in soil quality indicators between organic and sustainably managed potato fields  
418 in Eastern Canada. *Ecological indicators*, 37(A), 119–130.
- 419 Neuhauser, S., Kirchmair, M., Bulman, S., & Bass, D. (2014). Cross-kingdom host shifts of phytomyxid parasites. *Bmc*  
420 *Evolutionary Biology*, 14.
- 421 Noyce, G. L., Fulthorpe, R., Gorgolewski, A., Hazlett, P., Honghi, T., & Basiliko, N. (2016). Soil microbial responses to wood ash  
422 addition and forest fire in managed Ontario forests. *Applied Soil Ecology*, 107, 368–380.
- 423 Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., ... de Vargas, C. (2012). CBOL Protist Working Group:  
424 Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *Plos Biology*, 10(11).
- 425 Pearson, W. R. (2000). Flexible Sequence Similarity Searching with the FASTA3 Program Package. In S. Misener & S. Krawetz  
426 (Eds.), *Bioinformatics Methods and Protocols* (pp. 185–218). New York, NY: Humana Press.
- 427 Pellissier, L., Niculita-Hirzel, H., Dubuis, A., Pagni, M., Guex, N., Ndiribe, C., ... Guisan, A. (2014). Soil fungal communities of  
428 grasslands are environmentally structured at a regional scale in the Alps. *Molecular Ecology*, 23(17), 4274–4290.
- 429 Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R., Martínez-Meyer, E., Nakamura, M., & Araújo, M. P. (2011). *Ecological*  
430 *niches and geographic distributions*. Princeton: Princeton University Press.
- 431 Pohlert, T. (2014). *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)* (Version 4.1). Retrieved from  
432 <http://CRAN.R-project.org/package=PMCMR>.
- 433 Pushkareva, E., Johansen, J. R., & Elster, J. (2016). A review of the ecology, ecophysiology and biodiversity of microalgae in  
434 Arctic soil crusts. *Polar Biology*, 39(12), 2227–2240.
- 435 Reymond, A., Purcell, J., Cherix, D., Guisan, A., & Pellissier, L. (2013). Functional diversity decreases with temperature in high  
436 elevation ant fauna. *Ecological Entomology*, 38(4), 364–373.
- 437 Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics.  
438 *PeerJ*, 4.
- 439 Rose, J. M., & Caron, D. A. (2007). Does low temperature constrain the growth rates of heterotrophic protists? Evidence and  
440 implications for algal blooms in cold waters. *Limnology and Oceanography*, 52(2), 886–895.
- 441 Santoyo, G., Hernandez-Pacheco, C., Hernandez-Salmeron, J., & Hernandez-Leon, R. (2017). The role of abiotic factors modulating  
442 the plant-microbe-soil interactions: toward sustainable agriculture. A review. *Spanish Journal of Agricultural Research*, 15(1).
- 443 Schmidt, O., Dyckmans, J., & Schrader, S. (2016). Photoautotrophic microorganisms as a carbon source for temperate soil  
444 invertebrates. *Biology Letters*, 12(1).
- 445 Seppely, C. V. W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E. A. D., & Lara, E. (2017). Distribution patterns  
446 of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling.  
447 *Soil Biology & Biochemistry*, 112, 68–76.
- 448 Singer, D., Lara, E., Steciow, M. M., Seppely, C. V. W., Paredes, N., Pillonel, A., ... Belbahri, L. (2016). High-throughput  
449 sequencing reveals diverse oomycete communities in oligotrophic peat bog micro-habitat. *Fungal Ecology*, 23, 42–47.
- 450 Singer, D., Kosakyan, A., Seppely, C. V. W., Pillonel, A., Fernandez, L. D., Fontaneto, D., ... Lara, E. (2018). Environmental  
451 filtering and phylogenetic clustering correlate with the distribution patterns of cryptic protist species. *Ecology*, 99(4), 904–914.
- 452 Spehn, M., & Körner, C. (2009). *Data mining for global trends in mountain biodiversity*. Boca Raton, FL: CRC Press.
- 453 Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H., & Richards, T. A. (2010). Multiple marker parallel  
454 tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*,  
455 19(s1), 21–31.
- 456 Weber, A. A. T., & Pawlowski, J. (2013). Can Abundance of Protists Be Inferred from Sequence Data: A Case Study of  
457 Foraminifera. *Plos One*, 8(2).
- 458 Yashiro, E., Pinto-Figueroa, E., Buri, A., Spangenberg, J. E., Adatte, T., Niculita-Hirzel, H., ... van der Meer, J. R. (2016).  
459 Local Environmental Factors Drive Divergent Grassland Soil Bacterial Communities in the Western Swiss Alps. *Applied and*  
460 *Environmental Microbiology*, 82(21), 6303–6316.
- 461 Zaric, S., Schulz, M., & Mulitza, S. (2006). Global prediction of planktic foraminiferal fluxes from hydrographic and productivity  
462 data. *Biogeosciences*, 3(2), 187–207.
- 463 Zhang, T., Jia, R., & Yu, L. Y. (2016). Diversity and distribution of soil fungal communities associated with biological soil crusts  
464 in the southeastern Tengger Desert (China) as revealed by 454 pyrosequencing. *Fungal Ecology*, 23, 156–163.
- 465 Zimmermann, N. E., & Kienast, F. (1999). Predictive mapping of alpine grasslands in Switzerland: Species versus community  
466 approach. *Journal of Vegetation Science*, 10(4), 469–482.

