

## RESOURCE ARTICLE

# A needle in a haystack: A new metabarcoding approach to survey diversity at the species level of Arcellinida (Amoebozoa: Tubulinea)

Rubén González-Miguéns<sup>1</sup>  | Emilio Cano<sup>1,2</sup> | Antonio Guillén-Oterino<sup>3</sup> |  
Antonio Quesada<sup>4</sup> | Daniel J. G. Lahr<sup>5</sup>  | Daniel Tenorio-Rodríguez<sup>3</sup> |  
David de Salvador-Velasco<sup>3</sup> | David Velázquez<sup>4</sup> | María Isabel Carrasco-Braganza<sup>3</sup> |  
R. Timothy Patterson<sup>6</sup> | Enrique Lara<sup>1</sup>  | David Singer<sup>6,7</sup> 

<sup>1</sup>Department of Mycology, Real Jardín Botánico (RJB-CSIC), Madrid, Spain

<sup>2</sup>Research Support Unit, Real Jardín Botánico (CSIC), Madrid, Spain

<sup>3</sup>Estación Biológica Internacional Duero-Douro (EUROPARQUES-EBI), Buque hidrográfico Helios-Cousteau en el Lago de Sanabria, Castilla y León, Spain

<sup>4</sup>Department of Biology, Universidad Autónoma de Madrid, Madrid, Spain

<sup>5</sup>Department of Zoology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil

<sup>6</sup>Ottawa-Carleton Geoscience Centre and Department of Earth Sciences, Carleton University, Ottawa, Ontario, Canada

<sup>7</sup>UMR CNRS 6112 LPG-BIAF, Laboratory of Planetology and Geosciences, Angers University and Nantes Université, Angers, France

## Correspondence

Rubén González-Miguéns, Real Jardín Botánico (RJB-CSIC), Plaza Murillo 2, 28014 Madrid, Spain.  
Email: [rgmiguens@rjb.csic.es](mailto:rgmiguens@rjb.csic.es)

Enrique Lara, Real Jardín Botánico (RJB-CSIC), Plaza Murillo 2, 28014 Madrid, Spain.  
Email: [enrique.lara@rjb.csic.es](mailto:enrique.lara@rjb.csic.es)

## Funding information

Atracción de Talento Investigador, Grant/Award Number: PGC2018-094660B-I00 /10.13039/501100011033/; FORESTAT, Grant/Award Number: 3976-CT\_RD\_AMI\_18\_SURV\_FORESTAT; National Geographic Society, Grant/Award Number: EC-86022R-21

**Handling Editor:** Pierre Taberlet

## Abstract

Environmental DNA-based diversity studies have increased in popularity with the development of high throughput sequencing technologies. This permits the potential simultaneous retrieval of vast amounts of molecular data from many different organisms and species, thus contributing to a wide range of biological disciplines. Environmental DNA protocols designed for protists often focused on the highly conserved small subunit of the ribosome gene, that does not permit species-level assignments. On the other hand, eDNA protocols aiming at species-level assignments allow a fine level ecological resolution and reproducible results. These protocols are currently applied to organisms living in marine and shallow lotic freshwater ecosystems, often in a bioindication purpose. Therefore, in this study, we present a species-level eDNA protocol designed to explore diversity of Arcellinida (Amoebozoa: Tubulinea) testate amoebae taxa that is based on mitochondrial cytochrome oxidase subunit I (COI). These organisms are widespread in lentic water bodies and soil ecosystems. We applied this protocol to 42 samples from peatlands, estuaries and soil environments, recovering all the infraorders in Glutinoconcha (with COI data), except for Hyalospheniformes. Our results revealed an unsuspected diversity in morphologically homogeneous groups such as *Cylindrothecina*, *Excentrostoma* or *Sphaerothecina*. With this protocol we

David Singer and Enrique Lara are joint senior authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

expect to revolutionize the design of modern distributional Arcellinida surveys. Our approach involves a rapid and cost-effective analysis of testate amoeba diversity living in contrasted ecosystems. Therefore, the order Arcellinida has the potential to be established as a model group for a wide range of theoretical and applied studies.

**KEYWORDS**

Arcellinida, eDNA, metabarcoding, mitochondrial genes, species-level, testate amoeba

## 1 | INTRODUCTION

The use of environmental DNA (eDNA) has revolutionized biodiversity studies (Taberlet et al., 2012; Yoccoz, 2012), culminating with the development of high throughput sequencing (HTS) technologies (Shokralla et al., 2012). The application of HTS to amplicon sequencing (i.e., metabarcoding) allowed the retrieval of massive amounts of genetic data from many different organisms. This breakthrough has contributed significantly to academic research, especially in the fields of ecology, evolution or biogeography (Beng & Corlett, 2020; Boenigk et al., 2018; Gaither et al., 2022) and, in applied research, to environmental quality assessment (Lezcano et al., 2017; Pawlowski et al., 2021), among others. Nevertheless, to draw sound ecological or biogeographical conclusions based on metabarcoding data, an accurate reference taxonomic database is required. This database must be as exhaustive as possible, with representatives of the main clades that constitute the surveyed groups. This is the case for the most popular eukaryotic metabarcoding marker, the gene coding for the small subunit of the ribosome known as 18S rRNA gene (Berney et al., 2017; del Campo et al., 2018). The eukaryotic 18S reference database PR<sup>2</sup> (Guillou et al., 2013) is very large and currently comprises (last accessed 06 June 2022) almost 200,000 taxonomic annotated sequences, which makes possible a relatively accurate placement of newly obtained environmental sequences into known clades. This extensive taxonomic database and the universal presence of 18S rRNA in all eukaryotes has permitted the development of generic eDNA PCR protocols to retrieve a large part of environmental biodiversity without requiring prior knowledge of the taxonomic composition of communities (but see Vaulot et al., 2022). However, despite these attributes, the 18S rRNA gene is too conserved to be used for species identification in both micro- and macroscopic sized eukaryotes (Heger, Pawlowski, et al., 2011; Lara et al., 2022; Pinseel et al., 2020; Tang et al., 2012). Due to this the taxonomic resolution of the 18S rRNA gene is usually treated as operational taxonomic units (OTUs), which clusters the sequences with regards to their similarity, but lumps organisms with divergent ecological functions (Lara et al., 2022). As species are the only truly biologically and ecologically meaningful (and comparable) units of diversity in eukaryotes (Bickford et al., 2007; de Queiroz, 2005; Wiley, 1978), specific protocols designed to discriminate down to species-level and curated databases are necessary in eDNA biodiversity studies.

Species-level metabarcoding protocols based on fast evolving genes have already been successfully (and routinely) applied in

macro-organism (e.g., plants, animals or fungi; Ficetola et al., 2008; Richards et al., 2012). Metabarcoding protocol applications to bioindication (Czechowski et al., 2020), conservation biology (Cilleros et al., 2019; Pflieger et al., 2016), invasive species monitoring (Ardura et al., 2015) or biogeography assessment (West et al., 2021) allow the retrieval of large amounts of molecular data without direct observation of the species, facilitating fieldwork (Christianson et al., 2022). In protists, species level metabarcoding protocols are only seldom applied. Most diversity surveys are based on 18S rRNA gene which remain necessarily relegated to more theoretical-generalist topics (Pawlowski et al., 2016). Still, protists could be subject to applied research, as for instance bioindication of environmental disturbance. These microorganisms are considered excellent bioindicators due to their ubiquity, rapid response to environmental changes and short lifecycles (Payne, 2013). As a starting point to develop species-level metabarcoding studies, barcoding protocols that allow species level discrimination have been proposed for protists, notably based on the nuclear internal transcribed spacer (ITS), the ribulose-bisphosphate carboxylase gene (*rbcl*) or the mitochondrial cytochrome oxidase subunit I (*COI*). These former barcodes were first tested against isolated and already characterized organisms, proving that they provide a good resolution at the species level (Heger, Pawlowski, et al., 2011; Kostka et al., 2017; Macher et al., 2021; Moniz & Kaczmarek, 2009). Despite the proven success of these approaches for (organism) barcoding, protocols have been seldom transferred to environmental studies. The exceptions have been ecological research involving diatom *rbcl* barcodes (Hamsher et al., 2011; Kermarrec et al., 2013), for which there is an extensive database (Rimet et al., 2019) and, most recently, *COI* in foraminiferans (Girard et al., 2022). Both of these groups are excellent ecological and paleoecological bioindicators in marine environments. In addition, diatoms are also studied in freshwater systems, including streams and a variety of shallow lotic settings. Diatoms are also widely used as important proxies for ecological status assessment (e.g., the European Water Framework Directive; WFD; European Commission, 2000). The last decade has seen the development of cutting-edge metabarcoding protocols for diatom profiling that promise to eventually provide rapid, affordable and reliable tools for routine environmental ecological assessment in both freshwater and marine ecosystems, (Kermarrec et al., 2014; Mortágua et al., 2019). Routine and wide-scale application of the technology has not occurred yet though as standardization of eDNA and bioinformatic protocols are still pending (Baillet et al., 2020). Other important groups have barely been studied though with

species-level metabarcoding protocols. We propose here that Arcellinida (Amoebozoa: Tubulinea), also known as lobose testate amoebae, an important intermediary trophic group widespread in lentic water bodies and soil ecosystems, are an ideal test bed upon which to develop as species-level eDNA barcoding protocol in these environments.

Arcellinida is a group of amoeboid protists with a worldwide distribution, that can be found in soil, freshwater sediments, plankton and, to a lesser extent, marine environments where they inhabit interstitial zones (Smith et al., 2009). These organisms have a characteristic self-constructed test or shell whose size, shape and composition is used for taxonomic identification (Figure 1). In most studies these organisms are primarily identified and enumerated morphologically and are routinely used to provide insights into both substrate character and the general ecological health of lakes (Patterson & Kumar, 2002). They are routinely used for specific applications such as monitoring perturbations in lacustrine systems including; heavy metal contamination (Nasser et al., 2020; Yang et al., 2011), road salt loading (Cockburn et al., 2020) and eutrophication (Macumber et al., 2020). In terrestrial habitats, they are used in peatland restoration management (Carballeira & Pontevedra-Pombal, 2021; Marcisz et al., 2016; Valentine et al., 2013), or to evaluate the impact of open-cast coal mining on soil (Wanner & Dunger, 2001). There is thus an extensive scientific literature on the sensitivity of these organisms to different environmental perturbations (Freitas et al., 2022). In addition to their use as bioindicators, Arcellinida have also been proposed as a model group for studying microbial biogeography (Heger, Lara, & Mitchell, 2011; Singer et al., 2019; Smith & Wilkinson, 2007).

Nevertheless, recent studies involving the application of molecular protocols have provided new insights on both the diversity and taxonomy of Arcellinida, which was previously only based on morphological identification. For example, the systematic barcoding of single species with the variable mitochondrial marker COI revealed the existence of cryptic taxa not identifiable through morphological analysis alone (González-Miguéns, Soler-Zamora, Villar-Depablo, et al., 2022; Kosakyan et al., 2012; Singer et al., 2015). Although morphologically (almost) indistinguishable, these so-called “pseudocryptic” species do differ from each other according to their ecological preferences (Singer et al., 2018) and their geographic distributions (Singer et al., 2019). The combination of barcoding, distributional and ecological evidence indicates that these taxa should be considered as true biological species (Lara et al., 2020). Given the difficulties with the identification of some taxa based solely on their morphology, a molecular approach is required for the recollection of reliable distributional data. In addition to reducing costs and handling time, the application of a metabarcoding protocol for Arcellinida biodiversity assessment allows to overcome the previously described morphological identification issues.

