

Environmental DNA COI barcoding for quantitative analysis of protists communities: A test using the *Nebela collaris* complex (Amoebozoa; Arcellinida; Hyalospheniidae)

Anush Kosakyan^{a,b,*}, Matthieu Mulot^a, Edward A.D. Mitchell^{a,c}, Enrique Lara^a

^aLaboratory of Soil Biology, University of Neuchatel, Rue Emile-Argand 11, CH-2000 Neuchatel, Switzerland

^bInstitute of Biosciences, Department of Zoology, University of Sao Paulo, Rua do Matão, Travessa 14, Cidade Universitária, 05508-090 - Sao Paulo, SP, Brasil

^cJardin Botanique de Neuchâtel, Chemin du Perthuis-du-Sault 58, CH-2000 Neuchâtel, Switzerland

Abstract

Environmental DNA surveys are used for screening eukaryotic diversity. However, it is unclear how quantitative this approach is and to what extent results from environmental DNA studies can be used for ecological studies requiring quantitative data. Mitochondrial cytochrome oxidase (COI) is used for species-level taxonomic studies of testate amoebae and should allow assessing the community composition from environmental samples, thus bypassing biases due to morphological identification. We tested this using a COI clone library approach and focusing on the *Nebela collaris* complex. Comparisons with direct microscopy counts showed that the COI clone library diversity data matched the morphologically identified taxa, and that community composition estimates using the two approaches were similar. However, this correlation was improved when microscopy counts were corrected for biovolume. Higher correlation with biovolume-corrected community data suggests that COI clone library data matches the ratio of mitochondria and that within closely-related taxa the density of mitochondria per unit biovolume is approximately constant. Further developments of this metabarcoding approach including quantifying the mitochondrial density among closely-related taxa, experiments on other taxonomic groups and using high throughput sequencing should make it possible to quantitatively estimate community composition of different groups, which would be invaluable for microbial food webs studies.

Keywords: Community ecology; Environmental DNA; Metabarcoding; Protist diversity; *Sphagnum* peatlands; Testate amoebae

Introduction

Environmental DNA surveys are revealing a huge unknown diversity of microbial eukaryotes both globally and

within individual samples (Epstein and Lopez-Garcia 2008; Lopez-Garcia et al. 2001; Pawlowski et al. 2012). This high diversity in turn suggests the existence of extremely complex, but mostly undocumented ecological interactions. To understand the ecological roles played by species quantitative estimates of their abundance and biomass are required. For soil protozoan groups such as ciliates and testate amoebae, this is currently done using highly time-consuming direct

*Corresponding author. Tel.: +5511 980846710.
E-mail address: anush.kosakyan@unine.ch (A. Kosakyan).

counting of known taxa (Foissner et al. 2002; Jassey et al. 2010; Payne and Mitchell 2009; Schwarz and Frenzel 2003) rather than environmental DNA approaches.

Assessment of environmental micro-eukaryotic diversity is almost exclusively done by sequencing partial or entire ribosomal genes (Pawlowski et al. 2012). Besides biases in DNA extraction, PCR and possibly cloning, copy numbers of ribosomal genes in eukaryotes are known to vary over more than four ranges of magnitude, from one in the picoplanktonic *Nannochloropsis salina* to 12,000 in the large dinoflagellate *Akashiwo sanguinea* (Zhu et al. 2005). Such biases, already known and documented in prokaryotes (Von Wintzingerode et al. 1997), can be expected to be even higher in protists given their larger, more complex and more variable genomes. Thus, in many cases the community structure or biomass evaluation inferred from rDNA studies are not always reliable (Crosby and Criddle 2003; Morgan et al. 2010; Weber and Pawlowski 2013; Von Wintzingerode et al. 1997). Selecting the optimal barcoding marker is not trivial either; ribosomal genes (SSU and LSU rRNA) are most useful for coarse taxon discrimination, while ITS has been shown to present intra-genomic variation in some groups and is therefore not suited as a barcoding gene for all eukaryotic groups (Pawlowski et al. 2012).

The analysis of protist communities by light microscopy is time-consuming. Precise identification of taxa is often hindered by poor taxonomy, and the existence of morphologically similar, but genetically distinct species, which may have different ecological niches. As the identification of many putative taxa is highly dependent on the observer, results are often difficult to compare among studies. This problem, often referred to as cryptic diversity, is recurrent in eukaryotic micro-organisms (Beszteri et al. 2005; Heger et al. 2010; Koch and Ekelund 2005; Kosakyan et al. 2012; Kucera and Darling 2002).

An alternative to existing approaches would be to apply molecular methods but to focus on small groups that are genetically and morphologically well-characterised. In such a context, the use of a variable marker is possible and even required to reach a higher level of taxonomic accuracy. Candidate markers have to be tested and validated for DNA barcoding before being applied to environmental DNA samples. A good candidate marker for such an approach is the mitochondrial cytochrome oxidase subunit 1 (COI) first used for Amoebozoa in species-level taxonomical studies of van-nelids Nasonova et al. 2010, and later in Arcellinida (Heger et al. 2013; Kosakyan et al. 2012, 2013).

With the aim of both overcoming limitations of both direct microscopy and environmental rRNA gene surveys, we developed and tested an environmental DNA based method to screen protist diversity, using a cloning-sequencing approach and COI as marker gene. This method allows the assessment of taxonomic diversity within a selected group of closely related protist species. We focussed here on the *Nebela collaris s.l.* species complex (Amoebozoa; Arcellinida; Hyalospheniidae), a common group of testate

amoebae in Northern Hemisphere *Sphagnum* peatlands and acidic forest humus.

