Original article:

Soil microorganisms behave like macroscopic organisms: patterns in the global distribution of soil euglyphid testate amoeba

Running header: Forest soil euglyphid diversity

Enrique Lara[1*], Ludovic Roussel-Delif [1, 2], Bertrand Fournier[3], David M. Wilkinson[4], Edward A. D. Mitchell[1,5]

[1] Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland.

[2] Present address: Laboratory of Microbiology, University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland.

[3] Institut des Sciences de l'Evolution, CNRS UMR 5554, Université de Montpellier II, Place Eugène Bataillon, CC 065, 34095, Montpellier Cedex 5, France

[4]Natural Science and Psychology, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK.

[5] Jardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

*correspondence: enrique.lara@unine.ch, phone: +41 32 718 23 27

ABSTRACT

Aim

Patterns of alpha and beta diversity of soil protist communities and the factors that shape them remain largely unknown. We undertook a worldwide survey of forest litter to investigate the patterns of diversity in a group of testate amoebae. We aimed to assess: (1) whether there is a latitudinal gradient in alpha diversity, and (2) whether beta diversity was correlated solely with environmental factors commonly used in soil biology research or if it was also independently explained by geographical barriers.

Location

Worldwide.

Methods

We studied the diversity of Euglyphida, a common group of testate amoebae, in 35 samples of forest litter and moss samples from a global survey, using small subunit rRNA gene sequences. We assessed the relationship between sample alpha diversity and latitude using generalized additive models (GAM). Furthermore, we determined the relationships between community composition and geographical models (distance-based Moran's Eigenvector Maps - db-MEM) using Generalized UniFrac distances (GUniFrac). We also investigated the relationship between individual measured soil parameters, WORLDCLIM data and diversity (alpha plus beta) using both raw data and synthetic variables obtained through principal components analysis.

Results

We recorded 245 phylotypes belonging to 6 out of 7 known Euglyphida families, plus four novel deep clades. Euglyphid alpha diversity was positively correlated with temperature and negatively with latitude and litter C/N ratio. Euglyphida community structure was correlated with the spatial

eigenvector Db-MEM31, independently of all measured environmental variables. Db-MEM31 corresponds to a natural barrier constituted by the Northern hemisphere desert belt. Beta diversity was correlated with other environmental variables, such as pH, isothermality and temperature in the coldest month of the year.

Main conclusions

Soil euglyphid alpha diversity displays a latitudinal gradient, and beta diversity is not only correlated with climatic and physicochemical parameters but also with geographic barriers. Such patterns of diversity were until recently believed to be characteristic only for macroscopic organisms.

Keywords: alpha diversity, beta diversity, climatic gradient, cosmopolitanism, environmental filters, geographical isolation, latitudinal gradient of diversity, protists, testate amoebae.

INTRODUCTION

Microbial biogeography has suffered from the difficulty of isolating and identifying species, and has traditionally been considered separately from macroscopic organisms biogeography (Martiny *et al.*, 2006). However, recently, observational and experimental studies on microorganisms have become commoner, and are increasingly using molecular approaches to attempt to overcome limitations inherent to morphology-based taxonomy (Lara & Acosta-Mercado, 2012). The result is that some paradigms are now being repeatedly questioned - especially the longstanding but provocative idea that different rules govern the diversity and distribution of microbes and macroscopic organisms (e.g. Finlay *et al.*, 2004). Indeed, recent studies have revealed the existence of a high diversity of microorganisms, often with limited distributions and showing evidence for historical contingencies and habitat preferences (Porazinska *et al.*, 2010; Heger *et al.*, 2011b; Naff *et al.*, 2013; Tedersoo & Smith, 2013; Lueke *et al.*, 2014). Identification of factors that affect the composition and diversity of soil microbial communities can improve our understanding of the impact of environmental changes on soil ecosystems; these have started to be investigated in bacteria, where communities were shown to be correlated with pH (Jackson & Fierer, 2006; Laubner *et al.* 2009). Protists, as major consumers of decomposers such as many bacteria and fungi (Old & Oros, 1980; Ekelund & Rønn, 1994) play a pivotal role in soil food webs and are therefore an excellent group for assessing these questions. Moisture has been suggested to influence global soil microeukaryotic diversity (Bates *et al.*, 2013), but no study has yet focussed in detail on a single, functionally homogeneous, group of protists at a global scale.

A well-documented pattern in many groups of macroscopic organisms is the positive correlation of diversity with latitude. Clearly latitude is a human construct – part of a grid system which can be used to define the position of a particular location and cannot in itself explain species richness. However, latitude is correlated with many other potentially significant environmental features, such as energy input into the system, which may potentially explain these patterns

(Sherratt and Wilkinson, 2009). This latitudinal pattern is often mirrored by a similar decrease in diversity with elevation, as observed in many plant and animal taxa, although patterns showing a peak in species richness at mid-elevations are also common (Adams & Woodward, 1989; Willig *et al.*, 2003; Hillebrand, 2004; Jablonski *et al.*, 2006; Cox & Moore, 2010). A range of factors can plausibly affect these patterns in diversity as well as energy input; for example water availability and evapotranspiration influence β-diversity in most animal and plant communities (Hawkins *et al.*, 2003) and historical process may also be important (Sherratt & Wilkinson, 2009).Currently, data supporting latitudinal diversity gradients are very limited for most groups of protists (Sherratt & Wilkinson, 2009), with the exception of large marine forms such as foraminiferans (Allen *et al.*, 2006; Yasuhara *et al.*, 2012) and polycistine radiolarians (Boltovskoy *et al.*, 2010). Using the excellent fossil record provided by benthic foraminiferans, Buzas *et al.* (2002) provided support for the idea that higher temperatures favoured faster speciation events through geological time. However, given the diversity of lifestyles encountered in protists, these conclusions cannot be reasonably extended to all other taxa.

The influence of distance and geographical barriers on microbial diversity, or in other terms the existence of geographically limited distributions, was hotly debated around the turn of the 21st century. Baas Becking's tenet "everything is everywhere, but, the environment selects" (Baas Becking, 1934), frequently referred to as "EiE" postulates that barriers to dispersal are not effective in preventing organisms from dispersing. Therefore, they do not play any role in the distribution of microorganisms, and only environmental filters operate on microbial communities. This viewpoint has been developed further and applied specifically to protists (Fenchel *et al.*, 1997; Finlay, 2002; Finlay *et al.*, 2004). However, the existence of "flagship" (i.e, morphologically conspicuous), species of microorganisms that have geographically restricted distributions, contradicts this view (Foissner, 2006; Smith & Wilkinson, 2007; Vyverman *et al.*, 2007; Heger *et al.*, 2011a). In addition, increasing evidence for the existence of substantial cryptic protist diversity suggests that even supposedly

cosmopolitan free-living taxa might in fact correspond to complexes of genetically distinct biological species, each of which may potentially have a restricted distribution (Darling *et al.*, 2007; Aurahs *et al.*, 2009; Casteleyn *et al.*, 2010; Watts *et al.*, 2011; Heger *et al.*, 2013). The idea that barriers to dispersal do not have any effect on diversity as suggested by the EiE hypothesis is at one extreme of a range of possible scenarios: in this case only ecological filtering would affect community composition. The question remains open whether pure spatial contingencies also have an effect, and if they do, to what extent.

To address this general question, we used as a model group the Euglyphida Cope, a monophyletic clade of predominantly bacterivorous testate amoebae that build a self-secreted siliceous test (shell) (Meisterfeld, 2002). Euglyphids are ubiquitous soil protists, found under a wide range of environmental conditions, some species even being psychrophilic (Smith, 1992; Santibañez *et al.*, 2011). We surveyed their molecular diversity in forest litter and mosses from 35 sites covering a broad range of climates, from all continents except Antarctica, using a specific PCR protocol to amplify selectively euglyphid SSUrRNA genes from environmental DNA extracts, an approach coined metabarcoding (Pompanon *et al.*, 2011). These molecular methods provide an alternative approach to the long-running debates on the validity of morphological criteria that have bedevilled the study of protist diversity (Finlay *et al.*, 2004; Mitchell & Meisterfeld, 2005; Heger, 2009). We evaluated the phylotype composition obtained at the different sites and determined to what extent it could be predicted based on soil characteristics, macroclimatic variables and the geographical position of the sampling sites. In addition, we determined which variables were most correlated with diversity and community structure and compared these patterns to existing data and theory derived from the study of macroscopic organisms.