In this study, we introduce a species-level protocol to characterize Arcellinida diversity based on eDNA samples. We describe an eDNA extraction and amplification protocol based on an Arcellinida-specific primer, using the mitochondrial barcode gene COI. Finally,

we describe a bioinformatic pipeline to analyse the resulting data. As a case study, we applied this protocol to a collection of samples from the Western Palearctic region covering sites with naturally high (peatlands) and low (estuaries) diversity/density of Arcellinida. We subsequently generated an environmental sequences database, which consists of a collection of ASVs that can be used in future applied studies. This database is built with both Sanger sequencing of cloned environmental amplicons as well as Illumina sequencing, HTS. We also compared our metabarcoding data with morphological observations to explore the capacities of our polyphasic methodology to draw conclusions on the environmental diversity of Arcellinida.

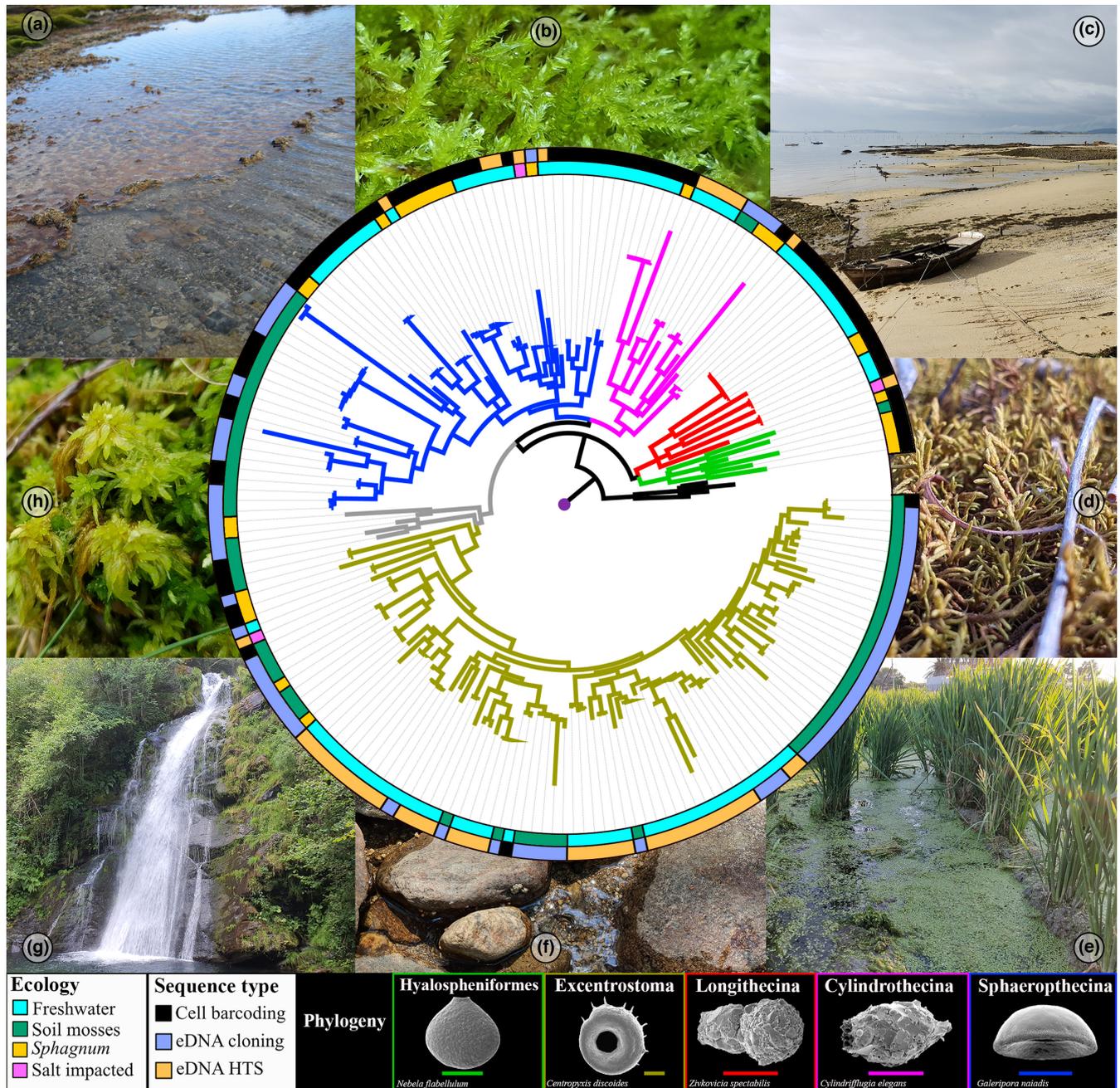
## 2 | MATERIALS AND METHODS

### 2.1 | Sampling

For this study we collected 42 samples from Europe, North America and the Antarctic Peninsula. We sampled freshwater sediments, terrestrial mosses, bogs (*Sphagnum*) and marine (interstitial) environments (Table S1). Soil and *Sphagnum* samples were collected following (González-Miguéns, Soler-Zamora, Villar-Depablo, et al., 2022) with the mosses and sediment being immersed in a falcon tube with distilled water. The moss was then squeezed to release the protist community, which was allowed to collect in the bottom of the tube to concentrate the testate amoebae. One ml of this concentrate was collected for subsequent analysis. Freshwater substrate samples were collected using a sterile plastic Pasteur pipette or a syringe (when scuba diving), collecting only the top-most few millilitres of sediment. Again, only 1 mL aliquot of this sample was used for subsequent analysis. Samples were fixed and conserved within LifeGuard Soil Preservation Solution (Qiagen) (in a proportion 1:1, with respect to the sample) or preserved at  $-20^{\circ}\text{C}$  until DNA extraction (Mazei et al., 2015). All samples were collected with different and sterile plastic Pasteur pipettes and Falcon tubes to avoid cross contaminations. Samples from Frog and Oromocto lakes (New Brunswick, Canada), and Nantes, Kérourag, Kergironé and Beg ar Graz estuaries (France) were sequenced using Illumina Novaseq and MiSeq platforms, respectively. For the rest of the samples, we cloned environmental amplicons and sequenced colonies by Sanger dideoxy method (Table S1).

### 2.2 | eDNA extraction and amplification

The sampling and laboratory works performed were inspired by the review of (Goldberg et al., 2016). Total eDNA was extracted using the DNeasy PowerSoil Pro Kit (Qiagen) and FastDNA Spin Kit for Soil (MP Biomedicals), following the instructions provided by the manufacturer. In each round of eDNA extraction, a negative control was performed to track potential contamination. The extracted eDNA was conserved at  $-20^{\circ}\text{C}$  until further processing. Then, we amplified a portion of the cytochrome oxidase subunit I (COI), using



**FIGURE 1** Maximum likelihood phylogenetic tree of Arcellinida based on cytochrome oxidase subunit I (COI) with the branch colours representing the infraorders in Glutinoconcha. The inner circle represents the ecosystem type from where organisms derive and the outer circle the origin of sequences. Pictures represents the localities where the samples were taken (a) Antarctic, (b) wet mosses, (c) mosses near the sea, (d) dry mosses, (e) eutrophicated freshwater, (f) sediment from lake, (g) sediment from a freshwater river and (h) *Sphagnum*. Scanning electron microscopy photographs showing the shell morphology of different species within Glutinoconcha.

a two-step nested polymerase chain reaction (PCR) protocol specifically designed for Arcellinida. This protocol can also be accessed from [10.17504/protocols.io.yxmv2389g3p/v1](https://doi.org/10.17504/protocols.io.yxmv2389g3p/v1):

(1) The first PCR had a final volume of 11  $\mu$ L containing: 1  $\mu$ L of DNA template, 1  $\mu$ L of distilled water, 1  $\mu$ L of bovine serum albumin (BSA), 6  $\mu$ L MyTaq Red DNA polymerase Mix (BioLine) and 1  $\mu$ L of each primer (10  $\mu$ mol). We used the universal COI primers pair developed by Folmer et al. (1994) LCO 1490 (5' GGTCACAAATCAT AAAGATATTGG 3'), as forward, and HCO 2198 (5' TAAACTTCAGG

GTGACCAAAAAATCA 3') as reverse. The PCR cycling profile was as follows: initial denaturation at 96°C for 5 min, followed by 40 cycles at 94°C for 15s; 40°C for 15s and 72°C for 90s and a final extension step at 72°C for 10 min. The resultant fragment was 692 bp long.

(2) The second PCR had a final volume of 21  $\mu$ L containing 1  $\mu$ L of product amplification of the first PCR, 6  $\mu$ L of distilled water, 12  $\mu$ L MyTaq Red DNA polymerase Mix (BioLine) and 1  $\mu$ L of each primer (10  $\mu$ mol). We used LCO 1490 as the forward primer and the Arcellinida-specific primer developed in (González-Miguéns,

Soler-Zamora, Villar-Depablo, et al., 2022) ArCOIR (5' CCACYNGA ATGWGCTARAATACC 3'), as reverse, with the following PCR cycling profile. There was an initial denaturation at 96°C for 5 min, followed by 40 cycles at 94°C for 15 s, 55°C for 15 s and 72°C for 90 s, and a final extension step at 72°C for 10 min. The derived fragment was 407 bp long. To multiplex these samples, we developed custom dual indexes with spacers primers LCO and ArCOIR. A list of these primers is included in Table S2.

Each PCR was replicated three times, and then all replicates were pooled together to mitigate the effects of the PCR biases. These PCRs were performed in a "PCR workstations" equipped with a UV hood and HEPA filter, using filter tips and negative-positive controls, to avoid and check for potential contaminations. We used 3 µL of each PCR reaction product and analysed them by electrophoresis on a 1% agarose gel. The PCR products of the second PCR were then purified, by band excision. Finally, we quantified the eDNA using a Qubit 3 fluorometer, with dsDNA high sensitivity (HS) assay kits (ThermoFisher) and stored at -20°C.

### 2.3 | Sequencing

First, we evaluated the specificity of the ArCOIR primer in eDNA samples by cloning the PCR products, based on 32 samples (Figure 1; Table S1). As the incorporation of the insert in the vector is a stochastic process depending on the inserts proportion and size, we consider that cloning/Sanger sequencing is a good approach to test the specificity of the primers prior to the application of HTS technologies. We used the pGEM T-Easy Vector System kit (Promega) following the manufacturer's recommendations. We selected 35 to 40 recombinant clones per library and amplified the insert with the plasmid primers T7 (5' TAATACGACTCACTATAGGG 3') and SP6 (5' GATTTAGGTGACACTATAG 3') universal primers, using the following volumes: the colony without performing a DNA extraction step, 6 µL of distilled water, 12 µL MyTaq Red DNA polymerase Mix (BioLine) and 1 µL of each primer (10 µmol). The resulting PCR of the colonies were sequenced with Sanger dideoxy-nucleotides technology, by the company Macrogen.

To test the robustness and the validity of our methods we used two different sequencing platforms and companies to sequence the amplicons. Samples from the Canadian lakes (FRO2, ORO2, Q1, Q2 and Q3) were sequenced using a NovaSeq 6000 (Illumina) by the company Novogene UK. The samples from the French estuaries (13, 35, 48, 78 and 79) were sequenced using MiSeq (Illumina) by the company ID-Gene Ecodiagnostics. In both cases the company performed a purification and library, by adapter ligation, and generated paired-end 250 bp reads.

### 2.4 | Data curation

The resulting cloning PCR sequences were checked and trimmed from cloning vector sequence and primer using the software Geneious Prime version 2019.0.4. We only considered sequences

which were obtained from at least three independent colonies. To corroborate the identity of the sequences as Arcellinida, we performed blastn analyses (Altschul et al., 1990) against both the whole GenBank database and our own database (González-Miguéns, Todorov, Blandenier, et al., 2022) (accessible in supplementary material as "pipeline.rar").