COI is a mitochondrial marker and, in contrast to ribosomal genes, number of copies should depend directly on mitochondrial density within the organisms. In multicellular organisms (including humans), this parameter can vary depending on the physiological activity of the tissues. For instance, intense muscular activity promotes mitochondrial division, as an adaptation to higher respiration rates required (Hoppeler et al., 1985). In Hyalospheniid cells, metabolism is not expected to vary as much as in the different tissues present in multicellular organisms. Indeed, these testate amoebae have, to our knowledge, never been found alive under anoxic or microaerophilic conditions. Living testate amoebae are only found in the uppermost, photosynthetic section of *Sphagnum* mosses and the young litter immediately below (Mitchell and Gilbert, 2004), where levels of dissolved oxygen can be expected to be high. In addition, while mass-specific activity (a parameter related to mitochondrial density) decreases linearly with body mass in Metazoa, it has been shown to remain constant in plants and protists (Makarieva et al. 2008). It can therefore be reasonably expected that mitochondrial density does not vary much during the life cycle of Hyalospheniid testate amoebae. Therefore, mitochondria numbers can be expected to increase proportionally with cell biovolume. If true, quantitative estimates of community structure from environmental samples would reflect relative biovolume rather than relative cell numbers and this discrepancy would increase with increasing contrast in cell size among taxa. Species of the *Nebela collaris* group show contrasted biovolumes that vary by as much as a factor 3 and thus represent a good model to test if the method can be used to obtain quantitative estimates of community composition from environmental samples. We tested if such a bias could exist within the *Nebela collaris s.l.* species complex by comparing the relative abundance of clones vs. known community composition using artificial communities (i.e. of known composition) and environmental samples. We assessed if the accuracy of COI-based quantitative community structure estimates increased with correction for cell biovolume and thus if COI-based environmental studies could provide quantitative estimates of community structure for the *Nebela collaris* group.

Material and Methods

Analyses of natural and artificial communities

Testate amoebae were extracted from *Sphagnum* samples collected from four peatlands in the Jura Mountains of Switzerland and France (Table 1). Testate amoeba cells were extracted and concentrated by sieving (150 μ m) and back sieving (20 μ m). This material was used for two complementary sets of experiments schematically shown in Fig. 1 and detailed hereafter:

Table 1. Sampling locations for environmental communities.

Sample name	Location and coll. dates	Country	Coordinates	Habitat
Sample 1	Les Pontins 15.04.2013	Switzerland	47°01'27.37"N; 6°98'96"E	Poor fen, <i>Sphagnum</i> mosses
Sample 2	Le Cachot 08.11.2012	Switzerland	47°00'20.93"N; 6°39'52.78"E	Centre of peatland with <i>Eriophorum vaginatum</i>
Sample 3	Le Russey 28.03.2012	France	47°10'128"N; 6°46'263"E	Poor fen, bog margin, <i>Sphagnum fallax</i> , <i>Eriophorum vaginatum</i>
Sample 4	Trédudon 16.01.2014	France	48°24'44.13"N; 3°50'33.21"W	Hummock in rich fen, <i>Sphagnum palustre</i>

For the first experiment the extracted community of each sample was split into two parts and left to settle in centrifuge tubes. One part was used for bulk environmental DNA extraction of natural communities: Approximately 0.40 g of microbial community extract was collected from the bottom of the centrifuge tube for environmental DNA extraction. The other was used for direct microscopy observation of communities: A total of 100–150 testate amoebae from *N. collaris s.l.* group were identified from each sample.

For the second experiment, individual species (Fig. 2) were isolated one at a time with a narrow diameter pipette under an inverted microscope from the same set of samples to build artificial communities. Each cell was documented carefully for species identification, washed several times with distilled water and deposited in an Eppendorf™ tube according to pre-defined community composition (Supplementary Table 2). Two contrasted artificial communities were created (namely Art-A and Art-B). DNA was then extracted from each of these samples.

DNA extraction, amplification and cloning

Power Soil DNA isolation kit (MO BIO) was used to extract the environmental DNA from both environmental and artificial communities following the manufacturer's instructions. Partial COI PCR products were obtained using the wide-spectrum primer LCO (Folmer et al. 1994) in combination with a specific primer TINCOX (CCATTCKATAHCCH-GGAAATTC) designed to amplify only *Nebela collaris s.l.* species (Kosakyan et al. 2013) allowing to amplify about 500 bp length of sequence. DNA was amplified in a total volume of 25 µl with an amplification profile consisting of a 5 min initial denaturation step in a 40 cycles programme of 15 s at 95 °C, 15 s at 43 °C, and 1 min and 30 s at 72 °C with the final extension at 72 °C for 10 min.

The PCR products were cloned into pCR2.1 Topo TA cloning vector (Invitrogen) and transformed into *E. coli* TOP10' One Shot cells (Invitrogen) according to the manufacturer's instructions. Cloned inserts were amplified with

vector M13F and M13R primers. PCR products were purified with the NucleoFasts 96 PCR Clean Up kit from Macherey-Nagel (Düren, Germany) and sequenced with an ABI PRISM 3700 DNA Analyzer (PE Biosystems, Genève, Switzerland) using a BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems).

COI sequences are deposited in GenBank with the following accession numbers XX-XX (nb. accession numbers will be added after paper acceptance).

Phylogenetic analyses

The data set used for phylogenetic analyses (300–665 bp) comprised 201 COI new environmental sequences that were analysed together with 31 COI *Nebela collaris s.l.* sequences taken from Kosakyan et al. 2013 sequence data. The sequences were aligned using ClustalW and refined manually using the BioEdit software. The alignment is available from the authors upon request. Trees were reconstructed using alternatively a maximum likelihood and a Bayesian approach. The maximum likelihood tree was built using the RAxML v7.2.8 algorithm (Stamatakis et al. 2008).

as proposed on the Black Box portal (<http://phylobench.vital-it.ch/raxml-bb/>) using the GTR + Γ + I model. Model parameters were estimated in RAxML over the duration of the tree search. We used sequences from *Certesella martiali* (GenBank number JN849064) and from *Nebela tubulosa* (JN849020, JN849021, JN849061) to root the tree, based on the fact that these species appear relatively closely related to the *N. collaris s.l.* group in the COI gene-based phylogeny of Hyalospheniidae (Kosakyan et al. 2012). Bayesian Markov Chain Monte Carlo analyses were performed using MrBayes v3.1 (Ronquist et al. 2005) with a general time reversible model of sequence evolution with four gamma-distributed rate variation across sites and a proportion of invariable sites. Bayesian MCMC analyses were carried out with two simultaneous chains, and 1000,000 generations were performed. The generations were added until the standard deviation of split frequencies fell below 0.01, according to the manual of