MATERIALS AND METHODS

Sampling and environmental data

126 Samples of soil litter (upper 3 cm composed essentially of organic matter) and mosses (growing on
127 the soil surface) were collected from 35 sites in forest ecosystems covering most biogeographical and
128 bioclimatic regions of the world (Fig. 1; see also Appendix S1 in Supporting Information). The
129 coordinates of each sampling location were recorded using a field GPS. We used two key metrics to
130 characterize soil chemistry: pH as a key factor explaining soil testate amoeba biodiversity and
131 community structure (Bonnet, 1964), and the C/N ratio as a measure of organic matter
132 decomposition (Bardgett, 2005). Total C and N contents determined by CHN analyser (CHN EA1109-
133 Elemental analyser, Carlo Erba Instruments) were used to calculate the C/N ratio and pH was
134 measured in a 1:1 aqueous slurry.

135 We computed spatial variables to investigate the geographical patterns of euglyphid diversity
136 and their possible relationship with known biogeographical barriers. These variables represent
137 spatial structures, from local to global scales, and can be easily used as spatial explanatory variables
138 in models and regression analyses. We computed these variables, based on the geographical
139 coordinates of the sampling points, following the method of Borcard & Legendre (2002) and Borcard
140 *et al.* (2004). We decomposed a Euclidean distance matrix among sites into 41 distance-based
141 Moran's Eigenvector Maps (db-MEM). We then chose the two db-MEMs that were most strongly
142 correlated with community composition using forward selection. These two variables - db-MEM6 and
143 db-MEM31 - were kept for further analyses. We used db-MEM as spatial descriptors because this
144 approach uses a multi-scale decomposition of a distance matrix that is more likely to capture the
145 relevant spatial structures (e.g. biogeographical barriers) than methods using distances based on raw
146 coordinates alone (Dray *et al.*, 2006).

147 We used the plot coordinates to extract biologically relevant bioclimatic variables from the
148 30 arc-second resolution grids of the WorldClim project (Hijmans, 2005). These variables comprise
149 various metrics based on monthly temperature and precipitation data that are biologically relevant
150 for fauna and flora (Elith *et al.*, 2006), protosteloid amoebae (Aguilar *et al.*, 2011), and, as

hypothesized here, for Euglyphida. In order to reduce dimensionality, we then conducted a principal components analysis (PCA) of these data and used the obtained principal components as climatic variables in subsequent analyses (clim_PC1 and clim_PC2). Using PCA axes instead of environmental variables enables the consideration of multiple climatic variables at the same time and summarizes them into two synthetic variables that cover a greater part of the variance in climatic conditions than temperature and precipitation alone. Overall, we had three groups of two environmental variables representing spatial (db-MEM6 and db-MEM31), climatic (clim_PC1 and clim_PC2) and soils (pH and C/N) aspects, respectively.

Here Figure 1

DNA extraction, PCR, sequencing and phylogenetic analysis

DNA was extracted using, in combination, a MoBio Power SoilTM DNA extraction kit (Carlsbad, CA, USA) and a bead-beating apparatus (FP120 FastPrepTM cell disruptor, Savant Instruments, Inc., Hotbrook, NY). A total of 0.25g of sample and 60 µl of C1 solution were added to the Powerbead tube. The tube was inverted several times and then shaken for 30s at 5.5m.s^{-1} in the FastPrepTM cell disruptor as an alternative to the vortexing step recommended by the manufacturer. The other steps of the protocol followed the manufacturer's instructions. We chose to use a "classical" PCR, cloning and sequencing approach rather than next generation sequencing methods because longer sequences, and higher quality control, were deemed essential for building an accurate phylogenetic tree and to discriminate closely related phylotypes. This approach is optimal for the study of a given phylogenetic group such as the Euglyphida.

Amplification of the small subunit rRNA gene was performed in two steps; a first PCR was achieved with the specific primers EuglySSUF (forward) (5' GCGTACAGCTCATTATATCAGCA 3') and EuglyLSUR (reverse) (5' GTTTGGCACCTTAACTCGCG 3'), the latter primer placed on the LSU rRNA gene. The cycling profile was as follows: an initial denaturation at 94°C for 5 minutes, and then 40

cycles with 94°C for 15 seconds as the denaturation step, 62°C for 15 seconds as the primer annealing step, with a touchdown of 1°C per cycle for the eight first cycles, and 72°C for 150 seconds as an elongation step. The final elongation step was of 10 minutes at 72°C. A second, semi-nested PCR was carried out, again using EuglySSUF in combination with EuglySSUR (reverse) (5' GCACCACCACCCATAGAATCWAGAAAGATC 3'), with an initial denaturation at 94 °C for 3 minutes, and then 30 cycles with a denaturation step of 94°C for 30 seconds, then an annealing step at 59°C for 30s and an elongation step at 72°C for 60 seconds, followed by a final elongation at 72°C for 10 minutes. PCR reactions were carried out in 50 µl of reaction buffer containing 1 µl DNA template (around 1–5 ng), 1.5mM MgCl2, dNTPs (10 nmol each), 20 pmol of each primer, and 1 U TaqDNA polymerase (Promega). The resulting amplicon was 1100 bp long and comprised the variable region v4; it spanned approximately the two first thirds of the entire SSU rRNA gene.

Amplicons were cloned into pCR2.1 Topo TA cloning vector (Invitrogen) and transformed into *E. coli* TOP10' One Shot cells by heat shock (Invitrogen). Cells were spread onto LB agar medium containing 50µg/ml ampicillin and X-gal and IPTG according to the manufacturer's instructions for blue-white screening. Colonies were picked and the insert was amplified using PCR primers EuglySSUF and EuglySSUR. The presence of the expected size insert was checked on the white colonies by PCR amplification. Clone inserts were amplified with vector primers M13F and M13R and inserts from the expected size were sequenced directly with the specific primers EuglySSUF and EuglyLSUR. Between 18 and 73 clones per site were sequenced depending on the diversity encountered.

The obtained sequences were trimmed for ambiguities and aligned manually using the software BioEdit v. 7.0.9.0 (Hall, 1999). Chimerical sequences were eliminated by careful observation of group-specific signature sequences, as suggested by Berney *et al.* (2004). Each difference in a single nucleotide was considered as yielding a distinct phylotype in further analyses; community composition was expressed as a percentage of the total sequences, which has been shown to be well

correlated with individual abundances in other protist groups, such as marine Stramenopiles MAST-4 (Rodriguez-Martinez *et al.*, 2009) and rotaliid foraminiferans (Pawlowski & Weber, 2013). Clones were sequenced until saturation was reached: this was established using the software DOTUR (Schloss & Handelsman, 2005). Phylotype sequences were aligned with sequences derived from GenBank (for accession files please refer to Fig. 3). The phylogenetic tree was built using RAxMLv7.2.8 (Stamatakis *et al.*, 2008), as proposed on the Black Box portal (http://phylobench.vital-it.ch/raxml-bb/) using the GTR+Γ+I model and was performed on 994 characters. The tree was rooted with non-euglyphid Cercozoa taken from a wide array of Monadofilosa. Clone sequences have been deposited in GenBank under the names KT272446-KT272698, and KP892886-KP892888.

Euglyphid α *and* β *diversity*

We characterized euglyphid testate amoeba α–diversity using phylotype richness (total number of phylotypes per site) and the Shannon (Shannon & Weaver, 1949) and Simpson (Simpson, 1949) indices of diversity. We chose these measures because they are broadly used in ecological research and describe complementary aspects of α–diversity.

To characterize β-diversity, we measured the pairwise phylogenetic distances among euglyphid communities using a generalized version of UniFrac (Lozupone & Knight, 2005), called GUniFrac (Chen *et al.*, 2012), which allows modulation of the relative weight given to both rare and over- represented sequences. While UniFrac is now used routinely in the field of environmental microbiology (Lozupone & Knight, 2007; Lauber *et al.*, 2009), GUniFrac is its latest development and has been shown to have higher detection power than UniFrac or other measures of β-diversity (Chen et al. 2012). We calculated GUniFrac for $\alpha = 0$, $\alpha = 0.5$, $\alpha = 1$ as well as the unweighted and variance adjusted weighted version of UniFrac to detect any changes in euglyphid communities diversity.