The Illumina reads were curated using the following pipeline: (1) Primers trimming and demultiplexing was done with Cutadapt version 2.8 (Martin, 2011), following the pipeline "Cutadapt\_Pipeline.bash" (provided as supplementary materials "pipeline.rar"). (2) Reads were ordered in the same direction using MAFFT version 7.490 (Katoh & Standley, 2013); (3) the resulting reads per sample were analysed with the dada2 R package (Callahan et al., 2016), following the tutorial in [https://benjineb.github.io/dada2/tutorial\\_1\\_8.html](https://benjineb.github.io/dada2/tutorial_1_8.html), with minor modifications (supplementary material as "pipeline.rar"). First, we visualized the quality profiles of the forward and reverse reads, to filter and trim them, truncating the reads after positions 210 and 200 for the forward and reverse, respectively. We dereplicated the reverse and forward reads, and then merged the paired reads, generating amplicon sequence variants (ASV). Then, we removed chimeric sequences. We only considered the ASV that were composed of at least 10 reads to reduce the consequences of methodological errors such as "tag jumping" (Schnell et al., 2015). (4) For the taxonomic assignment we used VSEARCH version 2.14.1 (Rognes et al., 2016) using a curated database from González-Miguéns, Todorov, Blandenier, et al. (2022), including COI sequences from GenBank public database (Amoebozoa, Archaeplastida, Excavata, Opisthokonta and SAR) (supplementary materials "pipeline.rar"). The ASV with an identity value higher than 0.82 with the Arcellinida database were checked in GenBank to verify their taxonomic assignment.

### 2.5 | Phylogenetic analysis

The sequences of environmental clones and ASV that were assigned to Arcellinida were aligned with the Arcellinida COI sequences data of González-Miguéns, Todorov, Blandenier, et al. (2022), using the MAFFT auto algorithm (Katoh et al., 2002) as implemented in Geneious version 2019.0.4 (accessible in supplementary material as "Arcellinida\_COI\_alignment.phy"). Tree topologies and node supports were evaluated with maximum likelihood (ML), using IQ-TREE2 version 2.0 (Minh et al., 2020). The best substitution models were selected with ModelFinder (Kalyaanamoorthy et al., 2017), implemented in IQ-TREE2 version 2.0. Node supports were assessed with 5000 ultrafast bootstrap replicates approximation (Hoang et al., 2018; Minh et al., 2013). The trees obtained were edited in FigTree version 1.4.3. Scanning electron microscopy images in the figures were taken from Todorov and Bankov (2019).

### 2.6 | Seqlogo graphs

We used the ggseqlogo R package (Wagih, 2017) to create sequence logos, a graphical representation of the sequence conservation of

nucleotides for COI and SSU specific Arcellinida primers across the groups molecular diversity. These illustrations allow to graphically visualize mismatches between primers and binding sites that can generate potential amplification biases.

We implemented these representations for our COI specific primer ArcCOIR for each Glutinoconcha infraorder, as well as for the outgroup (Tubulinea amoebae). Arcellinida COI sequences were retrieved from (González-Miguéns, Todorov, Blandenier, et al., 2022) as well as from the GenBank database, from which we extracted the ArcCOIR primer region.

Moreover, we also represented potential mismatches with specific primers published in the literature that target the SSU gene. We also created Seqlogo graphs for each Arcellinida infraorder and suborder. The primers evaluated were 2345 and 2088 (Ruggiero et al., 2020), each one in its corresponding binding site. Arcellinida SSU sequences were retrieved from "Supplementary data 5" of González-Miguéns, Todorov, Blandenier, et al. (2022), as it includes the most comprehensive database for this gene, including both bar-coding and transcriptomics data.

## 2.7 | Shell morphology versus COI molecular marker

We compared our molecular data with Arcellinida biodiversity assessments based on morphology from the same Canadian lakes samples (Steele et al., 2018). This study was based on a total of 43 samples, all of them divided into three sectors "Q1", "Q2" and "Q3". Our samples "Q1", "Q2" and "Q3", sequenced by Illumina, correspond with a single sample per quadrant described in Steele et al. (2018), labelled with the same name. Those samples correspond to a replicate of that study (in preparation).

To compare Arcellinida morphotypes retrieved in Steele et al. (2018) with our molecular ASVs, we clustered them into operational taxonomic units (OTU); as most of these ASVs cannot be assigned into known species already. We define the OTUs as the groups of ASVs that share an uncorrected pairwise distances ("p") equal or less than 5% of genetic divergence. The OTU composition of the samples Q1–3 was compared with the morphospecies recovered in Steele et al. (2018) for each family recovered in the study (Table S3). As the families Diffflugidae and Cylindrifflugidae are practically impossible to differentiate morphologically from each other (González-Miguéns, Todorov, Blandenier, et al., 2022) we merged these taxa within a family "Difflugiidae". In order to evaluate the "coverage" of morphology versus molecular, we evaluate the coverage (C) as the ratio between the number of morphospecies retrieved in Steele et al. (2018) and the number of molecular OTUs (morphological OTUs/molecular OTUs). Where  $C = 1$  if the number of morphospecies is equal to the number of molecular OTUs,  $C > 1$  if morphological diversity is higher than the assayed molecular one and  $C < 1$  if molecular diversity is higher than morphological.

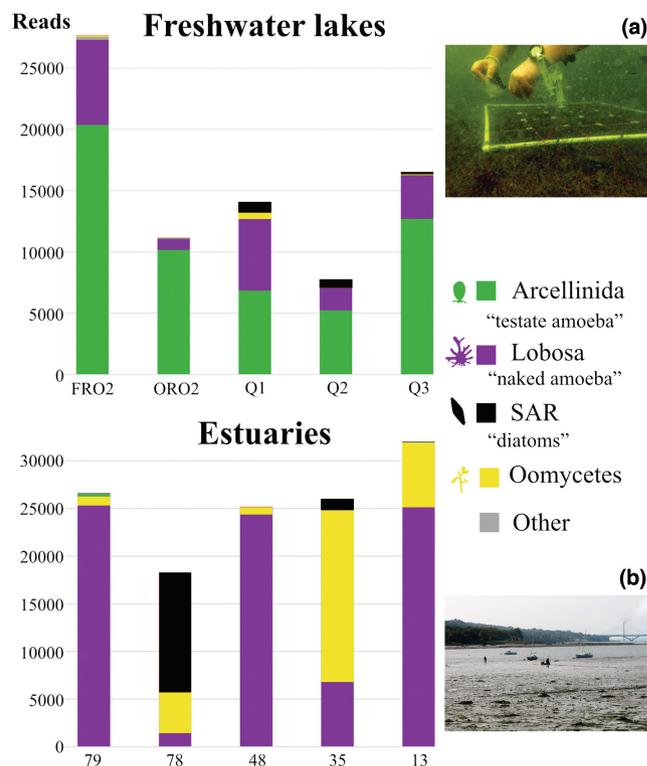


FIGURE 2 Number of reads recovered for each sample sequenced by Illumina. The colour represents the taxonomic composition. The pictures illustrate the type of environment in which samples were taken (a) Q1, Oromocto Lake (Canada) and (b) 13, Elorn (France).

## 3 | RESULTS

### 3.1 | Molecular results

We obtained a total of 82 Arcellinida clonal sequences, each of them being recovered in at least three independent colonies per sample (Figure 1 and Table S1). In all environments surveyed (freshwater sediments, soil and bogs), Arcellinida clonal sequences dominated together with a minor representation of sequences classified as diatoms or naked amoebae.

The number of Illumina reads per sample is detailed in (Table S4), as well as the number of reads assigned to Arcellinida (Figure 2 and Table S5). We obtained in total 61 ASV classified as Arcellinida. The samples from freshwater lake sediment provided the highest proportion of Arcellinida, representing 73%, 91%, 48%, 67% and 76% from the total of reads, in the samples FRO2, ORO2, Q1, Q2 and Q3, respectively. The samples from estuary sediment provided low numbers of Arcellinida reads: 1%, 0.08%, 0.03%, 0.04% and 0.1% from the total of reads, in the samples 79, 78, 48, 35 and 13, respectively. The best represented eukaryotic clade outside Arcellinida was Lobosea "naked amoebae" (Figure 2), with 25%, 7%, 41%, 23% and 21% from the total of reads, in the samples FRO2, ORO2, Q1, Q2 and Q3, respectively (freshwater sediments). This was also the case in estuary sediment samples with 95%, 7%, 96%, 26% and

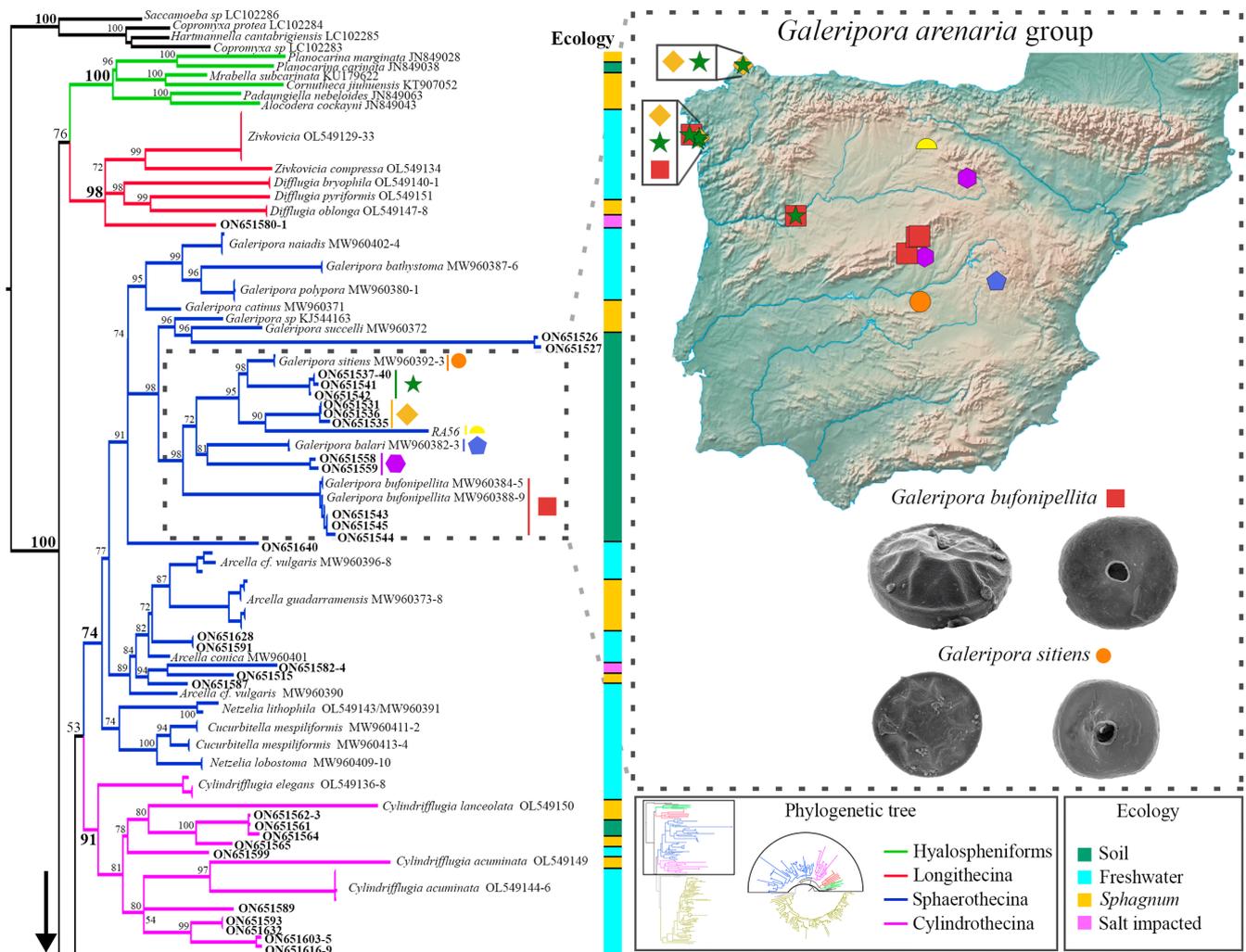
78% from the total of reads, in the samples 79, 78, 48, 35 and 13, respectively.