- 467 Zimmermann, N. E., Edwards, T. C., Moisen, G. G., Frescino, T. S., & Blackard, J. A. (2007). Remote sensing-based predictors  
468 improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, 44(5),  
469 1057–1067.
- 470 Zinger, L., Shahnava, B., Baptist, F., Geremia, R. A., & Choler, P. (2009). Microbial diversity in alpine tundra soils correlates  
471 with snow cover dynamics. *Isme Journal*, 3(7), 850–859.

## 472 Biosketch

473 The Laboratory of Soil Biodiversity (<https://www.unine.ch/biolso1>), led by Prof. Edward A.D.  
474 Mitchell, is interested in the diversity, biogeography and ecology of soil organisms with a strong focus  
475 on protists and links to other soil organisms and ecosystem ecology. The lab combines observational  
476 and experimental studies leading to applications in biomonitoring, palaeoecology, ecotoxicology and  
477 forensic sciences. The Spatial Ecology Group (<http://www.unil.ch/ecospat>), led by Prof. Antoine  
478 Guisan, is specialized in spatial modeling of biodiversity at the levels of species, communities, and  
479 ecosystems. Models are applied to the conservation of endangered species, the management of  
480 biological invasions, and the assessment of global change impact on biodiversity, with a special and  
481 long-term focus on above- and below-ground biota in the Western Swiss Alps.

482 Author contributions: E.Y., E.A.D.M., H.N.H., A.G., and E.L. conceived the idea; E.Y., E.A.D.M.,  
483 A.G. and E.L. provided the funding; A.B., E.Y., E.P.F. and A.G. collected the data; A.B., E.Y., E.P.F.,  
484 D.S., Q.B. and E.L. performed the laboratory work; C.V.W.S., O.B., A.B., A.G. and E.L. analysed  
485 the data and C.V.W.S., O.B., E.Y., D.S., Q.B., E.A.D.M., A.G. and E.L. wrote the manuscript. All  
486 authors gave final approval for publication.

## Soil protist diversity in the Swiss western Alps is better predicted by topo-climatic than by edaphic variables

Christophe V.W. Seppely<sup>1,2\*</sup>, Olivier Broennimann<sup>3,4</sup>, Aline Buri<sup>4</sup>, Erika Yashiro<sup>3,5</sup>, Eric Pinto-Figueroa<sup>3,5</sup>, David Singer<sup>1,6</sup>, Quentin Blandenier<sup>1,7</sup>, Edward A.D. Mitchell<sup>1,8</sup>, H el ene Niculita Hirzel<sup>9</sup>, Antoine Guisan<sup>3,4+</sup>, and Enrique Lara<sup>1,6+</sup>