Numerical analyses

To investigate the general biodiversity of euglyphid testate amoebae, we first computed rank abundance and species accumulation curves. From the species accumulation curve, we estimated the potential total size of the phylotype pool by comparing the obtained curve to simulated curves computed using various models (i.e. Chao, first and second order jack-knife (Smith, 1984; Chao, 1987)). Moreover, we computed rank-abundance curves and fitted commonly used models of abundance distribution (null or broken-stick, pre-emption, lognormal, and Mandelbrot; (Wilson, 1991)). The curves generated were compared visually.

We then investigated the correlation of euglyphid phylotype richness and Simpson and Shannon diversity with each selected environmental variable using generalized additive models (GAM). These models allow the discrimination of non-linear relationships and are commonly used in ecological research (Guisan *et al.*, 2002). These analyses allowed us to determine which variables were correlated with the α–diversity of euglyphids.

We finally used Permutational multivariate analysis of variance (PermanovaG), using the multiple distance matrices produced by GUniFrac, to test whether the soil, climatic and spatial variables influenced euglyphid β-diversity (Chen *et al.*, 2012). PermanovaG combines multiple distance matrices in a single test and thereby does not require an *a priori* knowledge of the type of changes of community composition.

All calculations were carried out within the R framework (R Development Core Team 2011) using packages “GUniFrac” (http://cran.r-project.org/web/packages/GUniFrac/index.html), “PCNM” (http://R-Forge.R-project.org/projects/sedar/) and “vegan” (http:// cran.r-project.org/package=vegan).

RESULTS

Euglyphid diversity

We obtained a total of 245 different euglyphid phylotypes. The number of phylotypes per sample varied between 1 (DR1, Dominican Republic) and 31 for the most diverse location (I2, India). The species accumulation curve and rarefaction analyses indicated that the global diversity of phylotypes did not attain saturation (Fig. 2a and b), indicating that total euglyphid diversity in forest litters is significantly higher than the total number of phylotypes recovered in this study. The flattening of the rarefaction curves was reached at 266, 351, and 557 phylotypes, using the Bootstrap, first order Jacknife, and Chao estimates, respectively. The rank abundance curve appeared different from the predicted null model curve (i.e. that should have been observed if the community composition was under neutral selection; see Fig. 2b), suggesting strong niche effects.

-Here Figure 2a and b-

The phylogenetic tree built on the clone sequences together with sequences retrieved from GenBank showed that the 245 phylotypes included representatives of all known euglyphid families except Cyphoderiidae (Lara *et al.*, 2007), a group that is associated with freshwater and marine intertidal habitats (Meisterfeld, 2002). Three phylotypes (CH2_2_11, Ma_44 and CD3_3) clustered together within the marine (and marginally freshwater) family Paulinellidae (Meisterfeld, 2002), a clade that has been reported recently for the first time from soils (Tarnawski & Lara, 2015). In addition, four well-supported clades (named here EEC1 to EEC4) appeared as new families represented only by environmental clone sequences obtained from this and previous studies (Fig. 3). Additional sequences retrieved from GenBank appear also on Fig. 3.

-Here Figure 3-

Environmental variables

For spatial aspects, db-MEM6 and 31 were retained by the forward selection procedure. For climatic aspects, principal components clim_PC1 and clim_PC2 represented ~70% of the total variance of the climatic data. Clim_PC1 (40.1% of variance) was explained mostly by variables associated with

temperature seasonality, including isothermality and coldest month temperature, whereas clim_PC2 (28.5%) was correlated with warmest month temperatures and mean yearly temperature (see Appendix S2 in Supporting information). Moreover, pH and C/N were negatively correlated with each other (see Appendix S3 in Supporting information).

Biodiversity patterns (α-diversity)

Richness, defined as the number of phylotypes, and diversity as expressed as either Shannon or Simpson indices were strongly correlated (r=0.92, P<0.001, Pearson correlations), therefore only phylotype richness will be discussed hereafter. Variations in phylotype richness were explained by the eigenvector map db-MEM6, a pattern that follows closely a latitudinal division of the Earth (P< 0.01, r^2=54.2%), but also by the climatic principal component clim_PC 2 (P=0.02, r^2=42.9%) and by C/N ratio (P=0.02, r^2=32.8%, Table 1, Fig. 4). Among the variables selected for clim_PC2, temperature of the warmest months was highly correlated with α-diversity (P>0.01, r^2=52.9%: Fig 5).

-Here Figure 4-

Community patterns and environmental drivers (β-diversity)

Euglyphid community composition was most strongly correlated with climatic principal component clim_PC1, which was in turn explained mostly by isothermality, and temperatures in the coldest month of the year (see Appendix S2 in Supporting information). The two distance-based Moran's Eigenvector Maps (db-MEM) models showing the best fit in the RDA selection procedure were db-MEM6 and db-MEM31.

The GUniFrac analysis showed that db-MEM6 and db-MEM31 (Table 1) were significantly correlated with the communities. While db-MEM6 (i.e. equivalent to a latitudinal gradient; Fig. 5) was obviously correlated with climate, db-MEM31 (i.e. showing a separation between North and

South of the Cancer tropic desert belt; Fig. 5) was not. Community composition was also significantly correlated with climatic PC1 (i.e. precipitation seasonality; $P< 0.01$), as well as with pH ($P< 0.01$).

-Here Figure 5- and -Table 1-

DISCUSSION

General considerations about euglyphid diversity in forest soils

Our data support the view that protists are highly diverse. Eukaryotic diversity still contains major clades only revealed by environmental DNA (eDNA) surveys (López-Garcia *et al.*, 2001; Massana *et al.*, 2004; Lara *et al.*, 2010), that have barely been investigated: most of these clades are protists. Indeed in this study we discovered four new major (family level, *sensu* Lara *et al.,* 2007) clades of euglyphids for which no data exist, either from cultured or from isolated taxa. In addition, as the barcoding gene used here, SSU rRNA, is highly conserved, the real specific diversity is very likely higher than the 245 phylotypes identified in our study. For instance, the mitochondrial cytochrome oxidase gene (COI) commonly used in animal studies is 3–5 times more variable than SSU rRNA in Euglyphida and has been shown to perform better in species level discrimination (Lara *et al.,* 2011; Heger *et al.*, 2011). Although some of the new clades may include taxa previously described morphologically, but lacking molecular data (e.g. family Psammonobiotidae), we consider it unlikely that all four clades belong to such taxa. The discovery in terrestrial habitats (i.e. non-wetland forests) of sequences related to the mostly marine or freshwater Paulinellidae (Nicholls, 2009) is also noteworthy and further illustrates the usefulness of eDNA surveys. This is a surprising finding, because the salinity barrier that separates freshwater and marine environments is difficult to cross for protists (Logares *et al.*, 2009), and soil protist communities are very distinct from those of aquatic environments (Foissner, 1987).

In this study, we chose to use a “conventional” molecular approach to screen richness, i.e. cloning and (Sanger) sequencing, in order to retrieve more phylogenetic information that could then

be used to build trees. The use of newer, high throughput sequencing strategies would probably have yielded higher richness. However, we believe that the observed patterns of diversity would have been the same, as our approach allowed the retrieval of the most dominant taxa.

Latitudinal gradient of diversity

Our data clearly support the existence of a latitudinal gradient of diversity in soil euglyphid testate amoebae. When richness per sample is plotted against latitude (and if sites above 500 m a.s.l. are removed as having atypical climates for their latitude, c.f. Ju *et al.*, 2014), these data show a unimodal distribution peaking at low latitudes (Fig. 4). In addition, the eigenvector map db-MEM6 (mostly a broad scale latitudinal pattern) was highly correlated with diversity and explained a large proportion of the distribution of phylotypes. The BIOCLIM variable that was best correlated with diversity was temperature of the warmest month (Fig. 4). This may seem counterintuitive, as soil protists are expected to be stressed during the hottest part of the year. However, as our sampling design included only forest soil sites that were shaded by trees, soil temperatures probably do not reach values that can be harmful to euglyphids. Possibly more importantly, it is also likely that humidity is generally sufficient in soil forest litter, thus promoting euglyphid activity even when absolute temperatures are high. Indeed, as long as water availability is not a limiting factor, high temperatures can increase enzymatic activity and, therefore, primary production. This is a likely explanation why only temperature, and not precipitations amount and/or regularity were correlated with species richness, as opposed to the predictions of the water-energy theory (O'Brien, 2006). The species–energy theory (Hawkins *et al.*, 2003) postulates that warm climates support higher individual numbers because of higher productivity, and therefore extinction rates are lower than in colder climates, thus diversity accumulates.