### 3.2 | Phylogenetic composition of Arcellinida communities

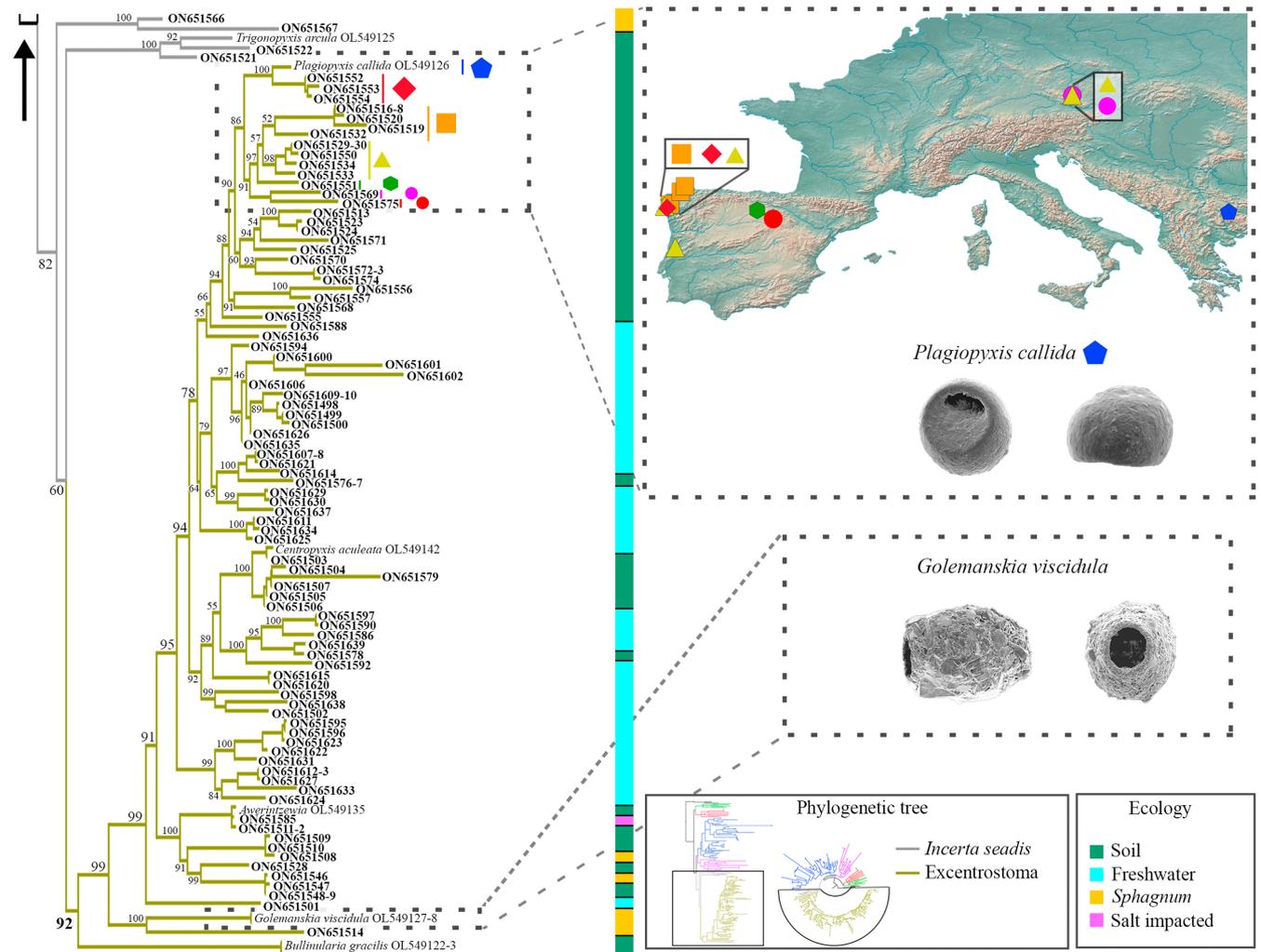
In total we obtained 143 environmental sequences assigned to Arcellinida by combining cloning/sequencing and Illumina (Table S1). These eDNA COI sequences were used in combination with sequences obtained from the single documented organisms to build a phylogenetic tree, whose topology was congruent with previous works (González-Miguéns, Todorov, Blandenier, et al., 2022) (Figure 1). Notably, we recovered the monophyly of Arcellinida with a maximum likelihood bootstrap (ML) of 100 (Figure 3). The tree also recovered the monophyly of the Glutinoconcha infraorders Hyalospheniformes ML = 100 (green colour in Figures 1 and 3),

Longithecina ML = 98 (red colour in Figures 1 and 3), Sphaerothecina ML = 74 (blue colour Figures 1 and 3), Cyndrothecina ML = 91 (purple colour in Figures 1 and 3) and Excentrostoma ML = 92 (yellow colour in Figures 1 and 4). Two clades did not classify among these infraorders; one of them included the characteristic species *Trigonopyxis arcula*; we labelled them Arcellinida *incertae sedis* (grey colour in Figures 1 and 4).

Environmental sequences were distributed among Glutinoconcha infraorders in the following way: Hyalospheniformes 0 (0 cloned [C] and 0 Illumina [I]), Longithecina 1 (0 C and 1 I), Sphaerothecina 20 (15 C and 5 I), Cyndrothecina 10 (4 C and 6 I), Excentrostoma 86 (48 C and 38 I) and *Incertae Sedis* 4 (4 C and 0 I) (Figures 1, 3 and 4). Some of these ASVs, 12 in total, corresponded to organisms that had been previously isolated and barcoded (González-Miguéns, Soler-Zamora, Useros, et al., 2022; González-Miguéns, Soler-Zamora, Villar-Depablo, et al., 2022). These taxa were *Galeripora bufonipellita* (3 ASVs assigned), *Centropyxis*



**FIGURE 3** Detail of the phylogenetic tree of Arcellinida based on cytochrome oxidase subunit I (COI) (Figure 1). The symbols on the map of the Iberian Peninsula represent the different localities where sequences of the group of *Galeripora arenaria* species complex have been found. At the bottom are SEM images for *Galeripora bufonipellita* and *G. sitiens*. The bold words at the tip of the tree represent the GenBank accessions for the eDNA ASVs obtained in this study.



**FIGURE 4** Detail of the phylogenetic tree of Arcellinida based on cytochrome oxidase subunit I (COI) (Figure 1) and the bottom part of the Figure 3. Symbols on the map of Europe represent the different localities where species from the *Plagiopyxis* clade have been found. At the bottom are SEM images for *Plagiopyxis callida* and *Golemanskia viscidula*. The bold words at the tip of the tree represent the GenBank accessions for the eDNA ASVs obtained in this study.

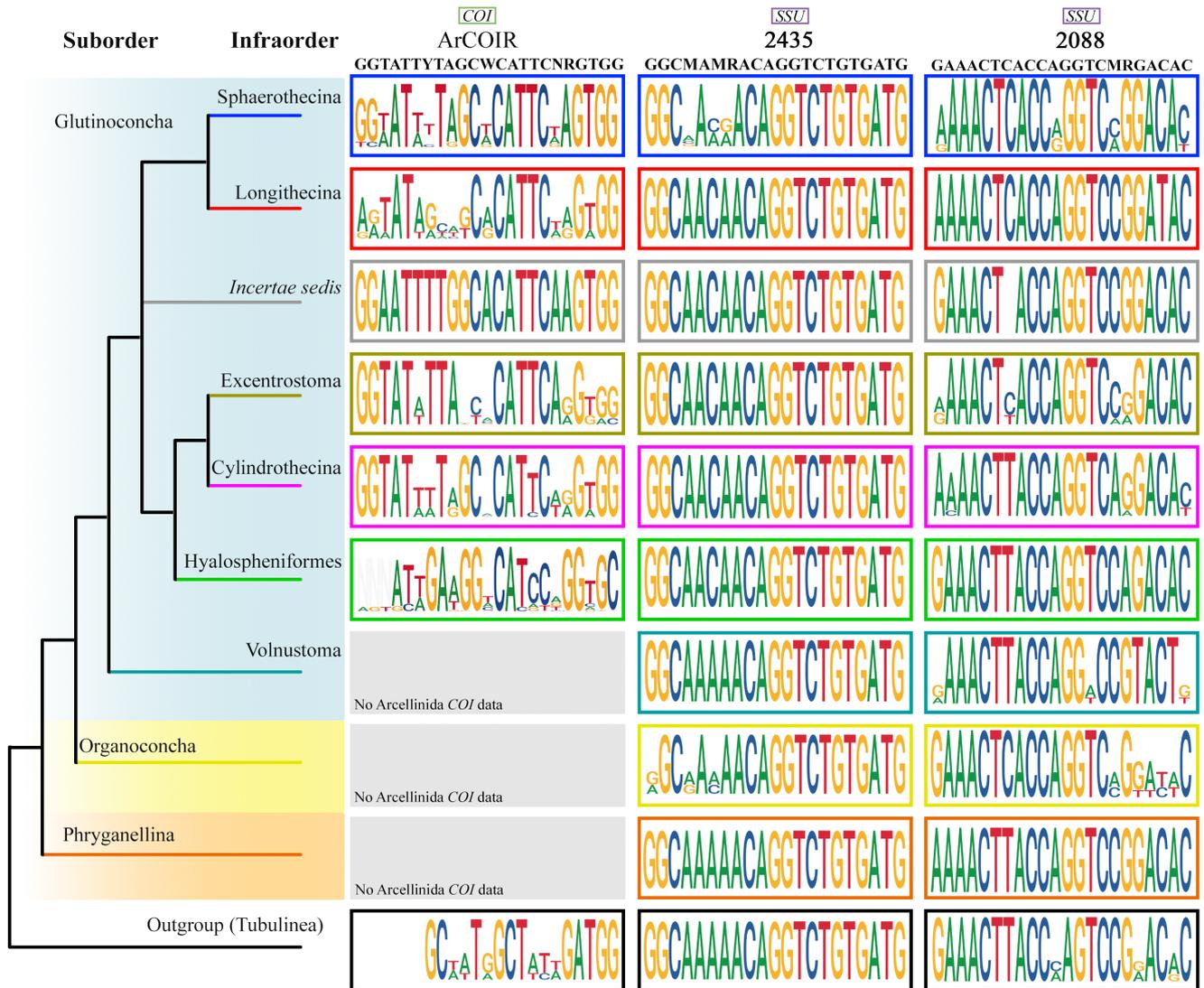
*aculeata* (6 ASVs assigned) and *Awerintzewia* sp (3 ASVs assigned). The lack of other species-level assignment of environmental clades is due to the COI database limited size. Nevertheless, future barcoding studies will refine the taxonomic assignment of the remaining ASVs.