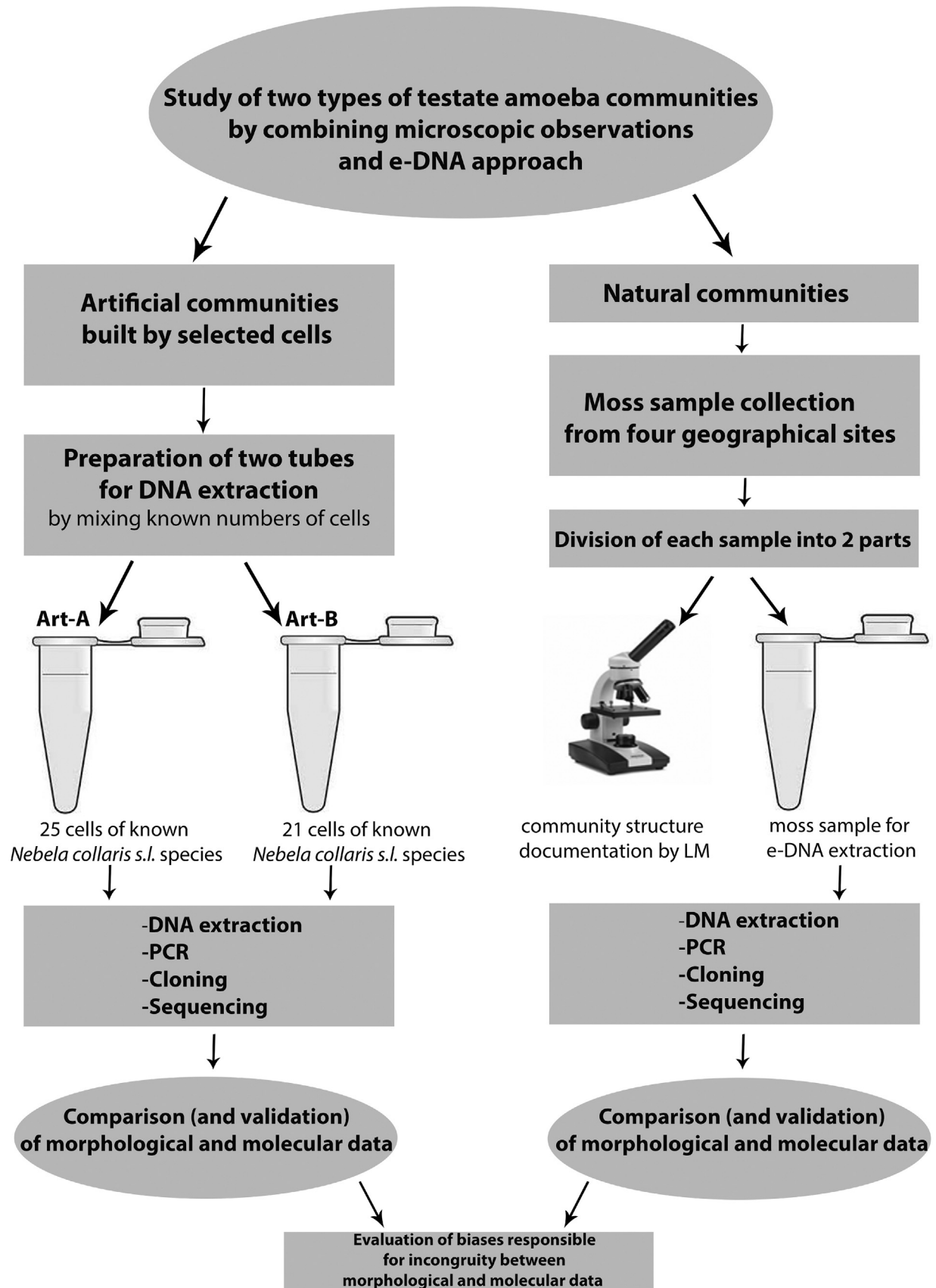


Fig. 1. Schematic flow of experiments used in this study to determine the relationship between clone-library data and microscopy data.