In addition, richness appears to be inversely correlated with C/N values, which is a commonly used measure of the speed of organic matter turnover, which is directly related to soil productivity

(Bardgett, 2005). Therefore, a low C/N value is associated with high biochemical energy, which suggests that the species–energy model may apply to soil euglyphid diversity. It has previously been argued that the species–energy paradigm convincingly explains latitudinal diversity gradient in marine organisms (Tittensor *et al.,* 2010). A convincing latitudinal gradient has also been found in marine bacteria (Fuhrman *et al.*, 2008) but not in a recent study of aquatic testate amoebae across China (Ju *et al.*, 2014). In soil bacteria, however, a latitudinal effect was not found; instead higher bacterial diversity was correlated with low C/N ratio (Fierer & Jackson, 2006). Among soil microbial eukaryotes, a similar latitudinal gradient was observed for dictyostelid amoebae (Swanson *et al.*, 1999; Perrigo *et al.*, 2013; Stephenson & Feest, 2013) and fungi (Treseder *et al.,* 2014), and higher yeast diversity was correlated with high temperatures (Vishniac, 2006). Protist-sized Metazoa (rotifers) from genus *Keratella* also exhibit higher diversity in the tropics, together with a high degree of local endemism (Seger & De Smet, 2008). Therefore a provisional conclusion would be that the latitudinal gradient of diversity applies to some microbial-sized organisms, but not to others. Life history traits probably explain these differences, but geographically uneven sampling efforts may also be important (Fontaneto *et al.*, 2012).

Non-cosmopolitanism of euglyphid phylotypes

Geographical variables such as the Equator/Cancer tropic desert belt (db-MEM31; Fig. 5) and a latitudinal gradient (db-MEM6; Fig. 5) significantly explained the geographical distribution of phylotypes (see also Table 1). While db-MEM6 is strongly correlated with temperatures (Fig. 3) and is therefore not independent of environmental parameters, db-MEM31 is not correlated with the other measured variables (i.e. pH and C/N). As pH and the C/N ratio often explain best the distribution of soil organisms (Ponge *et al.,* 1997; Ponge, 2003), db-MEM31 is therefore likely to be a primarily a purely spatial variable; it corresponds to the barrier of deserts surrounding the tropic of Cancer, which is also a major biogeographical limit for arcellinid testate amoeba genera such as *Apodera*, *Alocodera* and *Certesella* (Smith & Wilkinson, 2007; Smith *et al.*, 2008). This barrier is arguably

caused by the main wind regimes, which permit passive dispersal only along similar latitudes but prevent the easy crossing of the equator, even for small-sized organisms (Wilkinson *et al.*, 2012). These different lines of evidence therefore suggest that environmental barriers prevent testate amoeba phylotypes from spreading worldwide, thus allowing allopatric speciation to occur.

The distribution of euglyphid phylotypes was also shown to be correlated with other variables. The correlation between β-diversity and climate evenness (PC1) probably indicates that only certain phylotypes can tolerate extremes such as deep frost (and indeed minimum temperature of the coldest month was also strongly correlated with PC1). In this case, the ability of fast encystment (i.e. to enter a dormant stage) may play a crucial role in survival. Likewise, pH was significantly correlated with community composition, indicating the existence of specialists in acidic and/or alkaline substrates. Soil or water pH is known as one of the major factors explaining the structure of testate amoeba communities, together with moisture (Lamentowicz & Mitchell, 2005; Mitchell *et al.*, 2008). For example, in our survey, phylotype CH4_II20 (whose sequence is identical to *Assulina muscorum* AJ418791) has been found only in sites with marked seasonality, where it sometimes represents a large part of all phylotypes (in some sites more than 70% of the total number of sequences). The existence of specialist phylotypes is further corroborated by the rank-abundance curve of soil euglyphid data, which significantly diverged from the null model (Fig. 1). This shows that deterministic forces influence community composition, in contrast to the pattern found in some microbial Metazoa such as moss-dwelling rotifers (Fontaneto *et al.*, 2011). If niche effects influence community composition, evolving towards specialisation could be a winning strategy for euglyphids.

CONCLUSION

Our study revealed the existence of an unexpectedly high diversity of euglyphid testate amoebae, both at deep phylogenetic levels (i.e. presence of undetected clades) and within individual clades. Phylotype richness was significantly higher in low latitude, high-energy environments (i.e. with high

temperatures and fast nutrient cycling). In addition, soil euglyphid diversity and community structure were explained by a combination of both geographic isolation and ecological specialization (i.e. niche-driven community patterns). The geographical isolation observed in euglyphid testate amoebae is in clear contradiction to the cosmopolitan distribution models, which were believed to apply to free-living microorganisms. Our results from forest litter euglyphid testate amoebae thus suggest that the patterns of diversity and community structure of certain protists are rather similar to those observed for multicellular organisms. We predict that a more thorough investigation of the ecology of different groups of protists, taking into account their true diversity (i.e. beyond morphotypes) will reveal many similar cases. More generally, our study suggests that the rules that govern soil protist diversity are similar to those for larger organisms. There may be a unity to ecology that crosses the boundary created by the limitations of human vision that separates the macroscopic from the microbial world.

Acknowledgements

Several people provided samples that were used in this study: Anush Kosakyan, Pilar and Thomas Junier, Hilda Hollenstein, Thierry J. Heger, Béatrice Lecroq, Jan Pawlowski, Jason R. Grant, Gaëtan Martin, Alexandre Buttler, Martin Clerc, Luc Scherrer, Michalis and Yvonne Mouzakitis, Daniel Petitjean, Josée Blain and Laure Desplands. We are also indebted to Edita Elias, José Luis Rodríguez, Yuri Mazei, Anatoly Bobrov, Dimaris Acosta Mercado, Amelia L. Mateo Jiménez, Carlos J. Santos Flores, and Mariusz Lamentowicz for help in the field and Thibaud Goetschi for technical help. We thank Hannah O'Regan for editorial assistance and David Singer for the advice and help with the figures. Financial support was provided by an Ambizione fellowship from the Swiss National Science Foundation (PZ00P2_122042) to EL, the Science and Technology Cooperation Program Switzerland – Russia grant IZLR Z3_128338 to EM & EL, and the Swiss National Science Foundation (Scientific & Technological Cooperation Programme Switzerland Russia faculty exchange project) to EM.

415 **References**

416 Adams, J.M. & Woodward, F.I. (1989) Patterns in tree species richness as a test of the glacial
417 extinction hypothesis. *Nature*, **339**, 699-701.

418 Aguilar, M., Spiegel, F.W. & Lado, C. (2011) Microhabitat and climatic preferences of protosteloid
419 amoebae in a region with a Mediterranean climate. *Microbial Ecology*, **62**, 361-373.

420 Allen, A.P., Gillooly, J.F., Savage, V.M. & Brown, J.H. (2006) Kinetic effects of temperature on rates of
421 genetic divergence and speciation. *Proceedings of the National Academy of Sciences USA*,
422 **103**, 9130-9135.

423 Aurahs, R., Grimm, G.W., Hemleben, V., Hemleben, C. & Kucera, M. (2009) Geographical distribution
424 of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber*. *Molecular*
425 *Ecology*, **18**, 1692-1706.

426 Baas Becking, L.G.M. (1934) *Geobiologie of inleiding tot de milieukunde*. W.P. Van Stockum & Zoon,
427 The Hague, the Netherlands.

428 Bardgett, R. (2005) *The biology of soil*. Oxford University Press, Oxford.

429 Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R. & Fierer, N. (2013)
430 Global biogeography of highly diverse protistan communities in soil. *ISME Journal*, **7**, 652-
431 659.*

432 Berney, C., Fahrni, J. & Pawlowski, J. (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and
433 limitations of environmental DNA surveys *BMC Biology*, **2**, 13.

434 Boltovskoy, D., Kling, S.A., Takahashi, K. & Bjorklund, K. (2010) World atlas of distribution of recent
435 Polycistina (Radiolaria). *Palaeontologia Electronica*, **13,** 1-229.

436 Bonnet, L. (1964) Le peuplement thécamoebien des sols. *Revue d'Écologie et de Biologie du Sol*, **1**,
437 123-408.

438 Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal
439 coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51-68.