Sequences from Hyalospheniiformes were not recovered, even though these organisms could be observed microscopically in the samples. This was probably because these organisms have a codon deletion at the beginning of the primer ArCOIR (Figure 5). The low coverage of Longithecina is also probably due to a mismatch of their COI sequence with the primer ArCOIR (Figure 5). Reference sequences of the infraorders Volvnutoma (within Glutinocoencha) and of the suborders Organocoencha and Phryganellina are not yet barcoded. The expansion of the barcode database will allow finding the limits of our protocol at deeper taxonomic levels. The situation is different with SSU; indeed, the specifically designed primers 2435 and 2088 are located in a highly conserved region which does not vary between the different infra- and suborders of Arcellinida, as well as between other Amoebozoa (Figure 5).

This may explain their low performance in eDNA results (Ruggiero et al., 2020).

### 3.3 | Morphology versus molecular results

Calculation of coverage can provide hints regarding the potential and any limitations of our metabarcoding protocol by comparing both morphological classification in addition to the molecular data, for each clade. Even though the sampling intended for metabarcoding was much smaller than the one in Steele et al. (2018), we observed that some clades were better represented in metabarcoding than by direct counts. Indeed, the families with less coverage (better represented in molecular data than in microscopic counts), were Arcellidae  $C = 0$  (no morphological observations) and Centropyxididae  $C = 0.57$ . On the other hand, the families with better morphological coverage were Diffugiidae  $C = 5$  and Netzeliidae  $C = -$  (not recovered in the molecular data). The  $C$  values for all taxa with respect to (Steele et al., 2018) are given in Table 1.



**FIGURE 5** Schematic illustration of Arcellinida phylogeny based on the results of González-Miguéns, Todorov, Blandenier, et al. (2022) and Lahr et al. (2019). On the right panel, sequence logos represent the primer binding sites. Letter sizes are proportional to the ratio of nucleotides on each position. Primer binding sites illustrated are those for ArCOIR (COI mitochondrial) and 2088–2345 (SSU, nuclear) all in 5′–3′ direction, based on sequences from the respective clades.

**TABLE 1** Coverage between morphospecies and molecular operational taxonomic units (OTU).

Family	Morphospecies	Molecular OTU	Coverage
Arcellidae	0	3	0
Netzeiliidae	3	0	-
Centropyxididae	4	7	0.57
Diffflugidae	20	4	5

## 4 | DISCUSSION

### 4.1 | Prerequisites for species level protocols in eDNA

As species are the only biologically and ecologically meaningful (and comparable) units of diversity, species-level resolution

in metabarcoding studies has many advantages. Cross-taxa comparisons can be generated including information about each single species such as distribution, ecological optimum and tolerance or ecological function. One of the requisites for a “resolutive” marker gene must be variable enough to reflect the evolutionary history of each taxonomic unit, because each species has its own historical and future evolutionary trajectory with respect to other clades (Wiley, 1978). In Arcellinida, these barcodes were integrated with shell morphology (Singer et al., 2015), ecology (Singer et al., 2018), biogeography (Singer et al., 2019), or with several molecular markers (González-Miguéns, Todorov, Blandenier, et al., 2022), to trace species delimitation. The mitochondrial cytochrome oxidase subunit I (COI) was shown to be the best adapted marker for species-level discrimination, in Arcellinida but also in other Amoebozoa (Nassonova et al., 2010; Tekle, 2014). This gene is, to date, one of the most used barcoding genes in eukaryotes (excepted plants) for species identification, and also for species level metabarcoding.

The fine taxonomic resolution reached with our eDNA protocol, based on COI will make easier the evaluation of biodiversity patterns such as ecosystem changes in future studies on Arcellinida, as it has been proposed for other taxa (Pawlowski et al., 2018). Such an approach may also allow inferences on functional diversity as well (Laroche et al., 2021). Because eDNA facilitates the retrieval of high amounts of molecular data at species level with minimal effort, it represents an ideal approach to these applications, as well as to increase the existing knowledge of the ecology or distribution of species. These practices are much more common in plants and animals, and bring along many applications, for instance to invasive species monitoring (Ficetola et al., 2008), food regime studies (Roffler et al., 2021) and even to the rediscovery of critically endangered species (Turunen et al., 2021; Villacorta-Rath et al., 2022). The diversity of these applications suggests the immense potential that lies in the application of species level metabarcoding to protists.

In microbial sized organisms, eDNA is commonly the only way to retrieve information from elusive species, without direct observation (Pawlowski et al., 2012). Often, however, eDNA sequences cannot be assigned to known groups/species, because databases are incomplete and no close related organisms has been characterized molecularly. These sequences resulting from lack of taxonomic information on microbes form so-called “environmental clades”, whose number and importance depend on the degree to which biodiversity has been barcoded. In our data, we observed only one deep branching eDNA clade (to the level of an infraorder), which cannot be assigned yet to any known Glutinoconcha infraorder. We designated this clade as *incertae sedis*, pending morphological characterization of these organisms (Figures 1 and 4).

Here, we provide a large dataset of Arcellinida environmental COI sequences (obtained through both cloning/sanger sequencing and Illumina) which includes all infraorders described as for now within Glutinoconcha, except Volnustoma and Hyalospheniformes, which are exclusively soil-dwelling lineages. While we do not possess sequences from Volnustoma, many Hyalospheniformes typically have a codon deletion on the binding site of the primer ArcCOIR (Figure 5). This concerns most Hyalospheniformes but not the clade which comprises genera *Alabasta*, *Alocodera*, *Apodera* and *Padaungiella*. Therefore, our protocol can be considered as rather universal for Arcellinida with the exception of the above-mentioned infraorders.

## 4.2 | SSU and COI primer biases in Arcellinida

The primer choice is vital in metabarcoding studies, as each pair biases, in its own way, the vision of biodiversity analyses will provide (Hajibabaei et al., 2019). Instead, primers need to be chosen in function of the initial hypothesis, and there is most often a trade-off between specificity and taxonomic resolution (Govender et al., 2022). When group-specific metabarcoding protocols have to be developed, primers located in conserved genes may amplify many “unwanted”

taxa and provide a low taxonomic resolution. Such an approach causes a loss in sequencing depth for the groups of interest, especially if these are minority in the general community (like Arcellinida for this case). These primers are nevertheless well adapted to explore domain-wide communities (Lentendu et al., 2023). In protists, primers located in the most conserved regions of the SSU can be even used to retrieve sequences from totally unknown groups, which are absent from molecular databases. For this reason, SSU has been (and still is) the principal marker gene for protist metabarcoding. In the other hand, COI does not contain regions that are conserved enough to capture all eukaryotes, which implies that general primers need to be highly degenerated (Collins et al., 2019). This generates limitations for most metabarcoding applications (Deagle et al., 2014), preventing the development of general protocols for all eukaryotes. However, when the aim is to develop a clade specific approach, this weakness can be turned into a strength. Indeed, less unwanted taxa are amplified which increases the amount of information about the target organisms; this allows studies focussed on alpha diversity of protist groups, an approach that cannot be followed using SSU as a metabarcoding molecular marker.

Like in other protist groups, primers have been designed for specific amplification of Arcellinida SSU. These primers were used to retrieve long SSU sequences from molecularly undocumented species (see Annex 1), or to screen biodiversity in peat bogs (Ruggiero et al., 2020) which are ecosystems where Arcellinida are particularly abundant and diverse (Gilbert & Mitchell, 2006). However, the issue of the coamplification of unwanted taxa appeared as Ruggiero et al. (2020) obtained a minority of Arcellinida ASVs in peatbogs. This is illustrated here in Figure 5 where we show that SSU “specific” primers 2088–2435 (Ruggiero et al., 2020) capture most Arcellinida diversity, but also other Amoebozoa and eukaryotes, such as Fungi. These represent actually the majority of all sequences (Ruggiero et al., 2020). In contrast, there is a need for faster evolving genes, and, among them, only COI can rely on a sufficiently extensive database from Arcellinida to design primers and identify ASVs (González-Miguéns, Todorov, Blandenier, et al., 2022). Mitochondrial editing, which has been characterized in Arcellinida, increases the amount of insertions and deletions in COI (Yang et al., 2017), which facilitates specific primers design. In Arcellinida, concretely, two codons have been deleted with respect to other Tubulinea (Figure 5). This is the region where the primer ArcCOIR has been designed, allowing a high degree of specificity with respect other eukaryotes, and reducing the co-amplification of other taxa, especially when Arcellinida are numerous (Figure 2).

## 4.3 | Implications on the current knowledge of Arcellinida's environmental diversity

The application of our metabarcoding protocol has challenged previous perceptions of Arcellinida diversity. The most conspicuous outcome is the immense diversity of infraorder Excentrostoma, which includes the COI-barcoded genera *Centropyxis*, *Plagiopyxis*,

*Bullinularia*, *Awerintzewia* and the recently described genus *Golemanskia* (Figures 1 and 4). This order includes a vast diversity of shell morphologies (González-Miguéns, Todorov, Blandenier, et al., 2022) well-beyond the typical compressed, off-centred apertured shell (Lahr et al., 2019). Here, we retrieved an immense diversity in all environments surveyed (including in estuaries), suggesting an unexpected diversity overall (Figures 1 and 4). As might be expected, the diversity of Arcellinida observed was closely related to the environment they were found in. For instance, *Galeripora arenaria* complex species are only recorded in soil (Figure 3) as observed with the *Plagiopyxis* group (Figure 4). On the other hand, several other well supported environmental groups belonged exclusively in aquatic sediments (Figure 4). This result suggests a widespread occurrence of phylogenetic niche conservatism that should be tested. This trend has been demonstrated previously in Hyalospheniformes, where tolerance towards nitrogen depletion in *Sphagnum* peatbogs is inherited within clades in the genus *Nebela* (Singer et al., 2018).

By comparing the presence of Arcellinida in freshwater versus estuary sediments it appears obvious that the first contain the highest diversity and most probably the highest individual numbers (Figure 2). Estuaries, because of their fast changing salinity and their dynamic nature, contain generally a low diversity for most taxa, including the macroscopic ones (Benfield, 2012). Arcellinida have been recorded only on a limited number of occasions in high salinity environments (Smith et al., 2009); they typically become scarcer in anchialine systems as the salinity increases (Van Hengstum et al., 2008). The few species characterized in these environments correspond to morphologies typical of Longithecina or Cylindrothecina, Excentrostoma, or Phryganellina. Therefore, the low number of reads and low diversity recovered for Arcellinida in this study (Figure 2), is supported by the direct observations, which advocates against potential biases due to primers mismatches. It seems therefore that Arcellinida species are primarily a freshwater taxon. Nevertheless, we recovered an ASV in the estuarine sediment which corresponds exactly to *Awerintzewia* sp., an isolate obtained from salt impacted bryophytes growing on a beach. The same sequence was also found again in a clone from a similar environment (Figure 4). This suggests a trend towards halophily in this species, and therefore that some Arcellinida at least are capable of crossing the salinity barrier and establish populations in higher salinity environments. Two other ASVs were also recovered in estuarine sediment, corresponding respectively to a member of Sphaerothecina as well as an unidentified Longithecina. The salinity barrier is one of the greatest ecological barriers (González-Miguéns, Soler-Zamora, Useros, et al., 2022), and those organisms that cross it must face important challenges through their evolutionary history. This characteristic has been shown in another group of (filose) testate amoebae, *Cyphoderia* (González-Miguéns, Soler-Zamora, Useros, et al., 2022). Therefore, it can be expected that further investigations in saline environments, both inland (saline endorrheic) and marine systems will reveal new insights on Arcellinida diversity.