<sup>1</sup>Laboratory of Soil Biodiversity, University of Neuch atel,  mile-Argand 11, 2000 Neuch atel, Switzerland

<sup>2</sup>Department of Arctic and Marine Biology, University of Troms , Framstredet 39, 9019 Troms , Norway

<sup>3</sup>Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

<sup>4</sup>Institute of Earth Surface Dynamics, University of Lausanne, G opolis, 1015 Lausanne, Switzerland

<sup>5</sup>Department of Fundamental Microbiology, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

<sup>6</sup>Department of Zoology, Institute of Biosciences, University of S o Paulo, Mat o 14, 05508-090 S o Paulo, Brazil

<sup>7</sup>Real Jard n Bot nico, CSIC, Plaza de Murillo 2, 28014 Madrid

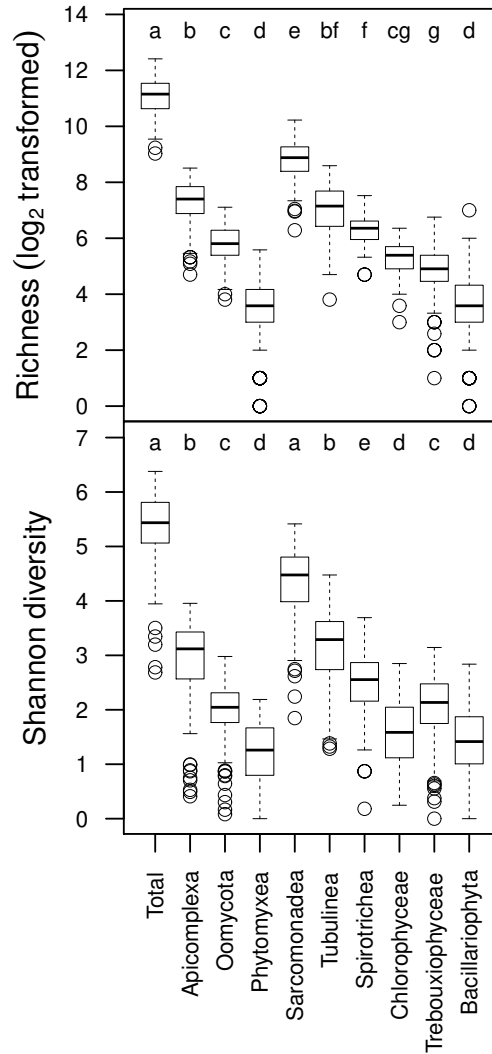
<sup>8</sup>Botanical Garden of Neuch atel, Perthuis-du-Sault 58, 2000 Neuch atel, Switzerland

<sup>9</sup>Institute for Work and Health, Universities of Lausanne and Geneva, 1066 Epalinges-Lausanne, Switzerland

\*Corresponding author: [christophe.seppely@uit.no](mailto:christophe.seppely@uit.no)

+Co-last authors

## Figure and tables



**Figure 1:** Shannon diversity and richness (log transformed) distributions of protist operational taxonomic units communities retrieved from 178 plots in the Swiss western Alps. The distributions are shown for the total community as well as for nine broad taxa. The letters above the boxplots represent groups according to a multiple comparison mean rank sums test (Nemenyi test  $P < 0.05$ ).