440 Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004) Dissecting the spatial structure of
441 ecological data at multiple scales. *Ecology*, **85**, 1826-1832.
442 Buzas, M.A., Collins, L.S. & Culver, S.J. (2002) Latitudinal difference in biodiversity caused by higher
443 tropical rate of increase. *Proceedings of the National Academy of Sciences USA*, **99**, 7841-
444 7843.
445 Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A.E., Kotaki, Y., Rhodes, L., Lundholm, N., Sabbe, K. &
446 Vyverman, W. (2010) Limits to gene flow in a cosmopolitan marine planktonic diatom.
447 *Proceedings of the National Academy of Sciences USA*, **107**, 12952-12957.
448 Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability.
449 *Biometrics*, **43**, 783–791.
450 Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., Collman, R.G., Bushman, F.D. &
451 Li, H.Z. (2012) Associating microbiome composition with environmental covariates using
452 generalized UniFrac distances. *Bioinformatics*, **28**, 2106-2113.
453 Cox, C.B. & Moore, P.D. (2010) *Biogeography*. 8th ed. John Wiley and Sons, Hoboken, NJ.
454 Darling, K.F., Kucera, M. & Wade, C.M. (2007) Global molecular phylogeography reveals persistent
455 Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National*
456 *Academy of Sciences USA*, **104**, 5002-5007.
457 Dray, S., Legendre, P., & Peres-Neto, P. R. (2006) Spatial modelling: a comprehensive framework for
458 principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling,* **196**, 483-
459 493.
460 Ekelund, F. & Rønn, R. (1994) Notes on protozoa in agricultural soil with emphasis on heterotrophic
461 flagellates and naked amoebae and their ecology. *FEMS Microbiology Reviews*, **15**, 321-353.
462 Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F.,
463 Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C.,
464 Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K.,
465 Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E.

(2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Fenchel, T., Esteban, G.F. & Finlay, B.J. (1997) Local versus global diversity of microorganisms: cryptic diversity of ciliated protozoa. *Oikos*, **80**, 220-225.

Fierer, N. & Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences USA*, **103**, 626-631.

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

Finlay, B.J., Esteban, G.F. & Fenchel, T. (2004) Protist diversity is different? *Protist*, **155**, 15-22.

Foissner, W. (1987) Soil protozoa: Fundamental problems, ecological significance, adaptation in ciliates and testaceans, bioindicators, and guide to the literature. *Progress in Protozoology*, **2**, 69-212.

Foissner, W. (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozoologica*, **45**, 111-136.

Fontaneto, D., Barbosa, A.M., Segers, H. & Pautasso, M. (2012) The 'rotiferologist' effect and other global correlates of species richness in monogonont rotifers. *Ecography*, **35**, 174-182.

Fontaneto, D., Westberg, M. & Hortal, J. (2011) Evidence of weak habitat specialisation in microscopic animals. *Plos One*, **6** : e23969

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L. & Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences USA*, **105**, 7774-7778.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89-100.

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

490 Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach,
491 G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and
492 broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105-3117.
493 Heger, T.J., Lara, E. & Mitchell, E.A.D. (2011a) Arcellinida testate amoebae (Arcellinida: Amoebozoa):
494 model of organisms for assessing microbial biogeography. *The importance of being small:*
495 *does size matter in biogeography?* (ed. by D. Fontaneto). pp. 111-129. Cambridge University
496 Press, Cambridge.
497 Heger, T.J., Booth, R.K., Sullivan, M.E., Wilkinson, D.M., Warner, B.G., Asada, T., Mazei, Y.,
498 Meisterfeld, R. & Mitchell, E.A.D. (2011b) Rediscovery of *Nebela ansata* (Amoebozoa:
499 Arcellinida) in eastern North America: biogeographical implications. *Journal of Biogeography*,
500 **38**, 1897-1906.
501 Heger, T.J., Pawlowski, J., Golemanski, V., Todorov, M., Lara, E. & Mitchell, E.A.D. (2011c) Comparing
502 potential COI and SSU rDNA barcodes for assessing the diversity and phylogenetic
503 relationships of cyphoderiid testate amoebae (Rhizaria: Euglyphida). *Protist*, **162,** 131-141.
504 Heger, T.J., Mitchell, E. A. D. & Leander, B. S. (2013) Holarctic phylogeography of the testate amoeba
505 *Hyalosphenia papilio* (Amoebozoa: Arcellinida) reveals extensive genetic diversity explained
506 more by environment than dispersal limitation. *Molecular Ecology*, **22,** 5172-5184.
507 Heger, T.J., Mitchell, E. A. D., Ledeganck, P., Vincke, S., Van De Vijver, B. & Beyens, L. (2009) The curse
508 of taxonomic uncertainty in biogeographical studies of free-living terrestrial protists: a case
509 study of testate amoebae from Amsterdam Island. *Journal of Biogeography*, **36**, 1551-1560.
510 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution
511 interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**,
512 1965-1978.
513 Hillebrand, H. (2004) On the generality of the latitudinal diversity gradient. *The American Naturalist*,
514 **163**, 192-211.

515 Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: Evolutionary dynamics of the
516 latitudinal diversity gradient. *Science*, **314**, 102-106.
517 Ju, L.H., Yang, J., Liu, L.M. & Wilkinson, D.M. (2014) Diversity and distribution of freshwater testate
518 amoebae (Protozoa) along latitudinal and trophic gradients in China. *Microbial Ecology*, **68**,
519 657-670
520 Lamentowicz, M. & Mitchell, E.A.D. (2005) The ecology of testate amoebae (protists) in *Sphagnum* in
521 North-Western Poland in relation to peatland ecology. *Microbial Ecology*, **50**, 48-63.
522 Lara, E. & Acosta-Mercado, D. (2012) A molecular perspective on ciliates as soil bioindicators.
523 *European Journal of Soil Biology*, **49**, 107-111.
524 Lara, E., Heger, T.J., Scheihing, R. & Mitchell, E.A.D. (2011) COI gene and ecological data suggest size-
525 dependent high dispersal and low intra-specific diversity in free-living terrestrial protists
526 (Euglyphida; *Assulina*). *Journal of Biogeography,* **38,** 640-650.
527 Lara, E., Moreira, D. & Lopez-Garcia, P. (2010) The environmental clade LKM11 and *Rozella* form the
528 deepest branching clade of Fungi. *Protist*, **161**, 116-121.
529 Lara, E., Heger, T.J., Mitchell, E.A.D., Meisterfeld, R. & Ekelund, F. (2007) SSU rRNA reveals a
530 sequential increase in shell complexity among the Euglyphid testate amoebae (Rhizaria:
531 Euglyphida). *Protist*, **158**, 229-237.
532 Lauber, C.L., Hamady, M., Knight, R. & Fierer, N. (2009) Pyrosequencing-based assessment of soil pH
533 as a predictor of soil bacterial community structure at the continental scale. *Applied and*
534 *Environmental Microbiology*, **75**, 5111-5120.
535 Logares, R., Bråte, J., Bertilsson, S., Clasen, J., Shalchian Tabrizi, K., & Rengefors, K. (2009) Infrequent
536 marine-freshwater transitions in the microbial world. *Trends in Microbiology*, **17**, 414-
537 422.López-Garcia, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. (2001) Unexpected
538 diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, **409**, 603-607.
539 Lozupone, C. & Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial
540 communities. *Applied and Environmental Microbiology,* **71**, 8228-8235.

Lozupone, C.A. & Knight, R. (2007) Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences USA*, **104**, 11436-11440.

Lueke, C., Frenzel, P., Ho, A., Fiantis, D., Schad, P., Schneider, B., Schwark, L. & Utami, S.R. (2014) Macroecology of methane-oxidizing bacteria: the beta-diversity of pmoA genotypes in tropical and subtropical rice paddies. *Environmental Microbiology*, **16**, 72-83.

Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K. & Pedrós-Alió, C. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Applied and Environmental Microbiology*, **70**, 3528-3534.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Ovreas, L., Reysenbach, A.L., Smith, V.H. & Staley, J.T. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102-112.

Meisterfeld, R. (2002) Testate amoebae with filopodia. *The illustrated guide to the protozoa* (ed. by J.J. Lee, G.F. Leedale and P. Bradbury), pp. 1054-1084. Society of protozoologists, Lawrence, Kansas, USA.

Mitchell, E.A.D. & Meisterfeld, R. (2005) Taxonomic confusion blurs the debate on cosmopolitanism versus local endemism of free-living protists. *Protist*, **156**, 263-267.

Mitchell, E.A.D., Charman, D.J. & Warner, B.G. (2008) Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future. *Biodiversity and Conservation*, **17**, 2115-2137.

Naff, C.S., Darcy, J.L. & Schmidt, S.K. (2013) Phylogeny and biogeography of an uncultured clade of snow chytrids. *Environmental Microbiology*, **15**, 2672-2680.