#### 4.4 | Morphological-molecular integration and ArCOIR primer biases in Arcellinida applied studies

The comparison between morphological observations and OTUs (see morphology vs. molecular in Materials and Methods) through coverage indices reveals the great potential of our new approach for diversity exploration (Table 1). The families Arcellidae and Centropyxididae had low coverage indices (0 and 0.57, respectively, in Table 1), meaning that metabarcoding retrieved more diversity than the morphological observations in Steele et al. (2018). This was even more remarkable as our sampling design is much more reduced (3 vs. 43 samples) and the size of our samples was smaller (c. 1 cm<sup>3</sup> of unprocessed sediment versus about 5 cm<sup>3</sup> of concentrated sediment). Several explanations can be invoked to explain these differences between visual observations and metabarcoding results. Arcellinida have inconspicuous “naked states” in their lifecycles which are very difficult to observe within environmental samples (Dumack et al., 2020). In addition, many Arcellinida species are difficult to distinguish visually but are genetically different (cryptic diversity). The difficulty in species identification in Arcellinida, has already been observed by Foissner and Korganova (2000) when they hypothesized that *Centropyxis aerophila* was a cryptic species complex. These authors, by observing the immense morphological diversity continuum of the group, already hypothesized a diversity higher than expected. For practical reasons, these forms were pooled together in ecological studies, based on morphological data, into a “*Centropyxis aerophila*-type” (Amesbury et al., 2018).

In contrast, the family Difflogiidae shows a low coverage index (5 in Table 1), and no sequence has been detected for the family Netzeiliidae. Sample size certainly plays a role, as *Cucurbitella mespiliformis* has been detected visually but not by metabarcoding. Indeed, the primer ArcCOIR recognized only few mismatches with that species (Figure 5). Primer incompatibility may play a role in “Difflogiidae”, by reducing the likelihood of their COI genes being amplified during the PCR process (Figure 5). Indeed, several mismatches with ArCOIR can be found in some members of genus *Difflogia*, while this is less of an issue in *Cylindridifflogia* (Figures 1 and 2). Then, Arcellinida populations density are known to vary during the year, as their metabolic activity and reproduction rates are influenced by water temperature as shown for *Netzelia tuberspinifera* in the transcriptome-based study of Wang et al. (2022). A larger sampling design with repeated sessions along a full seasonal cycle should allow retrieving environmental sequences from the less represented families in this study.

## 5 | CONCLUSION

A well-designed metabarcoding protocol will recover most of the target diversity present in an environment. Our new Arcellinida metabarcoding protocol recovers all the infraorders in Glutinoconcha, with COI data (except for Hyalospheniformes and possibly Volnustoma) at the species level. It will allow for the rapid,

cost effective and reliable handling of data in potentially vast studies for the monitoring of ecosystem health, as well as ecological or biogeographic data from samples collected all over the world. In order to detect the remaining diversity, the primer ArcCOIR could be used in combination with infraorder specific primers (e.g., Hyalospheniformes-specific). The application of Arcellinida can fill the current monitoring gap on heterotrophic protists in lentic (e.g., lakes and reservoirs) and soil ecosystems, as their heterotrophic condition allows them to live where light does not reach. In addition, Arcellinida behave like microbial top predators in the ecosystems where they live, which means that they resume all the bioindicative information from the rest of the foodweb. Further studies that expand the reference COI database will improve the interpretative power of Arcellinida metabarcoding studies, and for example provide a deeper understanding of many areas of research with the group, such as the link between Arcellinida test traits and ecological functions (Marcisz et al., 2020). Expanding the database of barcoded species will provide researchers with a more accurate taxonomical identification of the ASVs, which will in turn allow for better ecological interpretations. Finally, the optimization of this protocol towards one-step PCR, with LCO-1490 and ArCOIR primers, could facilitate their use and reduce quantitative biases. Altogether, this information will contribute to make Arcellinida an ideal model group for theoretical and applied studies.

#### AUTHOR CONTRIBUTIONS

Rubén González-Miguéns, David Singer and Enrique Lara designed the study; Antonio Guillén-Oterino, Antonio Quesada, Daniel Tenorio-Rodríguez, David de Salvador-Velasco, David Singer, David Velázquez, Enrique Lara, María Isabel Carrasco-Braganza, Rubén González-Miguéns and R. Timothy Patterson carried out the fieldwork; Emilio Cano, David Singer and Rubén González-Miguéns performed the laboratory work; Rubén González-Miguéns prepared the images; Rubén González-Miguéns performed the analyses, interpreted and discussed the data together with David Singer and Enrique Lara; Rubén González-Miguéns and Enrique Lara wrote the manuscript and discussed together with Antonio Guillén-Oterino, Antonio Quesada, Daniel J. G. Lahr, Daniel Tenorio-Rodríguez, David de Salvador-Velasco, David Singer, David Velázquez, María Isabel Carrasco-Braganza and R. Timothy Patterson.

#### ACKNOWLEDGEMENTS

We express our gratitude to C. Soler-Zamora, F. Useros, Dr I. García-Cunchillos and M. Blázquez for helpful during the manuscript. We also acknowledge the help of Y. Ruiz-León and M. García-Gallo for the help in the molecular biology laboratory and M. Villar-De Pablo for the data processing and analyses. We are indebted to the following people that took part in the field work; A. Berlinches, A. Kosakyan, I. Cacabelos, Professor E. A. D. Mitchell, L. Hernandez, L. Lara-Hollenstein, M. Blázquez, M. Miguéns-Gomez, M. Villar-De Pablo, P. Lara-Hollenstein. We acknowledge Professor F. Jorissen and the project FORESTAT (funded by the OFB (Office français de la

biodiversité) and the University of Angers, grant number 3976-CT\_RD\_AMI\_18\_SURV\_FORESTAT) for the estuarine samples. This study was funded by the Spanish Government PGC2018-094660B-I00 and PID2021-128499NB-I00 10.13039/501100011033/, (MCIU/AEI/FEDER,UE) and an "Atracción de Talento Investigador" grant awarded by the Consejería de Educación, Juventud y Deporte, Comunidad de Madrid (Spain) (2017-T1/AMB-5210) to E.L.; a National Geographic Grant EC-86022R-21 to R. G. M. Fieldwork in New Brunswick Canada was funded by a Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant to R.T.P.

#### CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest.

#### DATA AVAILABILITY STATEMENT

The Arcellinida genetic data generated in this article have been made available in GenBank Nucleotide Database, can be accessed with the accession number ON651498-ON651640. Illumina Fastq files can be accessed through european nucleotide archive (ENA) by the accession number study PRJEB53915. The pipeline for the Cutadapt, dada2 and Vsearch programmes can be accessed in supplementary material with the file "pipeline.rar". The alignment used to create the phylogenetic tree of Figures 1, 2 and 3 can be accessed in the Supporting Information "Arcellinida\_COI\_alignment.phy".

#### BENEFIT-SHARING STATEMENT

Benefits generated: There are no benefits outlined in the Nagoya protocol associated with this study to report. All data produced in this research have been deposited in public databases as described above.

#### ORCID

Rubén González-Miguéns  <https://orcid.org/0000-0002-8274-266X>

Daniel J. G. Lahr  <https://orcid.org/0000-0002-1049-0635>

Enrique Lara  <https://orcid.org/0000-0001-8500-522X>

David Singer  <https://orcid.org/0000-0002-4116-033X>

#### REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Amesbury, M. J., Booth, R. K., Roland, T. P., Bunbury, J., Clifford, M. J., Charman, D. J., Elliot, S., Finkelstein, S., Garneau, M., Hughes, P. D. M., Lamarre, A., Loisel, J., Mackay, H., Magnan, G., Markel, E. R., Mitchell, E. A. D., Payne, R. J., Pelletier, N., Roe, H., ... Warner, B. G. (2018). Towards a Holarctic synthesis of peatland testate amoeba ecology: Development of a new continental-scale palaeohydrological transfer function for North America and comparison to European data. *Quaternary Science Reviews*, 201, 483–500. <https://doi.org/10.1016/j.quascirev.2018.10.034>
- Ardura, A., Zaiko, A., Martinez, J. L., Samulioviene, A., Semenova, A., & Garcia-Vazquez, E. (2015). EDNA and specific primers for early detection of invasive species – A case study on the

- bivalve *Rangia cuneata*, currently spreading in Europe. *Marine Environmental Research*, 112, 48–55. <https://doi.org/10.1016/j.marenvres.2015.09.013>
- Bailet, B., Apothéoz-Perret-Gentil, L., Baričević, A., Chonova, T., Franc, A., Frigerio, J.-M., Kelly, M., Mora, D., Pfannkuchen, M., Proft, S., Ramon, M., Vasselon, V., Zimmermann, J., & Kahlert, M. (2020). Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Science of the Total Environment*, 745, 140948. <https://doi.org/10.1016/j.scitotenv.2020.140948>
- Benfield, M. C. (2012). Estuarine Zooplankton. In *En Estuarine ecology* (pp. 285–302). John Wiley & Sons, Ltd.. <https://doi.org/10.1002/9781118412787.ch11>
- Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: Opportunities, challenges and prospects. *Biodiversity and Conservation*, 29(7), 2089–2121. <https://doi.org/10.1007/s10531-020-01980-0>
- Berney, C., Ciuprina, A., Bender, S., Brodie, J., Edgcomb, V., Kim, E., Rajan, J., Parfrey, L. W., Adl, S., Audic, S., Bass, D., Caron, D. A., Cochrane, G., Czech, L., Dunthorn, M., Geisen, S., Glöckner, F. O., Mahé, F., Quast, C., ... de Vargas, C. (2017). UniEuk: Time to speak a common language in Protistology! *Journal of Eukaryotic Microbiology*, 64(3), 407–411. <https://doi.org/10.1111/jeu.12414>
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3), 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Boenigk, J., Wodniok, S., Bock, C., Beisser, D., Hempel, C., Grossmann, L., Lange, A., & Jensen, M. (2018). Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding and Metagenomics*, 2, e21519. <https://doi.org/10.3897/mbmg.2.21519>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), Art. 7–Art. 583. <https://doi.org/10.1038/nmeth.3869>
- Carballeira, R., & Pontevedra-Pombal, X. (2021). Diversity of testate amoebae as an indicator of the conservation status of peatlands in Southwest Europe. *Diversity*, 13(6), Art. 6. <https://doi.org/10.3390/d13060269>
- Christianson, L. M., Johnson, S. B., Schultz, D. T., & Haddock, S. H. D. (2022). Hidden diversity of Ctenophora revealed by new mitochondrial *COI* primers and sequences. *Molecular Ecology Resources*, 22(1), 283–294. <https://doi.org/10.1111/1755-0998.13459>
- Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., Iribar, A., Taberlet, P., Vigouroux, R., & Brosse, S. (2019). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources*, 19(1), 27–46. <https://doi.org/10.1111/1755-0998.12900>
- Cockburn, C. F., Gregory, B. R. B., Nasser, N. A., & Patterson, R. T. (2020). Intra-Lake Arcellinida (testate Lobose amoebae) response to winter De-icing contamination in an eastern Canada road-side “Salt Belt” Lake. *Microbial Ecology*, 80(2), 366–383. <https://doi.org/10.1007/s00248-020-01513-w>
- Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., Genner, M. J., & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with *COI*. *Methods in Ecology and Evolution*, 10(11), 1985–2001. <https://doi.org/10.1111/2041-210X.13276>
- Czechowski, P., Stevens, M. I., Madden, C., & Weinstein, P. (2020). Steps towards a more efficient use of chironomids as bioindicators for freshwater bioassessment: Exploiting eDNA and other genetic tools. *Ecological Indicators*, 110, 105868. <https://doi.org/10.1016/j.ecolind.2019.105868>
- de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences*, 102(suppl\_1), 6600–6607. <https://doi.org/10.1073/pnas.0502030102>
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, 10(9), 20140562. <https://doi.org/10.1098/rsbl.2014.0562>
- del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L. F., Nenarokov, S., Massana, R., Guillou, L., Simpson, A., Berney, C., de Vargas, C., Brown, M. W., Keeling, P. J., & Parfrey, L. W. (2018). EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biology*, 16(9), e2005849. <https://doi.org/10.1371/journal.pbio.2005849>
- Dumack, K., Görzen, D., González-Miguéns, R., Siemensma, F., Lahr, D. J. G., Lara, E., & Bonkowski, M. (2020). Molecular investigation of Phryganella acropodia Hertwig et lesser, 1874 (Arcellinida, Amoebozoa). *European Journal of Protistology*, 75, 125707. <https://doi.org/10.1016/j.ejop.2020.125707>
- Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, 4(4), 423–425. <https://doi.org/10.1098/rsbl.2008.0118>
- Foissner, W., & Korganova, G. A. (2000). The Centropxyxis aerophila Complex (protozoa: Testacea). *Acta Protozoologica*, 39, 257–273.
- 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. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.
- Freitas, Y. d. G. C., Ramos, B. R. D., Miranda, V. B. D. S., da Silva, Y. G., Sampaio, G. S., Nascimento, L. D. S., & Branco, C. W. C. (2022). Testate amoebae: A review on their multiple uses as bioindicators. *Acta Protozoologica*, 61, Art. Volume 61. <https://doi.org/10.4467/16890027AP.22.001.15671>
- Gaither, M. R., DiBattista, J. D., Leray, M., & von der Heyden, S. (2022). Metabarcoding the marine environment: From single species to biogeographic patterns. *Environmental DNA*, 4(1), 3–8. <https://doi.org/10.1002/edn3.270>
- Gilbert, D., & Mitchell, E. A. D. (2006). Chapter 13 microbial diversity in sphagnum peatlands. In *En developments in earth surface processes* (Vol. 9, pp. 287–318). Elsevier. [https://doi.org/10.1016/S0928-2025\(06\)09013-4](https://doi.org/10.1016/S0928-2025(06)09013-4)
- Girard, E. B., Langerak, A., Jompa, J., Wangensteen, O. S., Macher, J.-N., & Renema, W. (2022). Mitochondrial cytochrome oxidase subunit 1: A promising molecular marker for species identification in foraminifera. *Frontiers in Marine Science*, 9, 809659. <https://www.frontiersin.org/article/10.3389/fmars.2022.809659>
- Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7(11), 1299–1307. <https://doi.org/10.1111/2041-210X.12595>
- González-Miguéns, R., Soler-Zamora, C., Useros, F., Nogal-Prata, S., Berney, C., Blanco-Rotea, A., Carrasco-Braganza, M. I., de Salvador-Velasco, D., Guillén-Oterino, A., Tenorio-Rodríguez, D., Velázquez, D., Heger, T. J., Sanmartín, I., & Lara, E. (2022). *Cyphoderia ampulla* (Cyphoderiidae: Rhizaria), a tale of freshwater sailors: The causes and consequences of ecological transitions through the salinity barrier in a family of benthic protists. *Molecular Ecology*, 31(9), 2644–2663. <https://doi.org/10.1111/mec.16424>
- González-Miguéns, R., Soler-Zamora, C., Villar-Depablo, M., Todorov, M., & Lara, E. (2022). Multiple convergences in the evolutionary history of the testate amoeba family Arcellidae (Amoebozoa: Arcellinida: Sphaerothecina): When the ecology rules the morphology.