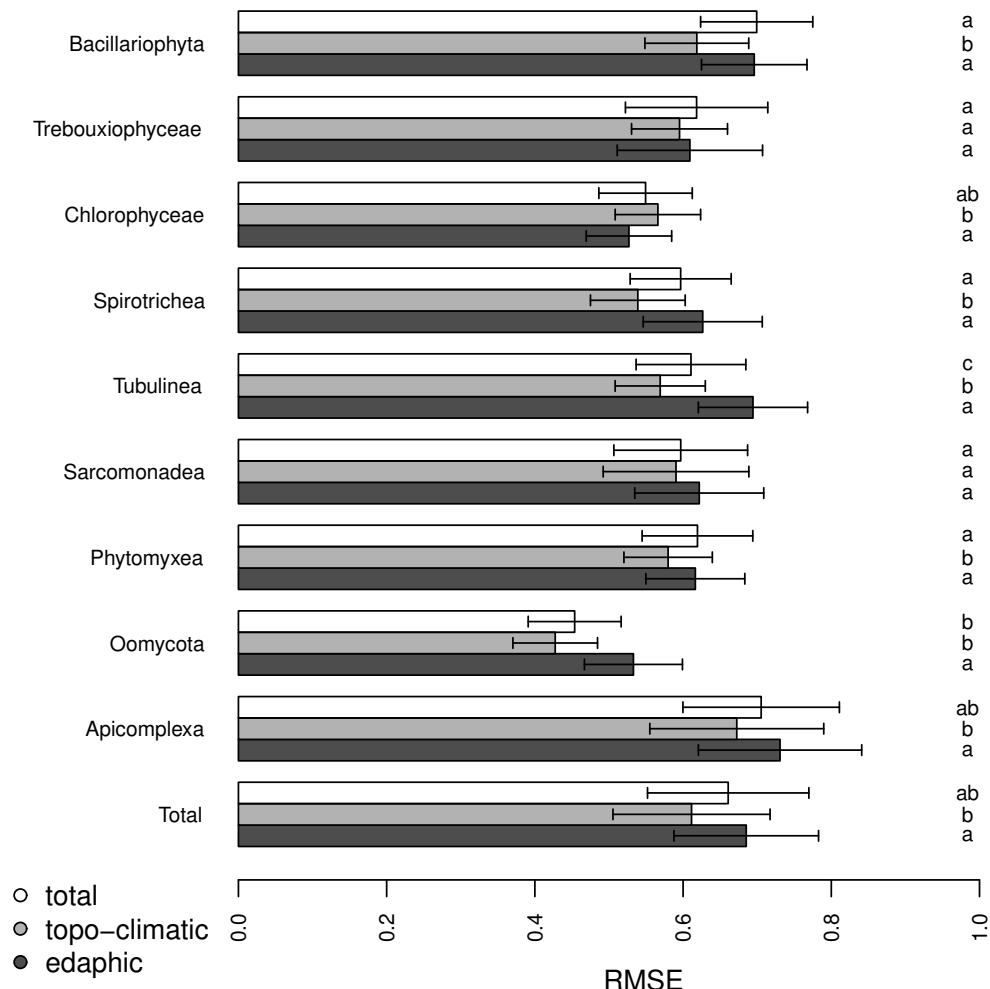
**Table 1:** Number of sequences and operational taxonomic units retrieved from 178 plots in the Swiss western Alps, through each step of the analysis for the total community and for the nine broad taxa. The numbers between parenthesis represent the percentage of the total community.

	number of sequences		number of OTUs	
quality check	24322487			
chimera removal	23724876			
singleton removal	17234091			
clustering	17110114		41050	
unwanted taxa removal	3303601		19605	
samples pooling	2752582		19260	
Apicomplexa	376533	(14%)	1215	(6%)
Oomycota	149331	(5%)	354	(2%)
Phytomyxea	18976	(1%)	113	(<1%)
Sarcomonadea	486121	(18%)	3469	(18%)
Tubulinea	141167	(5%)	1006	(5%)
Spirotrichea	130016	(5%)	377	(2%)
Chlorophyceae	130254	(5%)	235	(1%)
Trebouxiophyceae	27473	(1%)	171	(<1%)
Bacillariophyceae	14432	(1%)	193	(1%)

**Table 2:** Significance of edaphic (soil temperature: Soil temp, relative humidity: rh, pH, electroconductivity: EC, total phosphorus amount: P, carbon/nitrogen ratio: C/N, loss of ignition: LOI, shale percentage) and topo-climatic (topography: topo, slope southness: asp, slope steepness: slp, summer temperature average: tmean678) predictors on the diversity modelled (Generalized Additive Model) from total micro-eukaryotic community and nine broad taxonomic groups from operational taxonomic units gathered from 178 meadow soils in the Swiss western Alps. The + and – signs show if the diversity is positively or negatively associated to the predictor and the number of signs inform on the strength of the association (between parenthesis: P < 0.1, one sign: P < 0.05, two signs: P < 0.01, three signs: P < 0.001). The -+ and +- indicate minimum and maximum of diversity at mid-predictor value respectively. Details of the response of each taxonomic group to the different variables can be found in Fig. S1.3.