Nicholls, K.H. (2009) Six new marine species of the genus *Paulinella* (Rhizopoda: Filosea, or Rhizaria: Cercozoa). *Journal of the Marine Biological Association of the United Kingdom*, **89**, 1415-1425.

566 O'Brien, E.M. (2006) Biological relativity to water-energy dynamics. *Journal of Biogeography*, **33**,
567 1868-1888.
568 Old, K.M. & Oros, J. M. (1980) Mycophagous amoebae in Australian forest soils. *Soil Biology and*
569 *Biochemistry*, **12**, 169-175.
570 Pawlowski, J. & Weber, A.A.T. (2013) Can abundance of protists be inferred from sequence data: a
571 case study of Foraminifera. *PLos One*, **8**, e56739.
572 Perrigo, A.L., Baldauf, S.L. & Romeralo, M. (2013) Diversity of dictyostelid social amoebae in high
573 latitude habitats of Northern Sweden. *Fungal Diversity*, **58**, 185-198.
574 Pompanon, F., Coissac, E. & Taberlet, P. (2011) Metabarcoding a new way to analyze biodiversity.
575 *Biofutur*, **319**, 30-32.
576 Ponge, J.F., Arpin, P., Sondag, F. & Delecour, F. (1997) Soil fauna and site assessment in beech stands
577 of the Belgian Ardennes. *Canadian Journal of Forest Research-Revue Canadienne de*
578 *Recherche Forestière*, **27**, 2053-2064.
579 Ponge, J.F. (2003) Humus forms in terrestrial ecosystems: a framework to biodiversity. *Soil Biology &*
580 *Biochemistry*, **35**, 935-945.
581 Porazinska, D.L., Giblin-Davis, R.M., Esquivel, A., Powers, T.O., Sung, W. & Thomas, W.K. (2010)
582 Ecometagenetics confirms high tropical rainforest nematode diversity. *Molecular Ecology*, **19**,
583 5521-5530. Rodriguez-Martinez, R., Labrenz, M., del Campo, J., Forn, I., Jurgens, K. &
584 Massana, R. (2009) Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake
585 Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environmental*
586 *Microbiology*, **11**, 397-408.
587 Santibañez, P.A., Kohshima, S., Scheihing, R.A., Silva, R., Jaramillo, J.I., Labarca, P.J. & Casassa, G.
588 (2011) First record of testate amoebae on glaciers and description of a new species
589 *Puytoracia jenswendti nov. sp.* (Rhizaria, Euglyphida). *Acta Protozoologica*, **50**, 1-14.

590 Schloss, P.D. & Handelsman, J. (2005) Introducing DOTUR, a computer program for defining
591 operational taxonomic units and estimating species richness. *Applied and Environmental*
592 *Microbiology*, **71**, 1501-1506.
593 Segers, H. & De Smet, W.H. (2008) Diversity and endemism in Rotifera: a review, and *Keratella* Bory
594 de St Vincent. *Biodiversity and Conservation*, **17**, 303-316.
595 Shannon, C.E. & Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois
596 Press, Urbana.
597 Sherratt, T.N. & Wilkinson D.M. (2009) *Big questions in ecology and evolution.* Oxford University
598 Press, Oxford.
599 Simpson, E.H. (1949) Measurement of diversity. *Nature*, **163**, 688.
600 Smith, E.P. & van Belle, G. (1984) Nonparametric estimation of species richness. *Biometrics*, **40**, 119-
601 129.
602 Smith, H.G. (1992) Distribution and ecology of the testate rhizopod fauna of the continental Antarctic
603 zone. *Polar Biology*, **12**, 629-634.
604 Smith, H.G. & Wilkinson, D.M. (2007) Not all free-living microorganisms have cosmopolitan
605 distributions - the case of *Nebela* (*Apodera*) *vas* Certes (Protozoa : Amoebozoa : Arcellinida).
606 *Journal of Biogeography*, **34**, 1822-1831.
607 Smith, H.G., Bobrov, A. & Lara, E. (2008) Diversity and biogeography of testate amoebae. *Biodiversity*
608 *and Conservation*, **17**, 329-343.
609 Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A Rapid Bootstrap Algorithm for the RAxML Web
610 Servers. *Systematic Biology*, **57**, 758-771.
611 Stephenson, S.L. & Feest, A. (2013) Ecology of soil Eumycetozoans. *Acta Protozoologica*, **51**, 201-208.
612 Swanson, A.R., Vadell, E.M. & Cavender, J.C. (1999) Global distribution of forest soil dictyostelids.
613 *Journal of Biogeography*, **26**, 133-148.

614 Tarnawski, S.E. & Lara, E. (2015) From environmental sequences to morphology: observation and
615 characterisation of a paulinellid testate amoeba (Euglyphida, Paulinellidae) from soil using
616 fluorescent in situ hybridization. *Protist,* **166**, 264-270.
617 Tedersoo, L. & Smith, M.E. (2013) Lineages of ectomycorrhizal fungi revisited: Foraging strategies and
618 novel lineages revealed by sequences from belowground. *Fungal Biology Reviews*, **27**, 83-99.
619 Treseder, K.K., Maltz, M.R., Hawkins, B.A., Fierer, N., Stajich, J.E. & McGuire, K.L. (2014) Evolutionary
620 histories of soil fungi are reflected in their large-scale biogeography. *Ecology Letters*, **17**,
621 1086-1093.
622 Tittensor, D., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., van den Berghe, E. & Worm, B. (2010) Global
623 patterns and predictors of marine biodiversity across taxa. *Nature*, **466**, 1098-1101.
624 Vishniac, H.S. (2006) A multivariate analysis of soil yeasts isolated from a latitudinal gradient.
625 *Microbial Ecology*, **52**, 90-103.
626 Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, K., Sterken, M., Hodgson, D.A., Mann, D.G.,
627 Juggins, S., De Vijver, B.V., Jones, V., Flower, R., Roberts, D., Chepurnov, V.A., Kilroy, C.,
628 Vanormelingen, P. & De Wever, A. (2007) Historical processes constrain patterns in global
629 diatom diversity. *Ecology*, **88**, 1924-1931.
630 Watts, P.C., Martin, L.E., Kimmance, S.A., Montagnes, D.J.S. & Lowe, C.D. (2011) The distribution of
631 *Oxyrrhis marina*: a global disperser or poorly characterized endemic? *Journal of Plankton*
632 *Research*, **33**, 579-589.
633 Wilkinson, D.M., Koumoutsaris, S., Mitchell, E.A.D. & Bey, I. (2012) Modelling the effect of size on the
634 aerial dispersal of microorganisms. *Journal of Biogeography*, **39**, 89-97.
635 Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal gradients of biodiversity: Pattern,
636 process, scale, and synthesis. *Annual Review of Ecology Evolution and Systematics*, **34**, 273-
637 309.
638 Wilson, J.B. (1991) Methods for fitting dominance/diversity curves. *Journal of Vegetation Science*, **2**,
639 35-46.

640 Yasuhara, M., Hunt, G., Dowsett, H.J., Robinson, M.M. & Stoll, D.K. (2012) Latitudinal species diversity
641 gradient of marine zooplankton for the last three million years. *Ecology Letters*, **15**, 1174-
642 1179.

643

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1: List of all sampled sites with coordinates, country, climatic values and measured soil variables (pH and C/N)

Appendix S2: PCA of BIOCLIM data extracted from the sampling sites.

Appendix S3: Correlations between spatial, climatic and soil variables.

BIOSKETCH

Enrique Lara is a researcher at the University of Neuchâtel, Switzerland. His research interests encompass various aspects of microbial eukaryote evolution and ecology, with a special focus on testate amoebae.

Author contributions:

E.L. designed the study, L. R.-D. and E.L. performed cloning and sequencing, B.F. and E.L. analysed the data, and all authors discussed the results and wrote the manuscript.

Editor: Walter Jetz

659 **Figure legends**

660 Figure 1: Location of forest litter and mosses sampling sites for a worldwide study of euglyphid
661 environmental diversity, and codes used to identify them.

662 Figure 2a: Accumulation of soil euglyphid phylotypes based on 35 forest litter samples from a
663 worldwide sampling. The box and whiskers show the median, inter-quartile and 95%
664 confidence intervals of phylotype richness based on resampling of the data (100 iterations).