- Zoological Journal of the Linnean Society*, 194(4), 1044–1071. <https://doi.org/10.1093/zoolinlean/zlab074>
- González-Miguéns, R., Todorov, M., Blandenier, Q., Duckert, C., Porfirio-Sousa, A. L., Ribeiro, G. M., Ramos, D., Lahr, D. J. G., Buckley, D., & Lara, E. (2022). Deconstructing Diffugia: The tangled evolution of lobose testate amoebae shells (Amoebozoa: Arcellinida) illustrates the importance of convergent evolution in protist phylogeny. *Molecular Phylogenetics and Evolution*, 107557, 107557. <https://doi.org/10.1016/j.ympev.2022.107557>
- Govender, A., Singh, S., Groeneveld, J., Pillay, S., & Willows-Munro, S. (2022). Experimental validation of taxon-specific mini-barcode primers for metabarcoding of zooplankton. *Ecological Applications*, 32(1), e02469. <https://doi.org/10.1002/eap.2469>
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., ... Christen, R. (2013). The Protist ribosomal reference database (PR<sup>2</sup>): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41(Database issue), D597–D604. <https://doi.org/10.1093/nar/gks1160>
- Hajjibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLoS One*, 14(9), e0220953. <https://doi.org/10.1371/journal.pone.0220953>
- Hamsher, S. E., Evans, K. M., Mann, D. G., Pouličková, A., & Saunders, G. W. (2011). Barcoding diatoms: Exploring alternatives to COI-5P. *Protist*, 162(3), 405–422. <https://doi.org/10.1016/j.protis.2010.09.005>
- Heger, T. J., Lara, E., & Mitchell, E. A. D. (2011). Arcellinida testate amoebae (Amoebozoa: Arcellinida): Model of organisms for assessing microbial biogeography. In E. D. Fontaneto (Ed.), *Biogeography of microscopic organisms* (pp. 111–129). Cambridge University Press. <https://doi.org/10.1017/CBO9780511974878.008>
- Heger, T. J., Pawłowski, J., Lara, E., Leander, B. S., Todorov, M., Golemansky, V., & Mitchell, E. A. D. (2011). Comparing potential COI and SSU rDNA barcodes for assessing the diversity and phylogenetic relationships of Cyphoderiid testate amoebae (Rhizaria: Euglyphida). *Protist*, 162(1), 131–141. <https://doi.org/10.1016/j.protis.2010.05.002>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Kalyanamorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), art. 6. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.-M., Humbert, J.-F., & Bouchez, A. (2014). A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, 33(1), 349–363. <https://doi.org/10.1086/675079>
- Kermarrec, L., Franc, A., Rimet, F., Chaumeil, P., Humbert, J. F., & Bouchez, A. (2013). Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. *Molecular Ecology Resources*, 13(4), 607–619. <https://doi.org/10.1111/1755-0998.12105>
- 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(3), 415–434. <https://doi.org/10.1016/j.protis.2011.10.003>
- Kostka, M., Lares-Jiménez, L. F., Tymi, T., & Dyková, I. (2017). *Copromyxa laresi* n. Sp. (Amoebozoa: Tubulinea) and transfer of *Cashia limaoides* (page, 1967) to *Copromyxa* Zopf, 1885. *Journal of Eukaryotic Microbiology*, 64(2), 173–182. <https://doi.org/10.1111/jeu.12349>
- Lahr, D. J., Kosakyan, A., Lara, E., Mitchell, E. A., Morais, L., Porfirio-Sousa, A. L., Ribeiro, G. M., Tice, A. K., Pánek, T., Kang, S., & Brown, M. W. (2019). Phylogenomics and morphological reconstruction of Arcellinida testate amoebae highlight diversity of microbial eukaryotes in the Neoproterozoic. *Current Biology*, 29(6), 991–1001.
- Lara, E., Dumack, K., García-Martin, J. M., Kudryavtsev, A., & Kosakyan, A. (2020). Amoeboid protist systematics: A report on the “systematics of amoeboid protists” symposium at the VIIIth ECOP/ISOP meeting in Rome, 2019. *European Journal of Protistology*, 76, 125727. <https://doi.org/10.1016/j.ejop.2020.125727>
- Lara, E., Singer, D., & Geisen, S. (2022). Discrepancies between prokaryotes and eukaryotes need to be considered in soil DNA-based studies. *Environmental Microbiology*, 11, 3829–3839.
- Laroche, O., Pochon, X., Wood, S. A., & Keeley, N. (2021). Beyond taxonomy: Validating functional inference approaches in the context of fish-farm impact assessments. *Molecular Ecology Resources*, 21(7), 2264–2277. <https://doi.org/10.1111/1755-0998.13426>
- Lentendu, G., Lara, E., & Geisen, S. (2023). Metabarcoding approaches for soil eukaryotes, Protists, and microfauna. In E. F. Martin & S. Uroz (Eds.), *Microbial environmental genomics (MEG)* (pp. 1–16). Springer US. [https://doi.org/10.1007/978-1-0716-2871-3\\_1](https://doi.org/10.1007/978-1-0716-2871-3_1)
- Lezcano, M. Á., Velázquez, D., Quesada, A., & El-Shehawry, R. (2017). Diversity and temporal shifts of the bacterial community associated with a toxic cyanobacterial bloom: An interplay between microcystin producers and degraders. *Water Research*, 125, 52–61. <https://doi.org/10.1016/j.watres.2017.08.025>
- Macher, J.-N., Wideman, J. G., Girard, E. B., Langerak, A., Duijm, E., Jompa, J., Sadekov, A., Vos, R., Wissels, R., & Renema, W. (2021). First report of mitochondrial COI in foraminifera and implications for DNA barcoding. *Scientific Reports*, 11(1), Art. 1. <https://doi.org/10.1038/s41598-021-01589-5>
- Macumber, A. L., Roe, H. M., Prentice, S. V., Sayer, C. D., Bennion, H., & Salgado, J. (2020). Freshwater testate amoebae (Arcellinida) response to eutrophication as revealed by test size and shape indices. *Frontiers in Ecology and Evolution*, 8, 568904. <https://www.frontiersin.org/article/10.3389/fevo.2020.568904>
- Marcisz, K., Colombaroli, D., Jassey, V. E. J., Tinner, W., Kołaczek, P., Gałka, M., Karpińska-Kołaczek, M., Stowiński, M., & Lamentowicz, M. (2016). A novel testate amoebae trait-based approach to infer environmental disturbance in sphagnum peatlands. *Scientific Reports*, 6(1), 33907. <https://doi.org/10.1038/srep33907>
- Marcisz, K., Jassey, V. E. J., Kosakyan, A., Krashevska, V., Lahr, D. J. G., Lara, E., Lamentowicz, Ł., Lamentowicz, M., Macumber, A., Mazei, Y., Mitchell, E. A. D., Nasser, N. A., Patterson, R. T., Roe, H. M., Singer, D., Tsyganov, A. N., & Fournier, B. (2020). Testate amoeba functional traits and their use in paleoecology. *Frontiers in Ecology and Evolution*, 8, 575966. <https://doi.org/10.3389/fevo.2020.575966>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), Art. 1. <https://doi.org/10.14806/ej.17.1.200>
- Mazei, Y., Chernyshov, V., Tsyganov, A. N., & Payne, R. J. (2015). Testing the effect of refrigerated storage on testate amoeba samples. *Microbial Ecology*, 70(4), 861–864. <https://doi.org/10.1007/s0024-8-015-0628-1>
- Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models