	Edaphic								Topo-climatic				adj R <sup>2</sup>
	Soil temp (°C)	rh (‰)	pH	EC (µS/cm)	P (%)	C/N	LOI (%)	Shale (%)	topos	asp (%)	slp (%)	tmean678 (°C * 100)	
Total		(+)	(-)										0.18
Apicomplexa		(+)											0.27
Oomycota							+						0.11
Phycomycota					(+)		(+-)				++		0.16
Sarcomonadea			--						(-)	(-)	++		0.16
Tubulinea							-				+++		0.19
Spirotrichea			-			(-)						---	0.13
Chlorophyceae			+	+-	+								0.33
Trebouxiophyceae				(+-)		+-					+	-	0.25
Bacillariophyta											(-)	(+)	0.06

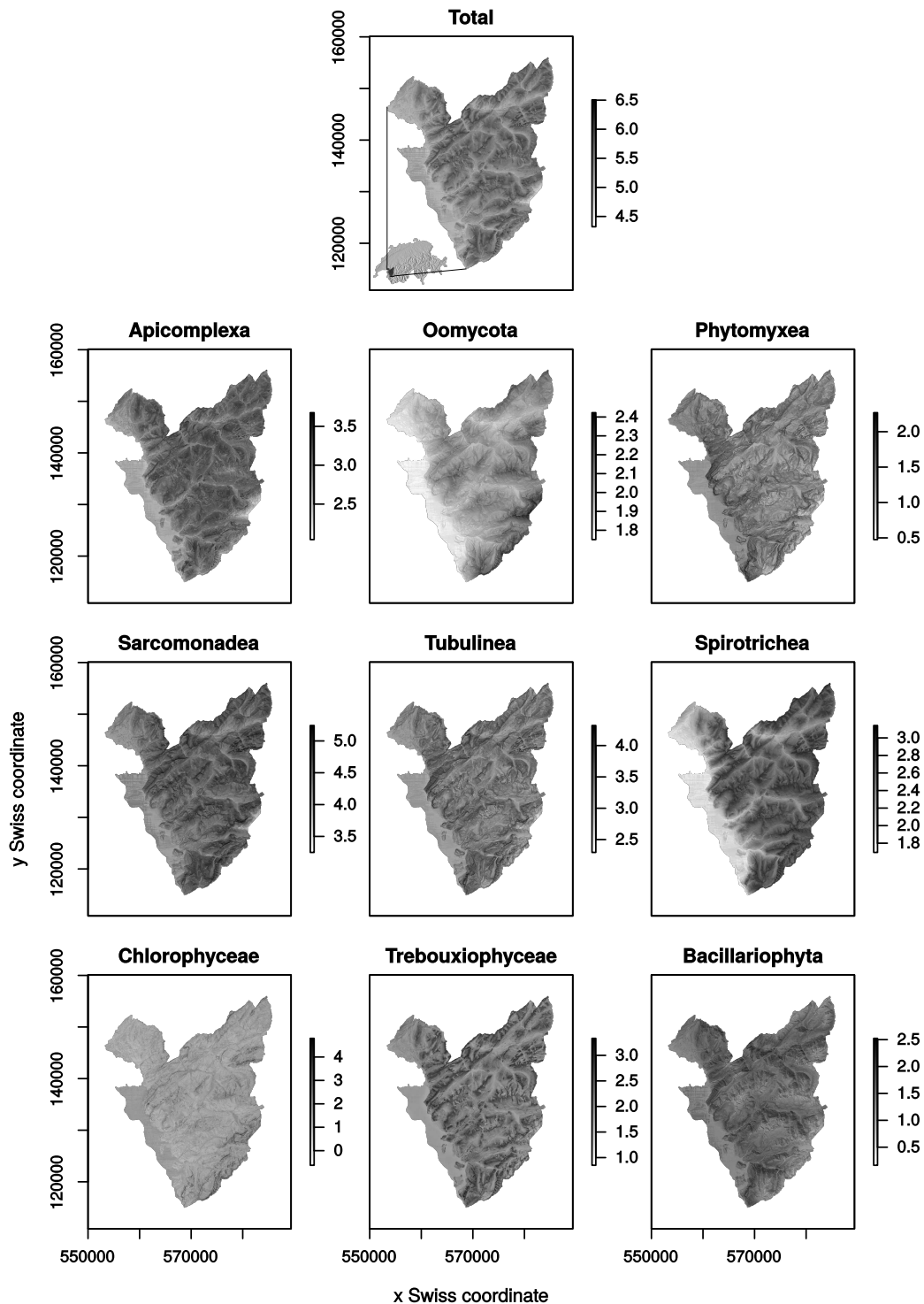
4



**Figure 2:** Predictive power (Root Mean Square Error: RMSE) of edaphic (dark grey), topo-climatic (pale gray) and overall (white) predictors calculated on the diversity of protist operational taxonomic units from the total community and nine broad taxa retrieved from 178 meadow soils in the Swiss western Alps. The RMSE were calculated on 100 cross validation of Generalized Additive Models performed with 20% of the samples as test dataset. The letters on the right of the barplot represent significantly different groups according to a multiple comparison mean rank sums test (Nemenyi test  $P < 0.05$ ) for each of the total communities and nine broad taxa.

**Table 3:** Average and standard deviation of the predictive power (Root Mean Square Error: RMSE) of edaphic, topo-climatic and overall predictors calculated on the diversity of protist operational taxonomic units from the total community and nine broad taxa as well as the functional groups the taxa were belonging to. The communities were retrieved from 