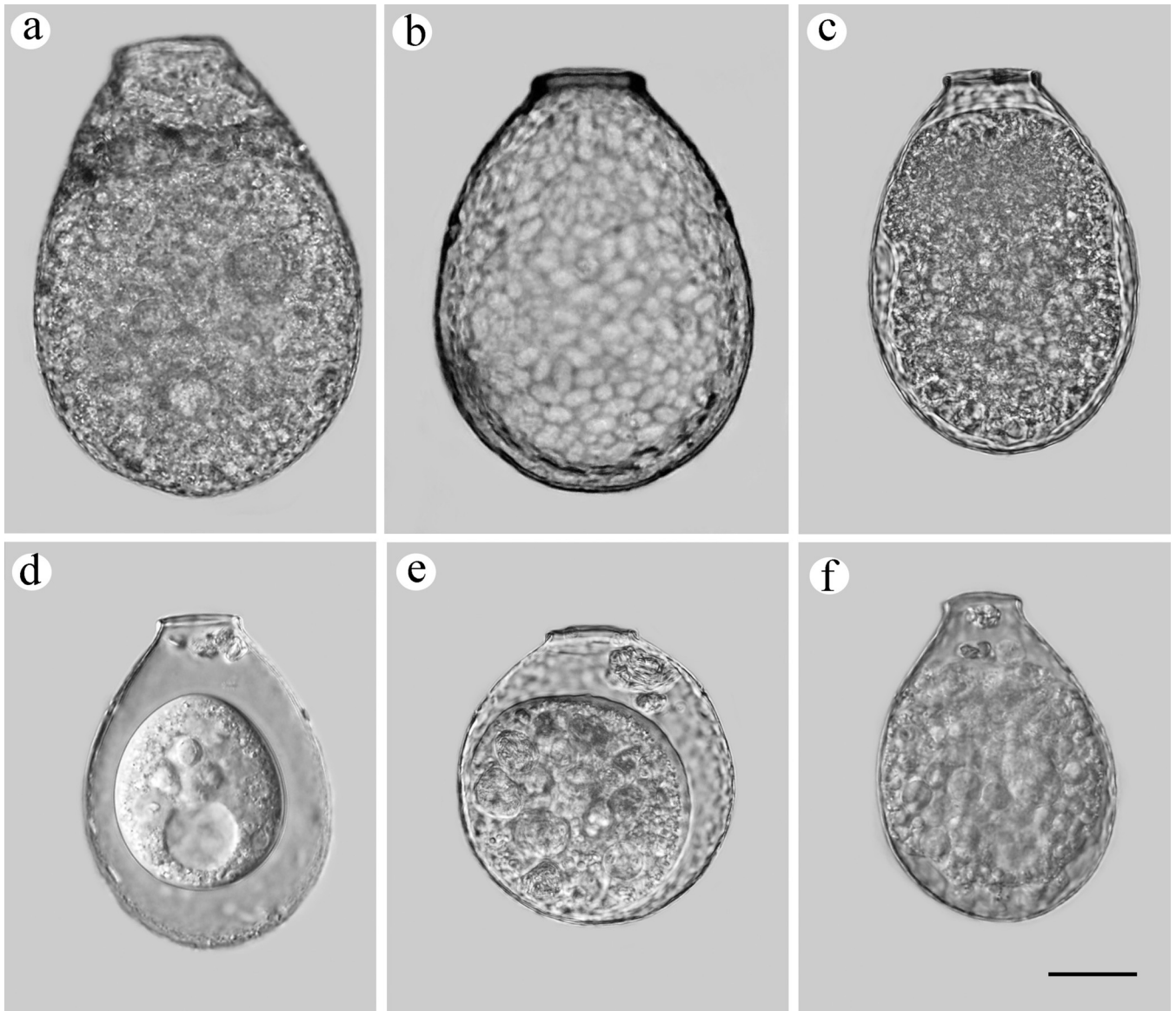


Fig. 2. Light microscopy images of the six taxa within the *Nebela collaris* s.l. group included in this study, (a) *Nebela collaris*, (b) *N. aliciae* (modified from Kosakyan et al. 2013), (c) *N. tincta*, (d) *N. pechorensis*, (e) *N. rotunda*, and (f) *N. guttata*.

MrBayes 3.1 (Ronquist et al. 2005). The tree with the best likelihood score was saved every 10 generations, resulting in 100,000 trees. The burn in value was set to 25%. Trees were viewed using FigTree (programme distributed as part of the BEAST package <http://tree.bio.ed.ac.uk/software/figtree/>).

Comparison of morphological and molecular data

We compared community composition as estimated by light microscopy to the proportion of clones recovered. Predicting that the number of mitochondria in a given cell (and, therefore, the number of COI gene copies) will be directly proportional to the biovolume of the amoeba, we compared both (1) raw proportions of clones and (2) proportions of

clones as normalised by the ratio between the largest biovolume (namely *N. collaris* s.str.) and the biovolume of the given cell after the following equation:

The biovolume was calculated according to Charrière et al. 2006 using the following equation:

$$B_x = \frac{2}{3} \times L \times W \times D \quad (1)$$

where B_x is the biovolume of species x [μm^3]; L is the length of the shell [μm]; W is the width of the shell [μm]; D is the depth of the shell [μm].

In order to calculate the *corrected clone number of given species in the community*, we used the following equation:

$$C_{x\text{-corr}} = \frac{B_{x(\text{max})}}{B_x \times C_x} \quad (2)$$

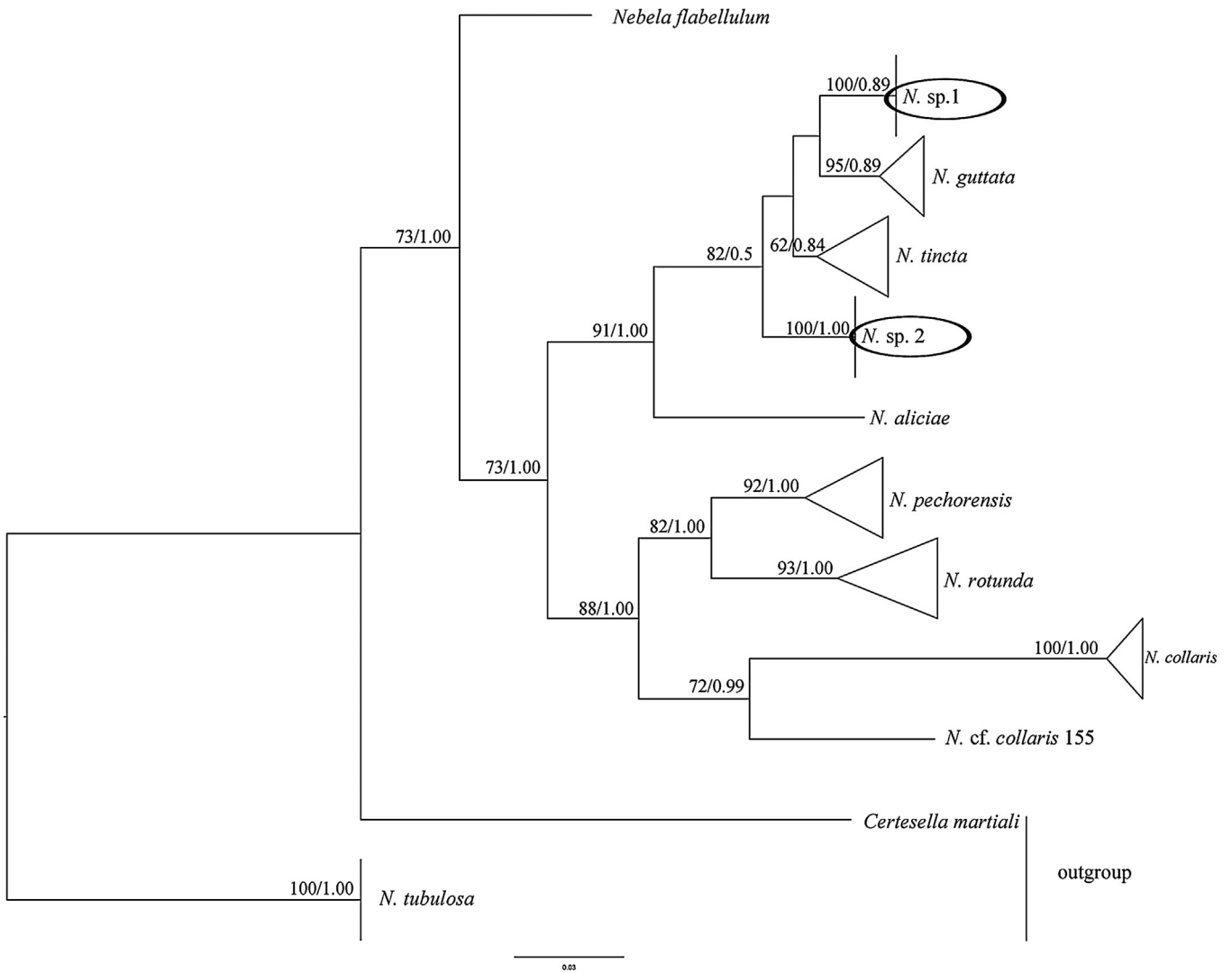


Fig. 3. Maximum likelihood bootstrap consensus tree including 31 sequences of *N. collaris* s.l. derived from single cell PCR and adding 201 environmental mtCOI sequences based on a 665-nucleotide alignment. The numbers along the branches represent, respectively, the bootstraps obtained by maximum likelihood method and the posterior probabilities as calculated with Bayesian analyses. Values under 50/0.50 are not shown.

where $C_{x,corr}$ is the corrected clone number of species x in the community; $B_{x(max)}$ is the biovolume of the largest species in the community (e.g. *Nebela collaris* s.str.) [μm^3]; B_x is the biovolume of species x [μm^3]; C_x is the number of clones of species x in the community (from clone library data).

Corrected clone numbers ($C_{x,corr}$) are then used to calculate corrected relative abundance [%].

Statistical analyses

Similarity in communities structure estimated by each method was explored using standard Pearson correlation test, according to the guidelines provided in Puth et al. (2014). In addition we computed linear regression of the relative abundances of each species obtained by microscopic count (x), against their relative abundances obtained by either molecular count or molecular count corrected with biovolume (y). The

regression formula being, if two communities are identical $a = 1$ and $b = 0$.

Results

Most of the obtained clones were affiliated to known species of *Nebela collaris* s.l., as defined by Kosakyan et al. 2013, with the exception of two groups of clones which in all likelihood form two independent taxonomic units (*Nebela* unknown sp. 1 and *N.* unknown sp. 2), respectively, sister to *N. guttata* and basal to both *N. tincta* and *N. guttata*. Both of these clades are fully supported (100 B and 1.00 PP; Fig. 3) and were detected in low numbers only in the environmental surveys.

Microscopy counts were significantly and strongly correlated (using paired Pearson correlation tests) to both raw

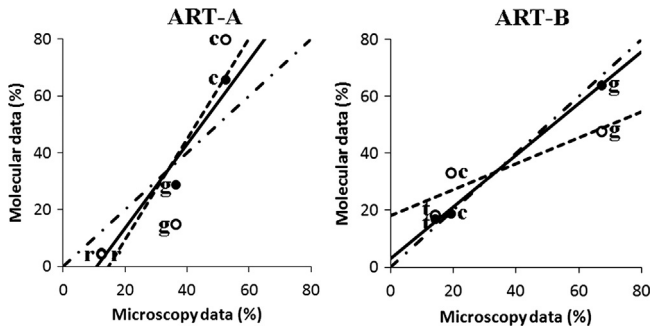


Fig. 4. Biplots of relative abundance of *Nebela collaris* s.l. taxa as assessed by direct microscopy observation vs. cloning and sequencing of the COI gene from two artificial communities before and after correction with biovolume. Semi dashed lines indicates the 1:1 slope corresponding to a perfect correlation. Dashed lines and full lines indicate actual correlation before (dashed) and after (full) biovolume correction. Species names are indicated as follows: c: *Nebela collaris*, g: *N. guttata*, r: *N. rotunda*, and t: *N. tincta*. Open and full symbols correspond to values before and after biovolume correction, respectively.

and biovolume-corrected proportions of clones ($r = 0.844$ and $r = 0.974$, respectively, $n = 5$, $p < 0.05$ in both cases) in the two artificial communities (Supplementary Tables 1 and 2, Fig. 4). A similar increase was observed in environmental sample 3 ($r = 0.900$ and $r = 0.962$, respectively, before and after correction with biovolume, $n = 7$, $p < 0.05$), and to a lesser degree in sample 2 ($r = 0.831$ and 0.852 , $p < 0.05$, $n = 7$) where an unknown species was found (*N. sp. 1*). In sample 4 we retrieved only *Nebela collaris* in the clone library in agreement with microscopic observations (Supplementary Table 3, Fig. 5). The correlation decreased in sample 1 ($r = 0.964$ and 0.902 , respectively, before and after correction by biovolume, $p < 0.05$ and $n = 7$), where a second unknown species was found (*N. sp. 2*).