665 Figure 2b: Rank-abundance curves of soil euglyphid phylotype data compared to five models of
666 species rank-abundance, null or broken-stick, pre-emption, lognormal, and Mandelbrot. Note
667 the log scale for the abundance axis of the rank-abundance graph.

668 Figure 3: Maximum likelihood phylogenetic reconstruction of the euglyphid clades recovered from
669 clones obtained through metabarcoding from a worldwide study of forest moss and litter soil
670 samples. The tree was built using sequences of the SSUrRNA gene obtained in this study, plus
671 data from GenBank (both environmental clones and sequences derived from isolates or
672 cultures). A diverse panel of Monadofilosea were chosen to root the tree. A total of 994
673 characters were used in the analysis.

674 Figure 4: Euglyphid phylotype richness as determined by a worldwide survey of euglyphid genetic
675 diversity in forest litter and mosses plotted against (a) maximum temperature of the
676 warmest month of the sites, (b) latitude and (c) sample C/N ratio. Residuals of fitted General
677 Additive Models are plotted against the latter variables. Twice-standard-error curves are
678 shown using dashed lines.

679 Figure 5: Distance-based Moran's eigenvector maps (db-MEM) generated on a world map based on
680 euglyphid community composition obtained through metabarcoding from a worldwide
681 sampling of forest litter and moss. Variables db-MEM6 and 31 are correlated with phylotype
682 distribution. Variable db-MEM6 corresponds generally with a latitudinal gradient, in contrast
683 with db-MEM31. The pattern shown by db-MEM31 suggests strong dissimilarities in

684 communities North and South from the desert belt around the Cancer tropic and/or the
685 inter-tropical convergence.
686

687 **Table**

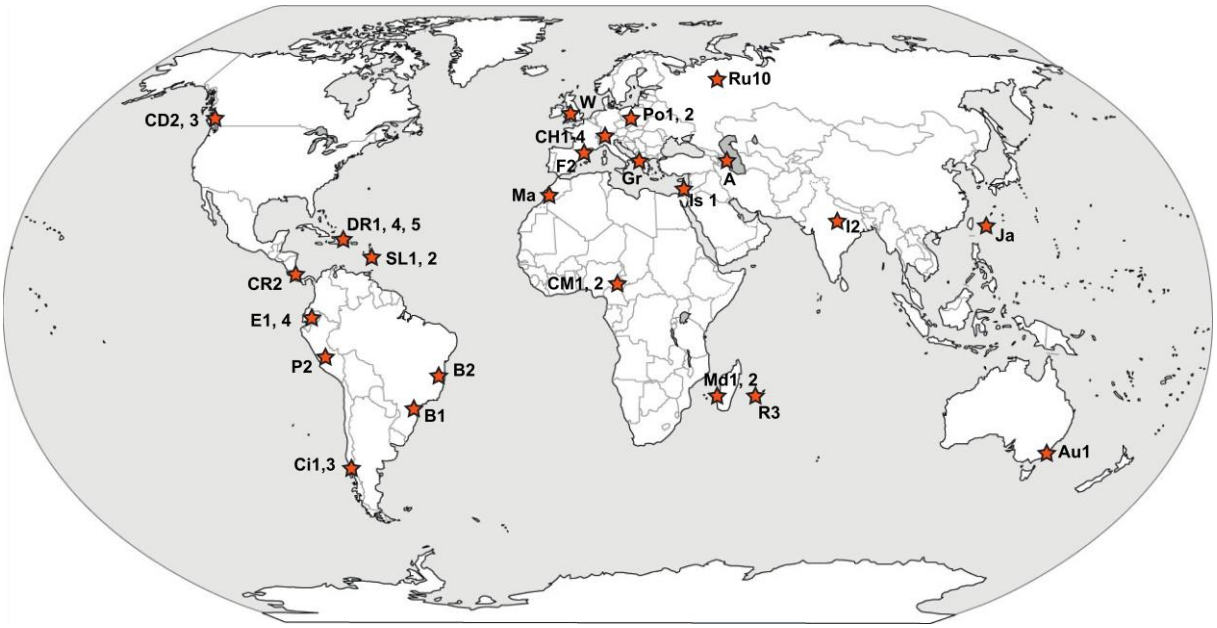
688 Table 1: Summary of statistics derived from GUniFrac and GAM analyses based on euglyphid
689 community composition obtained through metabarcoding from a worldwide sampling of
690 forest litter and moss. Significant values are indicated in bold. Spatial 1 and 2 are distance-
691 based Moran's eigenvectors that have been retained in the analysis, Climate 1 and 2 are the
692 first two principal components from the principal components analysis performed on data
693 extracted from the BIOCLIM dataset (values are shown in Appendix S1 in Supporting
694 Information).

		GAM - changes in species richness (alpha diversity)		GUniFrac - changes in community composition (beta diversity)
		r^2	*P* value (Wald test)	*P* value after 9999 permutations
Spatial 1	db-MEM31	0.01	0.95	**0.01**
Spatial 2	db-MEM6	**54.2**	**< 0.01**	**0.04**
Climate 1	clim_PC1 (PCA BIOCLIM)	16.5	0.08	**0.02**
Climate 2	clim_PC2 (PCA BIOCLIM)	**42.9**	**0.02**	0.15
Soil 1	pH	24.2	0.07	**< 0.01**
Soil 2	C/N	**32.8**	**0.02**	0.07

695

696

697

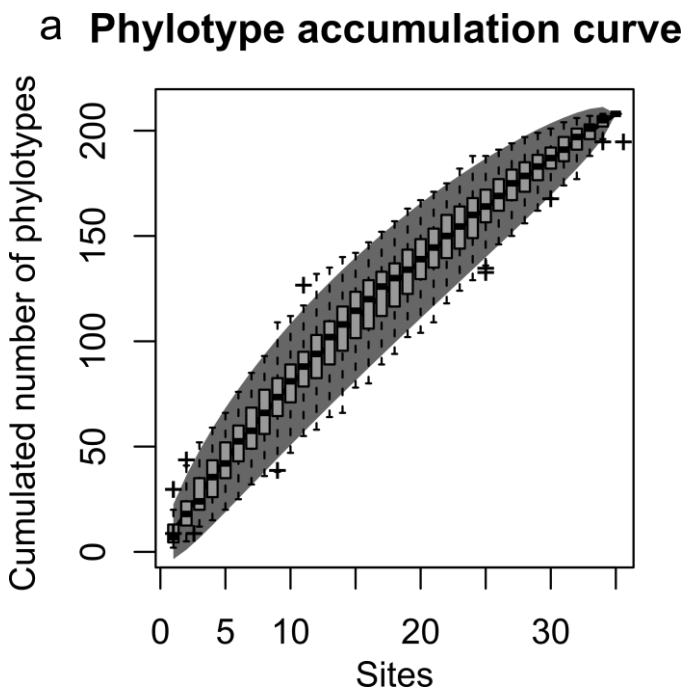


698

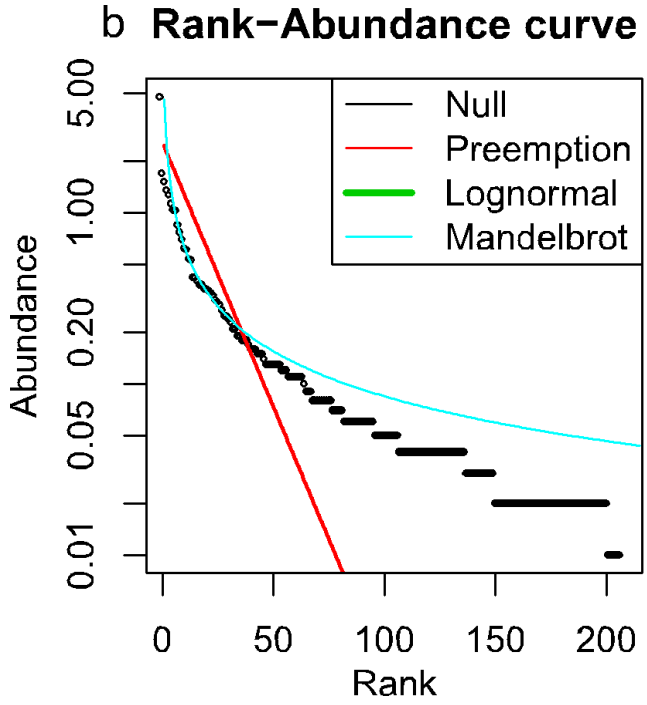
699 Figure 1 Grey scale in print, colour online.