- and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Moniz, M. B. J., & Kaczmarek, I. (2009). Barcoding diatoms: Is there a good marker? *Molecular Ecology Resources*, 9(s1), 65–74. <https://doi.org/10.1111/j.1755-0998.2009.02633.x>
- Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., João Feio, M., & Almeida, F. P. S. (2019). Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. *Ecological Indicators*, 106, 105470. <https://doi.org/10.1016/j.ecolind.2019.105470>
- Nasser, N. A., Patterson, R. T., Roe, H. M., Galloway, J. M., Falck, H., & Sanei, H. (2020). Use of Arcellinida (testate lobose amoebae) arsenic tolerance limits as a novel tool for biomonitoring arsenic contamination in lakes. *Ecological Indicators*, 113, 106177. <https://doi.org/10.1016/j.ecolind.2020.106177>
- 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(1), 102–115. <https://doi.org/10.1016/j.protis.2009.07.003>
- Patterson, R. T., & Kumar, A. (2002). A review of current testate rhizopod (thecamoebian) research in Canada. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 180(1–3), 225–251. [https://doi.org/10.1016/S0031-0182\(01\)00430-8](https://doi.org/10.1016/S0031-0182(01)00430-8)
- 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., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., ... de Vargas, C. (2012). CBOL Protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, 10(11), e1001419. <https://doi.org/10.1371/journal.pbio.1001419>
- Pawlowski, J., Bonin, A., Boyer, F., Cordier, T., & Taberlet, P. (2021). Environmental DNA for biomonitoring. *Molecular Ecology*, 30(13), 2931–2936. <https://doi.org/10.1111/mec.16023>
- Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M. J., Filipe, A. F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz, C., ... Kahlert, M. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment*, 637–638, 1295–1310. <https://doi.org/10.1016/j.scitotenv.2018.05.002>
- Pawlowski, J., Lejzerowicz, F., Apothéloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, 55, 12–25. <https://doi.org/10.1016/j.ejop.2016.02.003>
- Payne, R. J. (2013). Seven reasons why Protists make useful bioindicators. *Acta Protozoologica*, 52, 105–113.
- Pfleger, M. O., Rider, S. J., Johnston, C. E., & Janosik, A. M. (2016). Saving the doomed: Using eDNA to aid in detection of rare sturgeon for conservation (Acipenseridae). *Global Ecology and Conservation*, 8, 99–107. <https://doi.org/10.1016/j.gecco.2016.08.008>
- Pinseel, E., Janssens, S. B., Verleyen, E., Vanormelingen, P., Kohler, T. J., Biersma, E. M., Sabbe, K., Van de Vijver, B., & Vyverman, W. (2020). Global radiation in a rare biosphere soil diatom. *Nature Communications*, 11(1), Art. 1. <https://doi.org/10.1038/s41467-020-16181-0>
- Richards, T. A., Jones, M. D. M., Leonard, G., & Bass, D. (2012). Marine fungi: Their ecology and molecular diversity. *Annual Review of Marine Science*, 4(1), 495–522. <https://doi.org/10.1146/annurev-marine-120710-100802>
- Rimet, F., Gusev, E., Kahlert, M., Kelly, M. G., Kulikovskiy, M., Maltsev, Y., Mann, D. G., Pfannkuchen, M., Trobajo, R., Vasselon, V., Zimmermann, J., & Bouchez, A. (2019). Diat.Barcode, an open-access curated barcode library for diatoms. *Scientific Reports*, 9(1), Art. 1. <https://doi.org/10.1038/s41598-019-51500-6>
- Roffler, G. H., Allen, J. M., Massey, A., & Levi, T. (2021). Metabarcoding of fecal DNA shows dietary diversification in wolves substitutes for ungulates in an Island archipelago. *Ecosphere*, 12(1), e03297. <https://doi.org/10.1002/ecs2.3297>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. <https://doi.org/10.7717/peerj.2584>
- Ruggiero, A., Grattepanche, J., Weiner, A. K. M., & Katz, L. A. (2020). High diversity of testate amoebae (Amoebozoa, Arcellinida) detected by HTS analyses in a New England fen using newly designed taxon-specific primers. *Journal of Eukaryotic Microbiology*, 67(4), 450–462. <https://doi.org/10.1111/jeu.12794>
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15(6), 1289–1303. <https://doi.org/10.1111/1755-0998.12402>
- Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research: NEXT-GENERATION SEQUENCING FOR ENVIRONMENTAL DNA. *Molecular Ecology*, 21(8), 1794–1805. <https://doi.org/10.1111/j.1365-294X.2012.05538.x>
- Singer, D., Kosakyan, A., Pillonel, A., Mitchell, E. A. D., & Lara, E. (2015). Eight species in the Nebela collaris complex: Nebela gimlii (Arcellinida, Hyalospheniidae), a new species described from a Swiss raised bog. *European Journal of Protistology*, 51(1), 79–85. <https://doi.org/10.1016/j.ejop.2014.11.004>
- Singer, D., Kosakyan, A., Seppé, C. V. W., Pillonel, A., Fernández, L. D., Fontaneto, D., Mitchell, E. A. D., & Lara, E. (2018). Environmental filtering and phylogenetic clustering correlate with the distribution patterns of cryptic protist species. *Ecology*, 99(4), 904–914. <https://doi.org/10.1002/ecy.2161>
- Singer, D., Mitchell, E. A. D., Payne, R. J., Blandenier, Q., Duckert, C., Fernández, L. D., Fournier, B., Hernández, C. E., Granath, G., Rydin, H., Bragazza, L., Koronátova, N. G., Goia, I., Harris, L. I., Kajukato, K., Kosakyan, A., Lamentowicz, M., Kosykh, N. P., Vellak, K., & Lara, E. (2019). Dispersal limitations and historical factors determine the biogeography of specialized terrestrial protists. *Molecular Ecology*, 28(12), 3089–3100. <https://doi.org/10.1111/mec.15117>
- Smith, H. G., Bobrov, A., & Lara, E. (2009). Diversity and biogeography of testate amoebae. In E. W. Foissner & D. L. Hawksworth (Eds.), *Protist diversity and geographical distribution* (pp. 95–109). Springer Netherlands. [https://doi.org/10.1007/978-90-481-2801-3\\_8](https://doi.org/10.1007/978-90-481-2801-3_8)
- Smith, H. G., & Wilkinson, D. M. (2007). Not all free-living microorganisms have cosmopolitan distributions? The case of Nebela (Apodera) vas certes (protozoa: Amoebozoa: Arcellinida). *Journal of Biogeography*, 34(10), 1822–1831. <https://doi.org/10.1111/j.1365-2699.2007.01733.x>
- Steele, R. E., Nasser, N. A., Patterson, R. T., Gregory, B. R. B., Roe, H. M., & Reinhardt, E. G. (2018). An assessment of sub-meter scale spatial variability of Arcellinida (testate Lobose amoebae) assemblages in a temperate Lake: Implications for Limnological studies. *Microbial Ecology*, 76(3), 680–694. <https://doi.org/10.1007/s00248-018-1157-5>
- Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA: ENVIRONMENTAL DNA. *Molecular Ecology*, 21(8), 1789–1793. <https://doi.org/10.1111/j.1365-294X.2012.05542.x>
- Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences USA*, 109(40), 16208–16212. <https://doi.org/10.1073/pnas.1209160109>
- Tekle, Y. I. (2014). DNA barcoding in Amoebozoa and challenges: The example of *Cochliopodium*. *Protist*, 165(4), 473–484. <https://doi.org/10.1016/j.protis.2014.05.002>

- Todorov, M., & Bankov, N. (2019). An atlas of sphagnum-dwelling testate amoebae in Bulgaria, an atlas of sphagnum-dwelling testate amoebae in Bulgaria. *Advanced Books*, 1, 1–286. <https://doi.org/10.3897/ab.e38685>
- Turunen, J., Mykrä, H., Elbrecht, V., Steinke, D., Braukmann, T., & Aroviita, J. (2021). The power of metabarcoding: Can we improve bioassessment and biodiversity surveys of stream macroinvertebrate communities? *Metabarcoding and Metagenomics*, 5, e68938. <https://doi.org/10.3897/mbmg.5.68938>
- Valentine, J., Davis, S. R., Kirby, J. R., & Wilkinson, D. M. (2013). The use of testate amoebae in monitoring peatland restoration management: Case studies from north West England and Ireland. *Acta Protozoologica*, 52(3), 129–145. Special topic issue: "Protists as Bioindicators of Past and Present Environmental Conditions", Art. Volume 52, Issue 3, Special topic issue: "Protists as Bioindicators of Past and Present Environmental Conditions". <https://doi.org/10.4467/16890027AP.13.0013.1110>
- Van Hengstum, P. J., Reinhardt, E. G., Beddows, P. A., Huang, R. J., & Gabriel, J. J. (2008). Thecamoebians (*Testate amoebae*) and foraminifera from three anchialine cenotes in Mexico: low salinity (1.5–4.5 psu) faunal transitions. *The Journal of Foraminiferal Research*, 38(4), 305–317. <https://doi.org/10.2113/gsjfr.38.4.305>
- Vaulot, D., Geisen, S., Mahé, F., & Bass, D. (2022). pr2-primers: An 18S rRNA primer database for protists. *Molecular Ecology Resources*, 22(1), 168–179. <https://doi.org/10.1111/1755-0998.13465>
- Villacorta-Rath, C., Espinoza, T., Cockayne, B., Schaffer, J., & Burrows, D. (2022). Environmental DNA analysis confirms extant populations of the cryptic Irwin's turtle within its historical range. *BMC Ecology and Evolution*, 22(1), 57. <https://doi.org/10.1186/s12862-022-02009-6>
- Wagih, O. (2017). Ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics*, 33(22), 3645–3647. <https://doi.org/10.1093/bioinformatics/btx469>
- Wang, W., Gao, X., Ndayishimiye, J. C., Lara, E., Lahr, D. J. G., Qian, H., Ren, K., Chen, H., & Yang, J. (2022). Population and molecular responses to warming in *Netzelia tuberspinifera* – An endemic and sensitive protist from East Asia. *Science of the Total Environment*, 806, 150897. <https://doi.org/10.1016/j.scitotenv.2021.150897>
- Wanner, M., & Dunger, W. (2001). Biological activity of soils from reclaimed open-cast coal mining areas in upper Lusatia using testate amoebae (protists) as indicators. *Ecological Engineering*, 17(2–3), 323–330. [https://doi.org/10.1016/S0925-8574\(00\)00148-8](https://doi.org/10.1016/S0925-8574(00)00148-8)
- West, K., Travers, M. J., Stat, M., Harvey, E. S., Richards, Z. T., DiBattista, J. D., Newman, S. J., Harry, A., Skepper, C. L., Heydenrych, M., & Bunce, M. (2021). Large-scale eDNA metabarcoding survey reveals marine biogeographic break and transitions over tropical North-Western Australia. *Diversity and Distributions*, 27(10), 1942–1957. <https://doi.org/10.1111/ddi.13228>
- Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Systematic Zoology*, 27(1), 17–26. <https://doi.org/10.2307/2412809>
- Yang, J., Harding, T., Kamikawa, R., Simpson, A. G. B., & Roger, A. J. (2017). Mitochondrial genome evolution and a novel RNA editing system in deep-branching Heteroloboseids. *Genome Biology and Evolution*, 9(5), 1161–1174. <https://doi.org/10.1093/gbe/evx086>
- Yang, Z.-C., Wang, Z.-H., & Zhang, Z.-H. (2011). Biomonitoring of testate amoebae (protozoa) as toxic metals absorbed in aquatic bryophytes from the Hg-Tl mineralized area (China). *Environmental Monitoring and Assessment*, 176(1–4), 321–329. <https://doi.org/10.1007/s10661-010-1585-2>
- Yoccoz, N. G. (2012). The future of environmental DNA in ecology. *Molecular Ecology*, 21(8), 2031–2038. <https://doi.org/10.1111/j.1365-294X.2012.05505.x>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** González-Miguéns, R., Cano, E., Guillén-Oterino, A., Quesada, A., Lahr, D. J. G., Tenorio-Rodríguez, D., de Salvador-Velasco, D., Velázquez, D., Carrasco-Braganza, M. I., Patterson, R. T., Lara, E., & Singer, D. (2023). A needle in a haystack: A new metabarcoding approach to survey diversity at the species level of Arcellinida (Amoebozoa: Tubulinea). *Molecular Ecology Resources*, 00, 1–16. <https://doi.org/10.1111/1755-0998.13771>