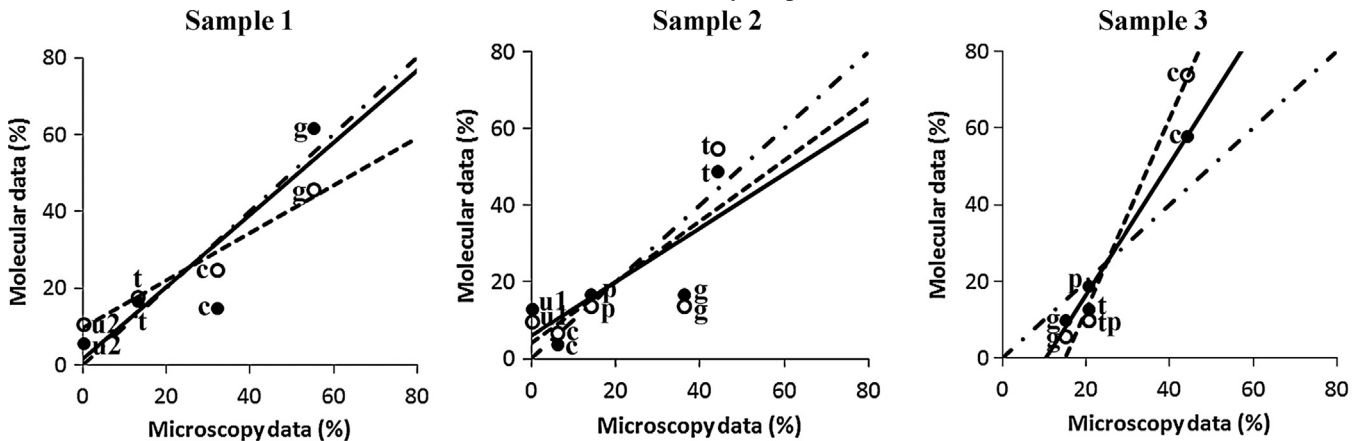


Fig. 5. Biplots of relative abundance of *Nebela collaris* s.l. taxa as assessed by direct microscopy observation vs. cloning and sequencing of the COI gene from natural communities before and after correction with biovolume. Semi dashed lines indicates the 1:1 slope corresponding to a perfect correlation. Dashed lines and full lines indicate actual correlation before (dashed) and after (full) biovolume correction. Species names are indicated as follows: c: *Nebela collaris*, g: *N. guttata*, r: *N. rotunda*, and t: *N. tincta*, u1: unknown species 1, u2: unknown species 2. Open and full symbols correspond to values before and after biovolume correction, respectively.

Assuming (1) that *N. sp. 1* and 2 were cryptic species that were morphologically undistinguishable from some other species in the sample, and (2) that they looked exactly like the species that showed the largest deviation to counted numbers proportions, we normalised their biovolume accordingly (i.e. respectively *N. collaris* in sample 1 for *N. sp. 2* and *N. guttata* in sample 2 for *N. sp. 1*). The resulting correlation increased for the overall dataset (Fig. 6).

Discussion

Community ecology requires reliable identification of species as well as an accurate assessment of the relative abundance of each species. Most communities are constituted of species differing in life history traits such as sizes and reproduction rates. An optimal method to assess community composition should therefore allow the precise identification of all taxa as well as of the relative biomass of each. The COI-based metabarcoding approach we tested on the *Nebela collaris* s.l. group shows that this method is promising. Other traits can only be included in analyses if enough is known about the organisms.

In this study our aim was to develop a method that allowed a quantitative screening of environmental communities at species level using a molecular approach. This is, to our knowledge, the first time that a metabarcoding approach that is not based on ribosomal genes is applied to a group of heterotrophic protists. The advantage of the COI-based metabarcoding approach is to provide a much finer level of taxonomic resolution than that the SSU rRNA gene usually used in environmental DNA studies. COI has also been shown to give high taxonomic resolution in many microbial eukaryotic groups (Barth et al. 2006; Chantangsi et al. 2007; Heger et al. 2010; Lin et al. 2009; Nasonova et al. 2010). For example, in the Hyalospheniidae, SSU rRNA does not allow to

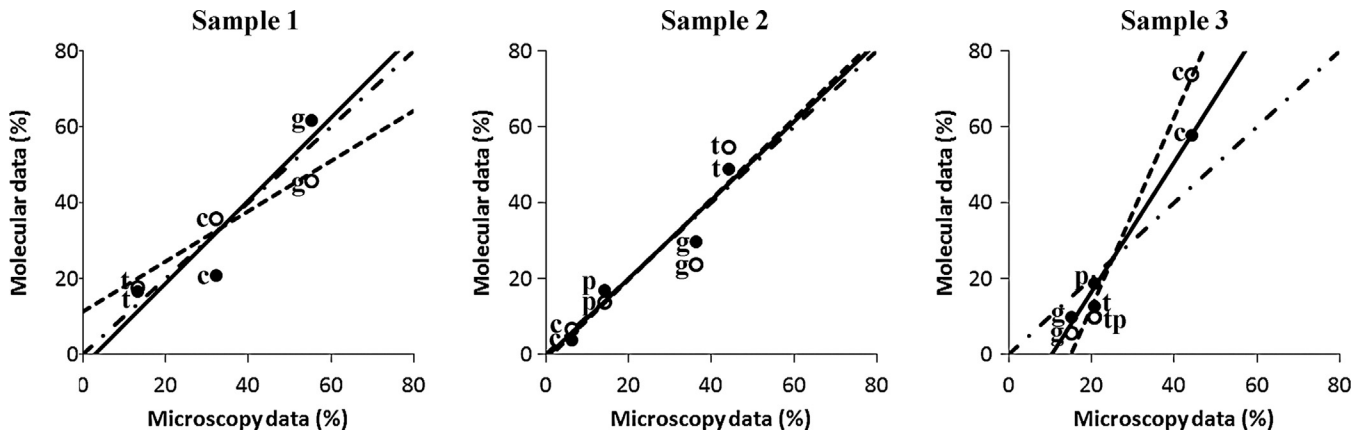


Fig. 6. Same data as in Fig. 5 but with unknown species 1 and 2 referred to *Nebela guttata* and *N. collaris*, respectively.

discriminating among closely related taxa (Lara et al. 2008) while COI does (Kosakyan et al. 2012, 2013). The COI-based metabarcoding approach allows discrimination among taxa that are generally interpreted as biological species, for Metazoa (Hebert et al. 2003; Ratnasingham and Hebert 2007) (but see Meyer and Paulay 2005; Moritz and Cicero 2004) and our recent work on Hyalospheniidae also supports this view (Heger et al. 2013; Kosakyan et al. 2013).