700

701



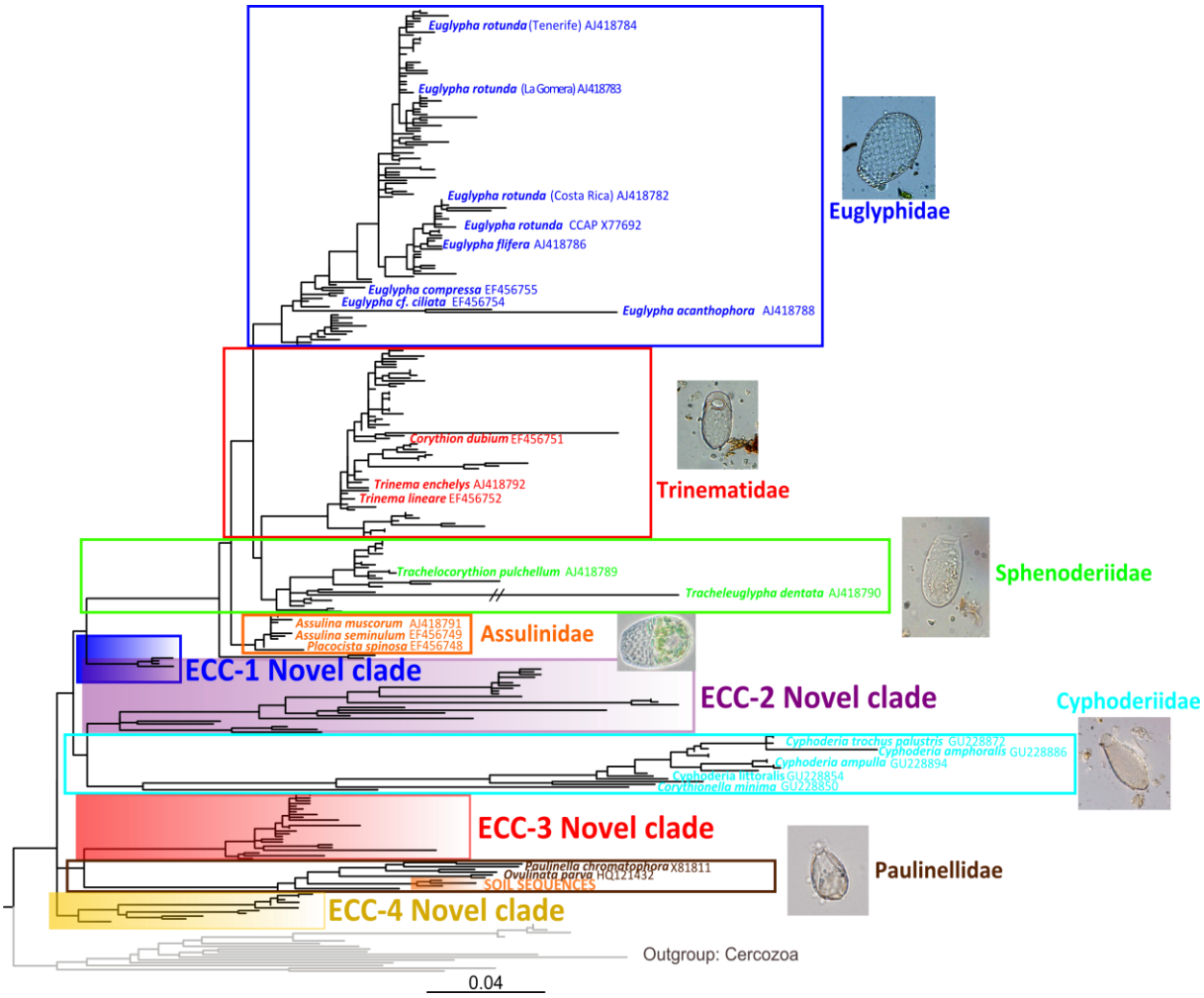
702



703

704 Figures 2 a and b Grey scale in print, colour online.

705

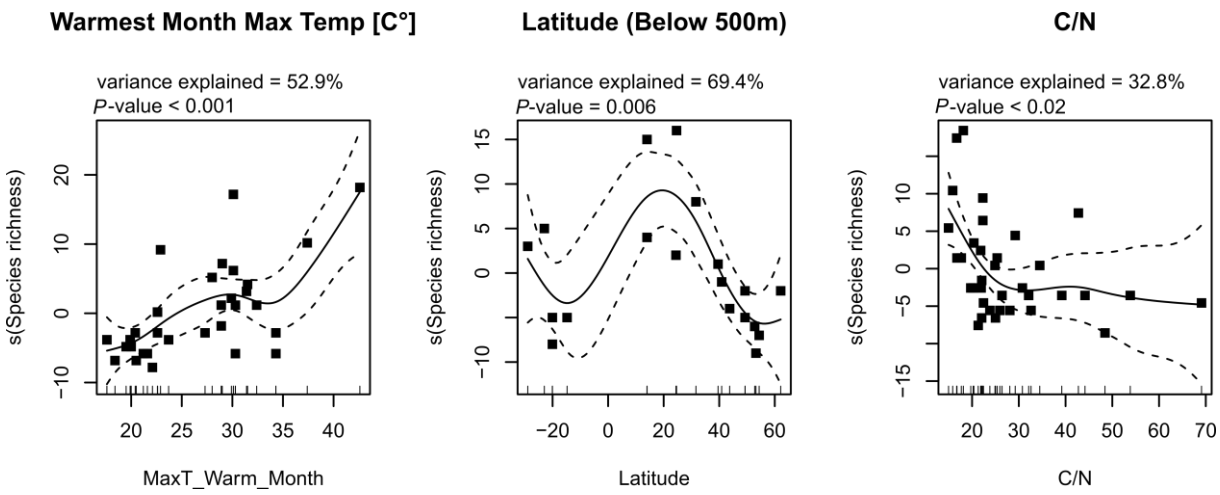


706

707 Figure 3 Grey scale in print, colour online.

708

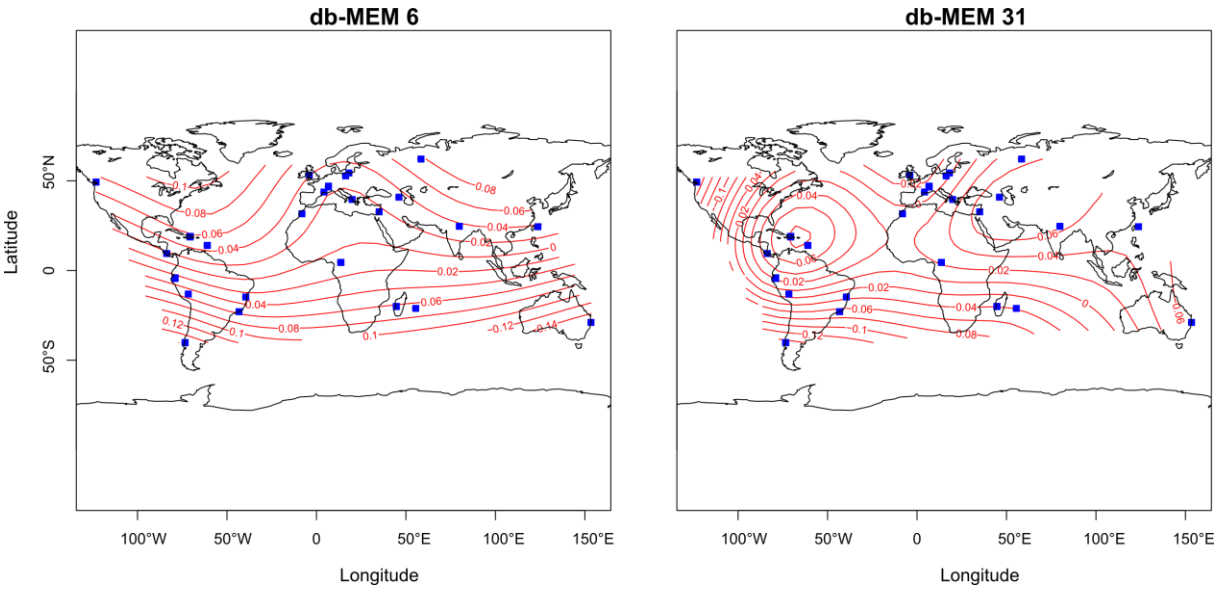
709



710

711 Figure 4 Grey scale in print, colour online

712



713

714

715 Figure 5 Grey scale in print, colour online