178 meadow soils in the Swiss western Alps. The RMSE were calculated on 100 cross validation of Generalized Additive Models performed with 20% of the samples as test dataset. The letters between parenthesis represent significantly different groups according to a multiple comparison mean rank sums test (Nemenyi test  $P < 0.05$ ) for each of the total communities and nine broad taxa or for the total community and the three functional groups.

	edaphic		topo-climatic		overall	
Total	0.69 ± 0.10	(a)	0.61 ± 0.11	(ab)	0.66 ± 0.11	(ab)
Apicomplexa	0.73 ± 0.11	(a)	0.67 ± 0.12	(c)	0.71 ± 0.11	(ac)
Oomycota	0.53 ± 0.07	(b)	0.43 ± 0.06	(d)	0.45 ± 0.06	(d)
Phycomyxea	0.62 ± 0.07	(c)	0.58 ± 0.06	(ab)	0.62 ± 0.07	(be)
Sarcomonadea	0.62 ± 0.09	(c)	0.59 ± 0.10	(ab)	0.60 ± 0.09	(e)
Tubulinea	0.69 ± 0.07	(a)	0.57 ± 0.06	(ae)	0.61 ± 0.07	(be)
Spirotrichea	0.63 ± 0.08	(c)	0.54 ± 0.06	(e)	0.60 ± 0.07	(e)
Chlorophyceae	0.53 ± 0.06	(b)	0.57 ± 0.06	(ae)	0.55 ± 0.06	(f)
Trebouxiophyceae	0.61 ± 0.10	(c)	0.60 ± 0.06	(ab)	0.62 ± 0.10	(be)
Bacillariophyta	0.70 ± 0.07	(a)	0.62 ± 0.07	(bc)	0.70 ± 0.08	(c)
Total	0.69 ± 0.10	(a)	0.61 ± 0.10	(a)	0.66 ± 0.10	(a)
Parasites	0.63 ± 0.10	(b)	0.56 ± 0.10	(b)	0.59 ± 0.10	(b)
Phagotrophs	0.65 ± 0.09	(c)	0.57 ± 0.08	(b)	0.60 ± 0.08	(bc)
Phototrophs	0.61 ± 0.10	(b)	0.59 ± 0.07	(a)	0.62 ± 0.10	(c)



**Figure 3:** Diversity of the total protist community and nine broad taxa predicted from Generalized Additive Model through the Swiss western Alps based on the topography, slope southness, slope steepness and average temperature from June to August.