Thus the first advantage of this approach is to overcome taxonomic limitation of morphology-based analyses and allow fine-level ecological studies regardless of the current state of taxonomy for individual groups. Indeed this approach can be used in the *Nebela collaris* s.l. group as well as other groups for which a representative amount of COI sequence data is available, even if the total diversity of the group is currently unknown. In this latter case this environmental metabarcoding approach is even the only available option to assess the ecological significance of hidden diversity. The discovery of two new phylotypes of species level shows that the full taxonomic diversity within the *Nebela collaris* group is still not known. This is not surprising given recent results of phylogenetic studies on this group. For example the existence of closely related but distinct species that have geographically limited distributions has recently been demonstrated in hyalospheniids for the *Hyalosphenia papilio* species complex (Heger et al. 2013). The application of our approach in other areas of the world will most likely allow the discovery of more unknown taxa. This could be achieved relatively easily using the environmental DNA-metabarcoding approach, either using cloning-sequencing as done here or using high throughput sequencing.

A second advantage of this method is that it allows the retrieval of reliable quantitative data, and our data suggests that raw clone proportions provide an accurate estimate of the relative biovolume of each species in a community, even for cryptic or unknown species whose biovolume is unknown. Although this may be seen as a bias if the goal is to estimate proportions of individuals it may actually be an advantage for community ecology studies. Community ecology studies of testate amoebae are indeed increasingly based on

biovolume-corrected community data obtained by direct microscopy (Jassey et al. 2013).

The rationale behind this approach is that biovolume-corrected community composition informs better on the true ecological impact of the species. As our results suggest that clone-library analysis directly provides community estimates that closely match a biovolume-corrected community count we thus believe that the clone-library results could be used for ecological analysis without any need for further corrections by biovolume.

This study clearly calls for more work, to test underlying assumptions and to explore the topic further:

1. In this study we accepted the assumption that in the *Nebela collaris* group mitochondrial density is approximately constant and thus increases proportionally with cell size. This has not been tested here and would constitute an obvious first follow-up study.
2. This approach can also be used to explore the full diversity of the *Nebela collaris* group and other protist communities in the environment and the degree to which closely related taxa differ in their ecological niche, and functional roles and to assess biotic interactions (e.g. competition). This study represents a first step towards these broader goals, taking as a model a particular group of hyalospheniid testate amoebae, *N. collaris sensu lato*.
3. Obviously, the approach presented in this work can be applied at first only to small groups of reasonably well-known protist species. Species must have been barcoded and their morphologies must have been properly documented. When new (i.e. non-barcoded) species appear within an otherwise well documented group the clone numbers may be used to estimate relative biovolume, assuming that the density of mitochondria per unit volume is approximately constant. It may thus be a useful tool to select morphotypes for classical barcoding on isolated individuals.
4. This approach can only be applied if members of the studied group share a similar lifestyle (e.g. aerobic) otherwise mitochondrial density might change and the assumption

of constant mitochondrial density should be tested. Relatively closely-related taxa that include species colonizing contrasted habitats (e.g. from high to low oxygen content) would constitute a good study topic to assess how variable mitochondrial density can be within a given taxonomic group.

We believe that this approach can easily be developed for other well-documented groups of testate amoebae, as well as other protists such as ciliates, diatoms, haptophytes etc., thus bringing an invaluable tool for answering many fundamental questions on the ecology of protist communities and revealing their true diversity.

Acknowledgements

This work was funded by Swiss NSF projects no. 205321-109709/1 and 205321-109709/2 to E. Mitchell; Science and Technology Cooperation Program Switzerland–Russia grant IZLR Z3 128338 to J. Pawlowski, E. Mitchell & E. Lara; Swiss NSF Ambizione grant no. PZ00P2 122042 to E. Lara; and Swiss SNF Early Mobility grant P2NEP1-155421 to A. Kosakyan. We thank Daniel Lahr, Frederick Spiegel and Jan Pawlowski for constructive comments on the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ejop.2015.06.005>.

References

- Barth, D., Krenek, S., Fokin, S.I., Berendonk, T.U., 2006. Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome *c* oxidase I sequences. *J. Eukaryot. Microbiol.* 53, 20–25, <http://dx.doi.org/10.1111/j.1550-7408.2005.0068.x>.
- Beszteri, B., Acs, E., Medlin, L.K., 2005. Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. *Protist* 156, 317–333, <http://dx.doi.org/10.1016/j.protis.2005.07.002>.
- Chantangsi, C., Lynn, D.H., Brandl, M.T., Cole, J.C., Hetrick, N., Ikononi, P., 2007. Barcoding ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. *Int. J. Syst. Evol. Microbiol.* 57, 2412–2425, <http://dx.doi.org/10.1099/ijs.0.64865-0>.
- Charrière, F., Pavillon, N., Colomb, T., Depaursing, C., Heger, T.J., Mitchell, E.A.D., Marquet, P., Rappaz, B., 2006. Living specimen tomography by digital holographic microscopy: morphometry of testate amoeba. *Opt. Express* 14, 7005–7013.
- Crosby, L.D., Criddle, C.S., 2003. Understanding bias in microbial community analysis techniques due to rrrn operon copy number heterogeneity. *Biotechniques* 34, 790.
- Epstein, S., Lopez-Garcia, P., 2008. “Missing” protists: a molecular prospective. *Biodivers. Conserv.* 17, 261–276, <http://dx.doi.org/10.1007/s10531-007-9250-y>.
- Foissner, W., Agatha, S., Berger, H., 2002. Soil ciliates (Protozoa, Ciliophora) from Namibia (Southwest Africa), with emphasis on two contrasting environments, the Etosha Region and the Namib Desert. *Denisia* 5, 1–1459.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.
- Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome *c* oxidase subunit I divergences among closely related species. *Proc. R. Soc. B: Biol. Sci.* 270, 96–99, <http://dx.doi.org/10.1098/rsbl.2003.0025>.
- Heger, T.J., Mitchell, E.A.D., Todorov, M., Golemansky, V., Lara, E., Leander, B.S., Pawlowski, J., 2010. Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. *Mol. Phylogenet. Evol.* 55, 113–122, <http://dx.doi.org/10.1016/j.ympev.2009.11.023>.
- Heger, T.J., Mitchell, E.A.D., Leander, B.S., 2013. Holarctic phylogeography of the testate amoeba *Hyalosphenia papilio* (Amoebozoa: Arcellinida) reveals extensive genetic diversity explained more by environment than dispersal limitation. *Mol. Ecol.* 22, 5172–5184, <http://dx.doi.org/10.1111/mec.12449>.
- Hoppeler, H., Howald, H., Conley, K., Lindstedt, S.L., Claassen, H., Vock, P., Weibel, E.R., 1985. Endurance training in humans-aerobic capacity and structure of skeletal muscle. *J. Appl. Phys.* 59, 320–327.
- Jassey, J.E.J., Chapuiso, G., Mitchell, E.A.D., Binet, P., Toussaint, M.-L., Gilbert, D., 2010. Fine-scale horizontal and vertical micro-distribution patterns of testate amoebae along a narrow fen/bog gradient. *Microb. Ecol.* 61, 374–385, <http://dx.doi.org/10.1007/s00248-010-9756-9>.
- Jassey, V.E.J., Chiapusio, G., Binet, P., Buttler, A., Laggoun-Defarge, F., Delarue, F., Bernard, N., Mitchell, E.A.D., Toussaint, M.-L., Francez, A.-J., Gilbert, D., 2013. Above- and belowground linkages in *Sphagnum peatland*: climate warming affects plant-microbial interactions. *Global Change Biol.* 19, 811–823, <http://dx.doi.org/10.1111/gcb.12075>.
- Koch, T.A., Ekelund, F., 2005. Strains of the heterotrophic flagellate *Bodo designis* from different environments vary considerably with respect to salinity preference and SSU rRNA gene composition. *Protist* 156, 97–112, <http://dx.doi.org/10.1016/j.protis.2004.12.001>.
- Kosakyan, A., Heger, T.J., Leander, B.S., Todorov, M., Mitchell, E.A.D., Lara, E., 2012. COI barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the *Hyalospheniidae* Schultze. *Protist* 163, 415–434, <http://dx.doi.org/10.1016/j.protis.2011.10.003>.
- Kosakyan, A., Gomaa, F., Mitchell, E.A.D., Heger, T.J., Lara, E., 2013. Using DNA-barcoding for sorting out protist species complexes: a case study of the *Nebela tinctoris-collaris-bohemica* group (Amoebozoa; Arcellinida, Hyalospheniidae). *Eur. J. Protistol.* 9, 222–237, <http://dx.doi.org/10.1016/j.ejop.2012.08.006>.
- Kucera, M., Darling, K.F., 2002. Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. *Philos. Trans., A: Math. Phys. Eng. Sci.* 360, 695–718, <http://dx.doi.org/10.1098/rsta.2001.0962>.

- Lara, E., Heger, T.J., Ekelund, F., Lamentowicz, M., Mitchell, E.A.D., 2008. Ribosomal RNA genes challenge the monophyly of the Hyalospheniidae (Amoebozoa: Arcellinida). *Protist* 159, 165–176.
- Lin, S., Zhang, H., Hou, Y., Zhuang, Y., Miranda, L., 2009. High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial *cox1* and *cob* for dinoflagellate DNA barcoding. *Appl. Environ. Microbiol.* 75, 1279–1290.
- Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C., Moreira, D., 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* 409, 603–607, <http://dx.doi.org/10.1038/35054537>.
- Makarieva, A.M., Gorshkov, V.G., Li, B.L., Chown, S.L., Reich, P.B., Gavrilov, V.M., 2008. Mean mass-specific metabolic rates are strikingly similar across life's major domains: evidence for life's metabolic optimum. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16994–16999.
- Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* 3, 2229–2238, <http://dx.doi.org/10.1371/journal.pbio.0030422>.
- Mitchell, E.A.D., Gilbert, D., 2004. Vertical micro-distribution and response to nitrogen deposition of testate amoebae in Sphagnum. *J. Eukaryot. Microbiol.* 51, 480–490.
- Morgan, J.L., Darling, A.E., Eisen, J.A., 2010. Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS ONE* 5, 10209.
- Moritz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. *PLoS Biol.* 2, 1529–1531, <http://dx.doi.org/10.1371/journal.pbio.0020354>.
- Nassonova, E., Smirnov, A., Fahrni, J., Pawlowski, J., 2010. Barcoding amoebae: comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist* 161, 102–115, <http://dx.doi.org/10.1016/j.protis.2009.07.003>.
- Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S.S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A.M., Gile, G.H., Holzmann, M., Jahn, R., Jirku, M., Keeling, P.J., Kostka, M., Kudryavtsev, A., Lara, E., Lukes, J., Mann, D.G., Mitchell, E.A.D., Nitsche, F., Romeralo, M., Saunders, G.W., Simpson, A.G.B., Smirnov, A.V., Spouge, J.L., Stern, R.F., Stoeck, T., Zimmermann, J., Schindler, D., de Vargas, C., 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10, e1001419.
- Payne, R., Mitchell, E.A.D., 2009. How many is enough? Determining optimal count totals for ecological and palaeoecological studies of testate amoebae. *J. Paleolimnol.* 42, 483–495, <http://dx.doi.org/10.1007/s10933-008-9299-y>.
- Puth, M.-T., Neuhäuser, M., Ruxton, G.D., 2014. Effective use of Pearson's product-moment correlation coefficient. *Anim. Behav.*, <http://dx.doi.org/10.1016/j.anbehav.2014.05.003>.
- Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7, 355–364, <http://dx.doi.org/10.1111/j.1471-8286.2006.01678.x>.
- Ronquist, F., Huelsenbeck, J.P., van der Mark, P., 2005. MrBayes 3.1. School of Computational Science. Florida State University, Tallahassee, Florida.
- Schwarz, M.V.J., Frenzel, P., 2003. Population dynamics and ecology of ciliates (Protozoa, Ciliophora) in an anoxic rice field soil. *Biol. Fertil. Soils* 38, 245–252, <http://dx.doi.org/10.1007/s00374-003-0644-z>.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771.
- Von Wintzingerode, F., Gobel, U.B., Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* 21, 213–229.
- Weber, A.A.T., Pawlowski, J., 2013. Can abundance of protists be inferred from sequence data: a case study of foraminifera. *PLoS ONE* 8, e56739, 10.1371.
- Zhu, F., Massana, R., Not, F., Marie, D., Vaultot, D., 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* 2, 79–92, <http://dx.doi.org/10.1016/j.femsec.2004.10.006>.